Parametrization of Force Field Bonded Terms under Structural Inconsistency

Anastasia Croitoru, Alexey Aleksandrov

▶ **To cite this version:**

Anastasia Croitoru, Alexey Aleksandrov. Parametrization of Force Field Bonded Terms under Structural Inconsistency. Journal of Chemical Information and Modeling, 2022, pp.4471-4482. 10.1021/acs.jcim.2c00950 . hal-03779485

HAL Id: hal-03779485
https://hal.science/hal-03779485

Submitted on 24 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Parametrization of Force Field Bonded Terms under Structural Inconsistency

Anastasia Croitoru[1] and Alexey Aleksandrov[1*]

[1]Laboratoire d'Optique et Biosciences (CNRS UMR7645, INSERM U1182), Ecole Polytechnique, Institut polytechnique de Paris, F-91128 Palaiseau, France

*Corresponding authors: alexey.aleksandrov@polytechnique.edu

Running title: Parametrization of Force Field Bonded Terms under Structural Inconsistency

ABSTRACT

Parametrization of the bonded part of molecular mechanics (MM) force fields (FFs) is typically done by fitting reference quantum mechanical (QM) energies or forces of representative structures. FFs for small molecules are constructed in incremental parametrization procedures, where parameters developed previously are retained for novel molecules, followed by optimization of missing, not previously optimized parameters. Equilibrium QM and MM geometries of molecules can deviate due to parameters transferred from existing molecules in the FF. In this work, we demonstrate that conventional parametrization methods based on fitting QM energies and/or forces to derive parameters for bond and angle terms produce largely suboptimal force constants when MM and QM equilibrium structures deviate. We further developed and tested a new method to derive CHARMM FF parameters based on the potential energy surface scans where a structural deviation between QM and MM optimized geometries is explicitly allowed during parametrization. The test of the new method was performed on a diverse set of 32 molecules. The results show that without any need for additional restraints the new method produces robust and largely transferable parameters for bond and angle terms. The new method also improves the agreement for the normal modes for all molecules in the test set, reducing the average error in the reproduction of QM normal mode frequencies from 9.5% computed with CGenFF parameters, to 6.8% computed with the new parameters. The new method will allow parametrization of molecules under structural deviations, common for force fields for small molecules, producing robust and transferable parameters.

INTRODUCTION

Molecular modeling techniques based on molecular mechanics (MM) are widely used in computer-aided drug design (CADD) and in studies of biological processes. The accuracy of MM based simulations in predicting the properties of chemical and biological systems strongly depends on the quality of the underlying force-field (FF) parameters. The MM model with associated FF parameters should be able to describe molecular equilibrium structures, conformational energies, and intermolecular interactions. The widely used Class I additive force fields for biological systems such proteins, nucleic acids and lipids include CHARMM[1–3], AMBER[4–8], GROMOS[9], and OPLS[10,11]. For CADD applications, force fields were extended for small molecules, such as OPLS[12], the CHARMM General force field (CGenFF) force field[13–15], the General AMBER Force Field (GAFF)[16], GROMOS[17] and Open Force Field.[18] The coverage of the vast chemical universe required for the field of CADD in these force fields is still limited. The main limitation is the complexity of developing force field parameters for novel chemical molecules, such as modified amino acids and nucleic acids.[19,20]

Force fields for biomolecules can be derived for molecules of interest from scratch[21] or more typically constructed in the incremental parametrization procedure.[22] Following this strategy, at each incremental step parameters developed previously are retained for novel molecules, followed by optimization of missing, not previously optimized, parameters.[23] However, such force fields critically rely on the transferability of parameters. Thus, the quality of FF parameters, in addition to the ability of the MM model to reproduce QM energies and structures for the molecule in which these parameters were derived, includes the ability to work equally well in all range of molecules, which share atom types and terms. To allow the rapid extension of the existing force field to novel molecules automated tools have been created to allow automatic atom typing and generating initial parameters for molecules, such as the CHARMM General Force Field program[13] for CGenFF, LigParGen for the OPLS-AA FF[24], and Antechamber for the Amber FF.[25] To further optimize the initial parameters, a number of tools have been proposed with the idea of automation.[26–33]

In this work, we mainly focus on bond and angle terms of the bonded part of the total FF energy. In widely used Class I additive force fields these terms are modeled by harmonic functions to describe deformations along bonds and angles around their equilibrium values. Parameters for these terms can be obtained by reproducing the experimental or/and QM vibrational spectrum, Hessian matrix[34,35], and deformation energies and forces.[36] In the later methods, also known as the Force Matching methods, MM parameters are fitted to reproduce the forces in non-equilibrium structures, which can be generated, for example, by classical MD simulations.[33,36–39] For the CHARMM General force field for small molecules, to parametrize stiff degrees of freedom a symbolic potential energy distribution (PED) analysis was performed in the internal coordinate space.[40] This allows estimating relative contributions of the valence coordinates to frequencies. These contributions are computed using the MM Hessian calculated using the trial parameters, and compared with the corresponding QM PED; parameters are iteratively varied until satisfactory agreement is reached. In practice, the fitting is difficult since QM and MM frequencies of a normal mode as well as QM and MM contributions of internal coordinates to the same normal

mode are different. With this, the quality of the fit is difficult to quantify and in addition, one needs to define a non-unique mapping between internal coordinates and normal modes, which is difficult to automate.

Apart from PED analysis, a method to determine force constants by three-point PES scans was used for CGenFF when the assignment of the internal coordinate contributions to the vibrations was ambiguous.[41] This method is also implemented in the Force Field Toolkit (ffTk), a VMD plugin[42] that can be used to parametrize the CHARMM force field for small molecules.[43] In this method, a small distortion in two opposing directions is generated and the corresponding increase in potential energy relative to the undistorted conformation is computed. The QM Hessian is used to compute QM energy for the small distortions about the minimized geometry. The energies are scaled to improve the agreement with experimental vibrational frequencies.[41] Since no optimization is done for the deformed structures, in principle different sets of parameters can reproduce QM energies equally well. Thus, the parameter optimization problem is ill-defined in this case, and requires a restraining strategy. Different such restraint strategies have been proposed.[41,43–45] However, introducing such artificial restraints may result in a poor transferability of parameters for molecules that were not used for the optimization. In particular, bond and angle parameters are developed typically only in one molecule and used for all other molecules in the chemical universe sharing the same term defined by atom types.

Equilibrium QM and MM geometries of molecules can deviate due to parameters transferred from existing molecules in the force field, even with FF parameters for new terms optimized against QM reference data. For example, in the previous work on the parametrization of the large set of nonstandard amino acids we found that bonds and angles deviate on average by 0.02 Å and 2º, respectively, for a large set of 189 compounds after optimization of new parameters not existing in CGenFF.[20] In this work, we demonstrate that, while these MM structural inconsistences relative to QM optimized structures can be negligible for applications, they strongly impact the quality of new parameters optimized in novel molecules. We further show that the conventional methods to derive parameters for bond and angle terms produce largely suboptimal force constants when MM and QM equilibrium structures deviate even slightly. This problem arises if the same structures (for example, QM optimized structures) are used during parametrization for QM and MM calculations, or QM and MM structures have the same value of the deformed bond or angle. We further developed and tested a new method to derive force field parameters based on the PES scans where a structural deviation between QM and MM structures is explicitly allowed and show that the new method produces stable and transferable parameters for bond and angle terms without any need for additional restraints.

METHODS AND MATERIALS

CHARMM potential energy function

For specifics, we will demonstrate the development and application of the new method using the non-polarizable all-atom CHARMM force field.[46] The potential energy function, used for the CHARMM36 and CGenFF force fields has two contributions:

$U = U_{inter} + U_{intra}$ [1]

The intermolecular or non-bonded energy is due to electrostatic and van der Waals (vdW) interactions and its form can be found in previous works.[46] The electrostatic and vdW energy are not computed for directly linked atoms as well as for atoms forming covalence angles; for 1-4 atoms, in the CHARMM force field, it is computed without scaling.[46] In this work, we are interested in the bonded or intramolecular part of the potential energy function in Equation 1, which is contributed by terms for the bonds, valence angles, dihedral angles, improper dihedral angles, and selected Urey-Bradley terms. The bonded contribution is given by:

$$U_{intra} = \sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_a(\theta - \theta_0)^2 + \sum_{Urey-Bradley} K_{UB}(r_{1-3} - r_{1-3;0})^2 + \sum_{dihedral} \sum_{n=1}^{N} K_n\,(1 + \cos(n\varphi - \delta_n)) + \sum_{improper} K_\varphi(\varphi - \varphi_0)^2$$

, [2]

where b_0, θ_0, $r_{1-3;0}$, and φ_0 are the bond, angle, Urey-Bradley, and improper angle equilibrium values, respectively; *K*'s are the force constants; and *n* and δ_n are the dihedral multiplicity and phase. A dihedral term is represented as a truncated Fourier series with *N* number of multiplicities. In addition, the bonded energy function includes the CMAP cross-term in the current version C36 of the CHARMM force field only for the peptide backbone.[2] The Urey-Bradley contribution, a cross-term accounting for angle bending and bond stretching was introduced to improve the agreement with the QM vibrational spectrum. However, we will not consider this term as in the current CGenFF force field it is not widely used.

QM potential energy surface scans

The reference QM data were generated using adiabatic PES scans for each selected degree of freedom. The same method was also used in CGenFF to determine force constants by three-point PES scans and in our previous work to parametrize a set of non-standard amino acids.[41,47] During the PES scans performed by varying one stiff degree of freedom, all rotatable dihedrals were constrained to the values corresponding to the minimum energy geometry of the molecule. We use the following method to ensure that only relevant regions of PES are used for parametrization and limit the maximum energy.[19] Initial values for distortions are used to estimate the ranges of deformations according to:

$$\Delta x' = \sqrt{2\,\Delta E_{max}/k}, \text{ [3]}$$

where $k = 2\,(E(\Delta x) - E_0)/\Delta x^{\,2}$. Δx and $\Delta x'$ are the initial and adjusted maximum distortions, respectively; E_0 and $E(\Delta x)$ are the minimum energy and energy of the deformed structure. ΔE_{max} defines the highest energy of points on PES. The structure deformed along the scanned degree of freedom by Δx is minimized to obtain $E(\Delta x)$, which is used in Equation 3 to get the new range of points $\Delta x'$. PES scans are performed for a series of points equally spaced in the range of $x \in [x_0 - \Delta x',\ x_0 + \Delta x']$, including the minimum energy structure at $x = x_0$. 2.0

kcal·mol^{-1} was used for ΔE_{max} in Equation 3. All PES scans were performed at the MP2/6-31G(d) model chemistry and MP2/6-311G(d) for anions.

The optimization of parameters is done using the Powell minimization algorithms from Numerical Recipes.[48] Each conformation for the MM calculation was extracted from the QM scan and minimized with a large restraint force constant of $5 \cdot 10^4$ kcal·mol^{-1}·Å^{-2} or $5 \cdot 10^4$ kcal·mol^{-1}·radian^{-2} on the target bond and valance angle, respectively. PES adiabatic scans were performed with CHARMM using a new set of MM parameters at each optimization iteration of bonded parameters. The MM parameters were adjusted until the target function could not be reduced further. The CHARMM program[49] was used for the MM calculations.

Optimization and cost function

Typically, to optimize bonded parameters a cost function is constructed using energy differences between QM and MM structures, for example from PES scans, which is further minimized to give an optimal set of bonded parameters. In the simplest form, the RMS deviation between QM and MM energies for a set of structures can be used for the cost function:

$$F^{ener} = \sqrt{\frac{1}{N}\sum_i (E^{QM}(\boldsymbol{x}_i) - E^{MM}(\boldsymbol{x}_i))^2}, \quad [4]$$

where the sum is over a set of *N* structures; E^{QM} and E^{MM} are QM and MM energies relative to the QM or MM minimum energy, respectively, computed using the same set of coordinates, $\boldsymbol{x}_i$. The structures can be generated in different ways, however, in this work, we will focus on potential energy surface scans, described above. For calculations with the force field model, structures can correspond to QM optimized structures, or typically, they can be re-optimized with the force field model. Thus, QM and MM structures can be different, however, the scanned valence coordinate (for example a dihedral angle) has the same value in the QM and MM structures. In this case, the RMS deviation is given by:

$$F^{ener} = \sqrt{\frac{1}{N}\sum_i (E^{QM}(\boldsymbol{x}_i^{QM}) - E^{MM}(\boldsymbol{x}_i^{MM}))^2}, \quad [5]$$

where the two sets of coordinates, $\boldsymbol{x}_i^{QM}$ and $\boldsymbol{x}_i^{MM}$ have the same value along the scanned internal degree of freedom, *k*: $x_{i,k}^{QM} = x_{i,k}^{MM}$, but can be different along other degrees of freedom in contrast to Equation 4.

Modification of the cost function

To remove the strong dependence of parameters on equilibrium parameters and to improve force constants optimization, we adjusted the optimization method.We allow structural deviations along the scanned degree of freedom, where energy differences are computed between different structures used for QM and MM calculations. These QM and MM structures are now different, in principle, in all coordinates. One of the requirements for such MM structures is that QM and MM structures, which energies are compared in Formula 5 are close to each other in the configurational space, in another words the MM and QM PES approximately match. We use a similar method described above where the MM structures are optimized with one internal coordinate constrained,

however, in the new method the value x_k^{MM} of the constrained coordinate, *k*, is allowed to deviate from x_k^{QM} and is given by:

$$x_k^{MM} = x_{0,k}^{MM} + (x_k^{QM} - x_{0,k}^{QM}), \text{ [6]}$$

where $x_{0,k}^{MM}$ and $x_{0,k}^{QM}$ are values in the MM and QM structures optimized structures without any constraints; $x_k^{QM}$is the value in the corresponding QM structure. The terms in the later formula can be regrouped:

$$x_k^{MM} = x_k^{QM} + \Delta x_{0,k}^{QM-MM}, \text{ [7]}$$

where $\Delta x_{0,k}^{QM-MM} = x_{0,k}^{MM} - x_{0,k}^{QM}$ is the deviation in the optimal values in the QM and MM optimized structures, which can be considered as a correction term. As mentioned before, one of requirements is that the QM and MM structures should be close on in the configurational space, which can be achieved by introducing an additional term, also needed to provide a bias in optimization of equilibrium parameters. In particular, we introduced an additional simple term given by:

$$F^{eq} = \sqrt{\frac{1}{N}\sum_i (x_{k,i}^{MM} - x_{k,i}^{QM})^2}, \text{ [8]}$$

where $x_{k,i}^{MM}$ and $x_{k,i}^{QM}$ is the scanned internal coordinate in the *i* MM and QM optimized structure, respectively.

Additional term for angle terms

For an atomic center, which has angle terms to parametrize, the following deformation-based term is included to the cost function in addition to the restraints on the scanned degree of freedom given by Formula 8:

$$F^{def} = \sqrt{\sum_j \sum_i ((\theta_{j,i}^{MM} - \theta_{j,i+1}^{MM}) - (\theta_{j,i}^{QM} - \theta_{j,i+1}^{QM}))^2}, \text{ [9]}$$

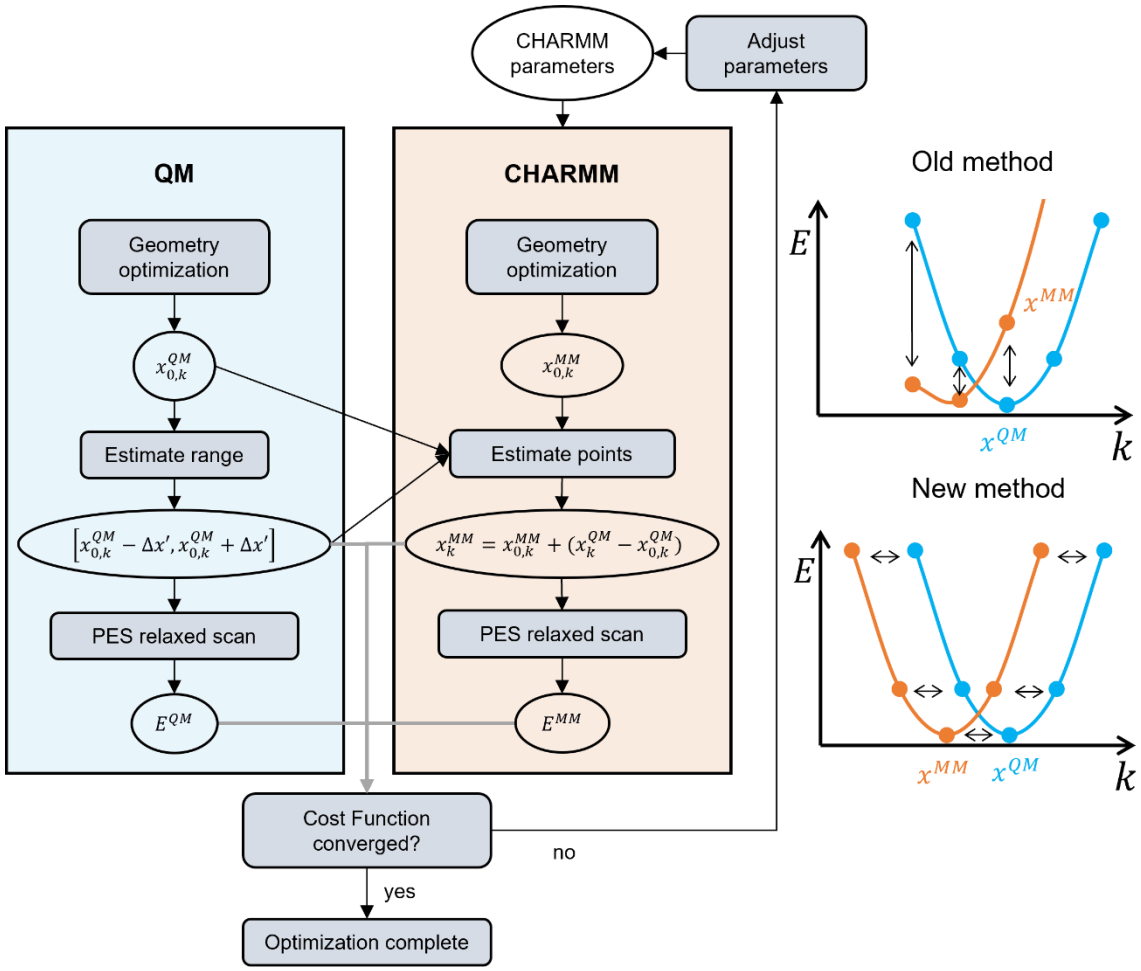
where the summing is done over all structures, *i*, and all angles, *j*, involving this atomic center with adjustable parameters in the FF model. In this formula, angles in the MM and QM structures, are subtracted between subsequent structures on PES similar to Equation 6 to remove strong dependences on the equilibrium values. The total cost function is a sum of different contributions given by Formulae 5, 8 and 9:

$$F^{cost} = \omega_{ener} \cdot F^{ener} + \omega_{eq} \cdot F^{eq} + \omega_{def} \cdot F^{def}, \text{ [10]}$$

where ω_{ener} , ω_{def}, and ω_{eq}, are corresponding weights. These weights were defined based on our previous experience with parametrizing a large set of molecules to reproduce average agreement for different energetic and structural properties. In particular, the weights were defined in such a way that the terms in Formula 10 give equal contributions (of a unity by definition) with the RMS deviations for energies and structural parameters obtained in the previous work.[20] The values for the weights are given in Table S2. All results presented in next sections were obtained with Equation 10, unless otherwise stated.

The flowchart of optimization with the new method is shown in Scheme 1. To produce QM reference data,

the geometry is first optimized without any constraints with QM to estimate equilibrium values, $x_{0,k}^{QM}$, for scanned degrees of freedom; another QM optimization is performed to estimate the range of distortions, $\Delta x'$, in accord with Equation 3, for the subsequent QM PES scan. CHARMM iterations also start with the geometry optimization without constraints to estimate equilibrium values, $x_{0,k}^{MM}$, which are used to correct points for the subsequent CHARMM PES scan in accord with Equation 6. Structures and energies are compared with the cost function 10, which is minimized in the space of CHARMM force field parameters.



Scheme 1. Flow chart of optimization of force field parameters with the new method. QM reference data are calculated in the left panel, while optimization of CHARMM parameters is performed in the CHARMM panel. The structural deviation allowed in the new method is shown schematically on the right.

Normal Mode Analysis

The quality of the optimized parameters was tested using normal mode analysis (NMA). QM NMA was performed for the molecules in the data set using the optimized structures at the MP2/6-31G(d) level of theory. The structures were fully optimized with quadratically convergent SCF procedure[50] and there were no normal modes with negative frequencies. The correspondence between QM and MM normalized normal modes was determined based on the dot product before further comparison. For this purpose, normal modes were considered as collinear if their

absolute dot product is >0.5. If a QM normal mode is contributed by several CHARMM normal modes, the MM mode with the largest dot product was considered.

To characterize the fit between QM and MM frequencies, Mean Percentage Absolute Error (MPAE) was calculated using:

$$\text{MPAE}= \frac{100\%}{n}\sum_i^n \left| \frac{\alpha\nu_i^{QM} - \nu_j^{MM}}{\alpha\nu_i^{QM}} \right|, \quad [11]$$

where n is number of colinear NM; ν_i^{QM} and ν_j^{MM} are the frequency of QM NM and the frequency of corresponding collinear CHARMM and QM NM; $\alpha = 0.9432$ is the vibrational scaling factor.[51] MPAE was adapted from a previous work.[52] The bonded parameters were optimized with the scaling factor of α^2 applied to QM energies to obtain scaling factor of α for QM frequencies. The comparison was also done with the QM frequencies scaled by this factor in Equation 11. Bonded parameters were derived starting from initial CHARMM parameters for force constants and equilibrium values obtained from *ab initio* optimized geometry of the molecule.

Test molecules and atom types

To test different parametrization methods, 32 molecules were selected with available CHARMM parameters. The set of molecules comprises molecules with diverse chemical structures and includes three, four, five and six-atom ring structures. In total, the set contains 62 CHARMM types that define 74 unique bond terms and 127 unique angle terms with an average of 4 and 6 terms per molecule respectively. Chemical structures of these molecules are shown in Figure 1 and their chemical names and formulae are given in Table S1 in Supporting Information.

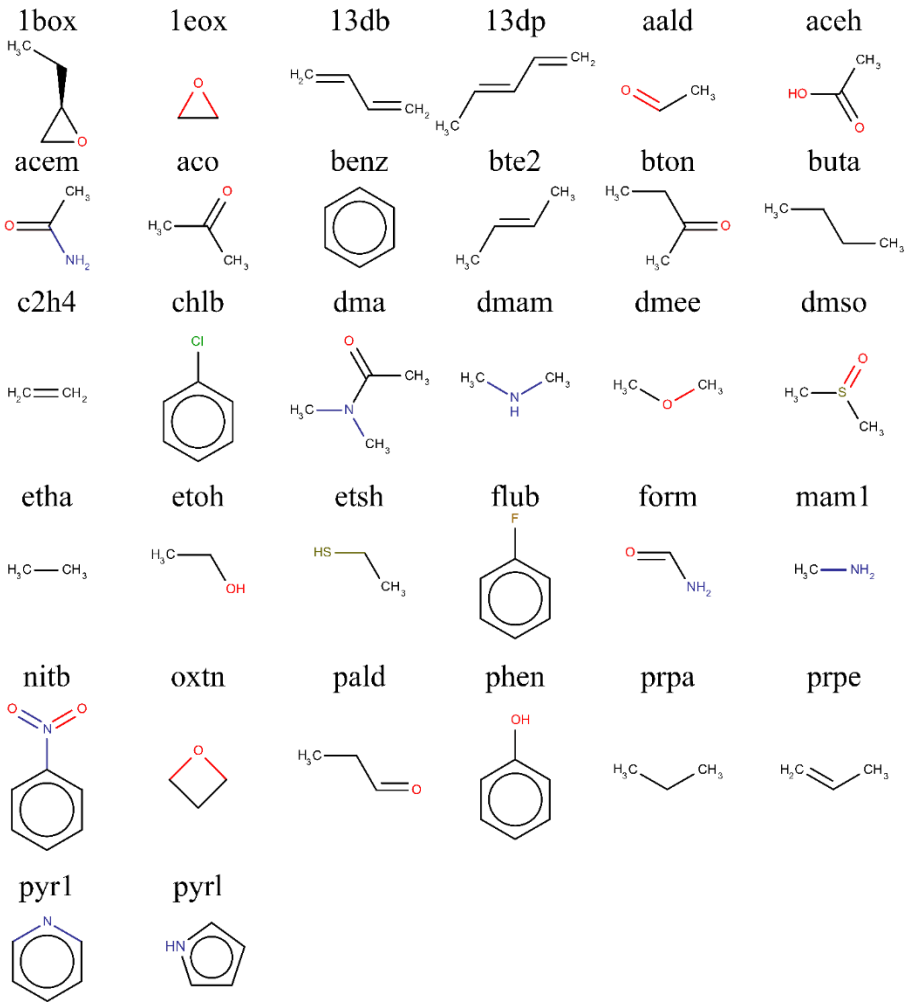


Figure 1. Chemical structures of 32 molecules used in this work. 2D representations were prepared with MarvinSketch.[53]

The force field parameters for the molecules in the test set were taken from CGenFF.[14] For simplicity, these parameters will be further referred as initial CHARMM parameters. The geometry of molecules was generated from the existing tables of internal coordinates in the CGenFF force field files. The geometries were further optimized at the MP2/6-31G(d) model chemistry, or MP2/6-311G(d) model chemistry for anionic molecules, the same theory level used to parametrize the CGenFF force field.[14] Adiabatic potential energy surface (PES) scans with QM were performed along selected degrees of freedom as described above. MM calculations were performed with CHARMM program and QM calculation were performed with Gaussian16.[54]

RESULTS

Deviation between QM and MM relaxed structures leads to suboptimal force constants

In this section, the arguments will be illustrated on selected molecules. Figure 2A shows the PE surface for C-C bond stretching in ethylene computed with QM and optimal CHARMM parameters. The Root Mean Square (RMS) deviation between QM and MM energies computed with Formula 5 is shown in Figure 2 (panels B and C) computed with different force field equilibrium bond distance values, b_0; and with different force constant values,

K_b. For these calculations, the optimal values for the term were used for b_0 and K_b, respectively; the other terms optimized in PES scans were also treated with the optimal parameters. The MM structures were optimized using the same value for the valence coordinate (bond or angle), which was used for the QM optimizations, i.e. $x_{i,k}^{QM} = x_{i,k}^{MM}$. As it can be seen, even relatively small deviations in b_0 lead to large deviations between QM and MM energies, and thus the RMS deviation. For example, a deviation of just 0.05 Å from the optimal value for b_0 leads to a value of 2.6 kcal·mol^{-1} for the RMS deviation (a deviation of 0.1 Å leads to ~8 kcal·mol^{-1}). In contrast, relatively large deviations in K_b produce only a small increase in the RMS deviation. For example, in ethylene, reducing the force constant by 300 kcal·mol^{-1}·Å^{-2} would increase the RMS deviation only by 0.35 kcal·mol^{-1}. Thus, if the RMS deviation is used for the cost function, the optimization would be largely balanced toward a better equilibrium distance b_0 to improve the RMS deviation, while the quality of the force constants could be sacrificed. In practice, this is strongly undesirable since force constants are important to reproduce the molecular flexibility. This can be further demonstrated on the heatmap of the RMS deviation between QM and MM energies computed with different b_0 and K_b on Figure 3A. In the previous work on the parametrization of a large and diverse set of molecules[Ref], a typical value observed for the deviation in equilibrium bond distances was 0.02 Å. With this deviation, as shown in Figure 3A, the optimized value for the force constant is on order of 300 kcal·mol^{-1}·Å^{-2}, which is ~300 kcal·mol^{-1}·Å^{-2} off from the optimal force constant needed to reproduce the C-C bond stretching in ethylene.

We shall consider another example of the valence angle in acetaldehyde. Figure 2 panel A shows the PE surface for O-C-C valence angle bending in acetaldehyde computed with QM and optimal CHARMM parameters; while the RMS deviation between QM energies and MM energies calculated at different equilibrium angle values, θ_0 and force constants, K_a is shown in panels B and C of the same figure. The Figure 3B shows the heatmap of the RMS deviation between QM and MM energies at different θ_0 and K_a. It can be seen that similar to the bond term relatively small deviations in θ_0 increase significantly the difference in QM and MM energies and hence the RMS deviation. For example, the RMS deviation of 0.6 kcal·mol^{-1} corresponds to just less than 3º deviation from the optimal equilibrium valence angle and the same RMSD corresponds to 38 kcal·mol^{-1}·rad^{-2} deviation for the force constant K_a (over 100%).

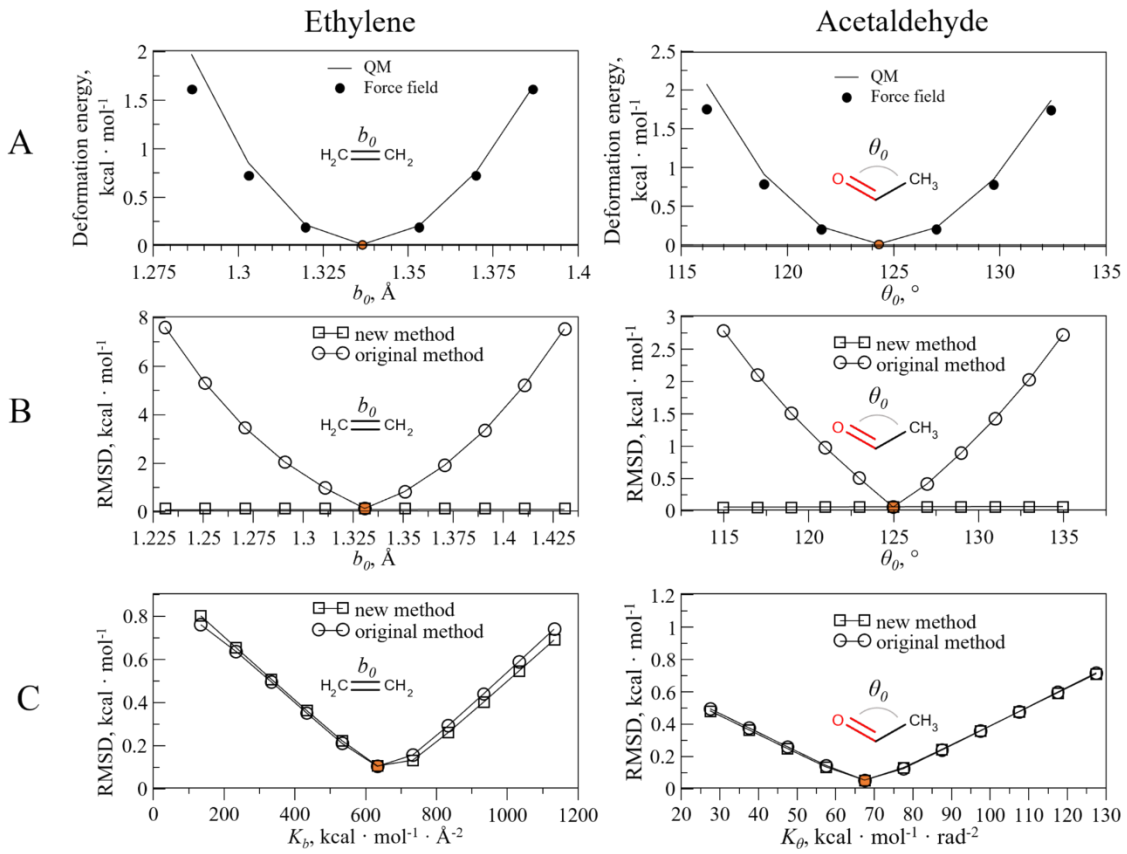


Figure 2. Potential energy surfaces and energy RMS deviations as a function of the force field parameters. (A) Left, the PE surface for the C-C bond stretching in ethylene, and right, for the O-C-C angle bending for acetaldehyde. (B) The RMS deviation between QM and CHARMM energies of structures from the PES scans as a function of the equilibrium bond distance, b_0, and valence angle, θ_0, for ethylene (Left) and acetaldehyde (Right), respectively. (C) RMS deviation between QM and CHARMM energies as a function of the force constant for the bond term in ethylene (left) and the force constant of the angle term in acetaldehyde, right, respectively. The optimal values for the equilibrium bond distance, valence angle and force constants are colored in orange.

The above reasoning was illustrated for PES energies, however, they can be generalized to force matching methods. The average RMS deviation between QM and MM atomic forces in structures from the PES scans, shown in Figure 2A, is shown in Figure 3. Overall, the behavior of the force RMS deviation is very similar to the energy RMS deviation, i.e. small deviations from the optimal equilibrium value for the bond in ethylene and the valence angle in acetaldehyde lead to significantly suboptimal force constants relative to the QM model. We find that deviations in equilibrium values for the bond and angle lead to the same force constants, similar to what was demonstrated using the energy RMS deviation in the previous paragraph: with a deviation of 0.02 Å in the equilibrium value, the optimized value for the force constant is on order of 300 kcal·mol^{-1}·Å^{-2}, which is ~300 kcal·mol^{-1}·Å^{-2} off from the optimal force constant needed to reproduce the C-C bond stretching in ethylene. For the valence angle in acetaldehyde 3º deviation from the optimal equilibrium valence angle results in ~40 kcal·mol^{-1}·rad^{-2} deviation for the force constant K_a.

In both considered cases for the bond and angle terms, structural deviations between the QM and MM

optimized structures along the corresponding bond and angle lead to force constants that are smaller compared to the values optimized in the absence of these deviations. In general, the resulting MM model under a structural inconsistency as demonstrated in Figure 3, is softer than the QM model, and in other words, it allows larger amplitudes of deformations at the same energy values compared to the QM model.

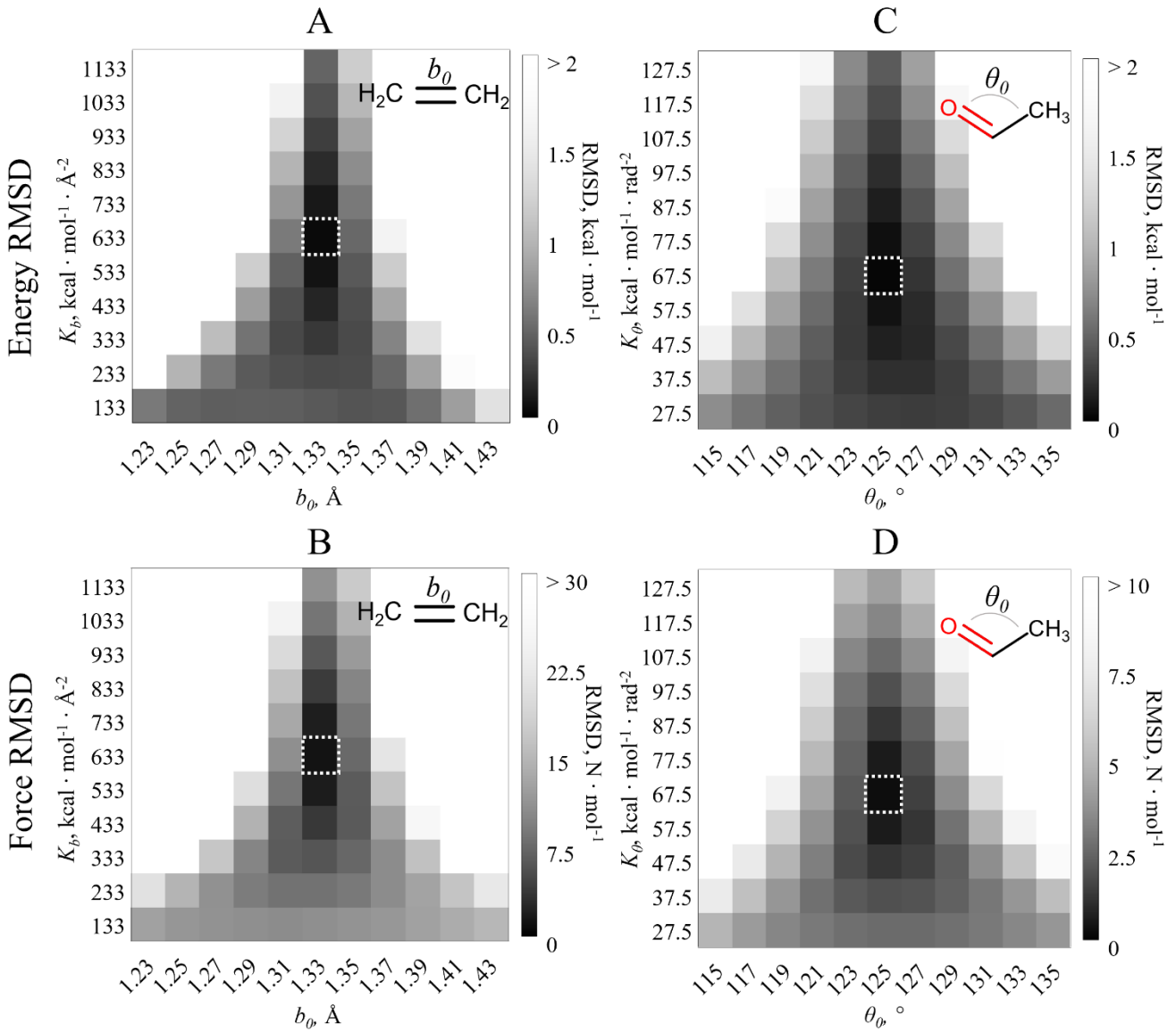


Figure 3. The RMS deviation between QM and CHARMM energies and forces of structures from PES scans as a function of FF parameters. (A) The energy RMS deviation and (B) the force RMS deviation as a function of the equilibrium bond distance, b_0, and force constant, K_b, of the C–C bond term in ethylene. (C) The energy RMS deviation and (D) the force RMS deviation as a function of the equilibrium angle, θ_0, and force constant, K_θ, of the O–C–C angle term in acetaldehyde. The optimal values for the force field parameters are shown in a white dotted square.

To remove the strong dependence on equilibrium parameters, demonstrated above, we allow structural deviations along the scanned degree of freedom, where energy differences are computed between different structures used for QM and MM calculations.

Force constants for angle terms sharing an atomic center

Let's consider deformations along the valence angles involving the ortho hydrogen atom of pyridine shown in Figure 4. In a previous work the conclusion was drawn that defining force constants for the valence angles for this atom is an ill-defined problem.[41] To derive this conclusion the ring structure was assumed to be rigid with all reference angles having 120°. In this case, indeed, the in-plane hydrogen bending can be described by only one valence angle, and energy in Eq. 2 can be expressed as one harmonic term with the force constant given by the sum of the force constants of the two angle terms involving the hydrogen atom ($K_{effective} = K_1 + K_2$). However, this is not the case if minimization is done as the pyridine ring is not rigid. The angles with the atomic center are defined as shown in Figure 4. In practice, with the angle θ_1 between N-C-H bent out of its equilibrium value, the other two angles θ_2 (C-C-N) and θ_3 (C-C-H) would also assume values different from their values in the minimized structures. In particular, the valence angle θ_3 would be different than 120º with in-plane hydrogen bending. Figure 4A shows the dependence of θ_2 and θ_3 on θ_1. For example, with θ_1 bending of 5º, the structure is bent along θ_2 and θ_3 by around 1º and 4º, respectively.

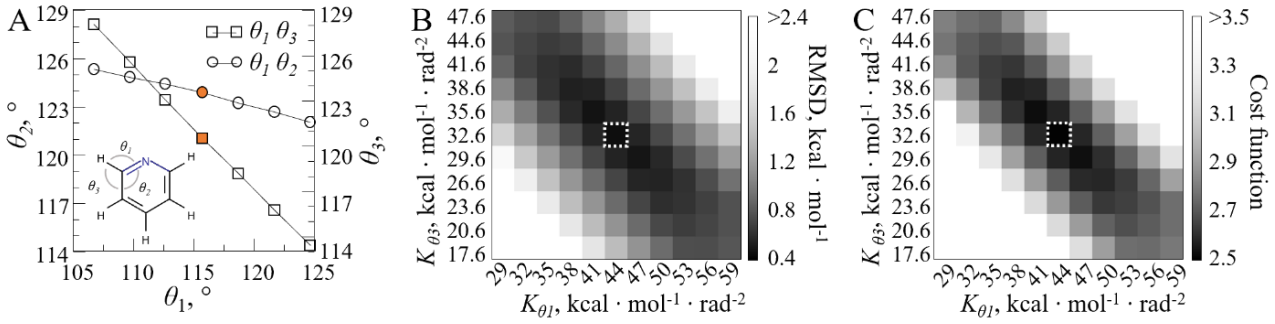


Figure 4. Angle bending involving ortho-hydrogen in pyridine. (A) Angles involving ortho-hydrogen in pyridine in the PES scan along θ_1 angle between N–C–H. Values corresponding to the minimum energy structure are in orange. (B) RMS deviation between QM and CHARMM energies from the PES scan as a function of the force constants of the two angle terms defined for ortho-hydrogen in pyridine (θ_1 and θ_3) (C) The cost function that includes the new term proposed in this work as a function of the two force constants K_1 and K_3. The optimal values for the force constants are marked by dotted line squares.

At a particular bending along θ_1 in the MM optimized structure, the other two angles depend on the force constants of the corresponding angle terms K_{θ_2} and K_{θ_3}, which should be sufficient to define these force constants. This conjecture we will test numerically in the next section. To improve further the distribution of force constants of angle terms, we tested an additional term to the cost function. We note that at a particular value of the angle, θ_1, the position of the atom N, and thus angles θ_2 and θ_3 depend on the corresponding force constants in the MM model. Thus, including the deformations along the adjacent angles, in principle, is expected to improve the distribution of the force constant. Figure 4B shows the RMS deviation between QM and MM energies as a function of the force constants of the two angle terms defined for ortho-hydrogen in pyridine ($\theta 1$ and $\theta 3$). As expected, the RMS deviation has the minimum very shallow along the sum of the two force constants, $K_{effective} = K_1 + K_2$ (diagonal on Figure 4B), making it difficult to define K_1 and K_2. By contrast, the cost function that includes the

RMS deviation as well as the new deformation based term, given by Eq. 10, shown in Figure 4C demonstrates a better defined minimum, which allows to resolve the force constants K_1 and K_2.

Stability relative to the initial parameters

Since we employ optimization as a method to derive force field parameters, in principle, different sets of bonded parameters can be obtained starting from different initial values. To characterize the stability of optimized parameters with the new method we performed a numerical test where optimization was initiated from different initial force constants. Force constants, K_b for bond terms were assigned randomly from a wide range of values between 100 and 800 kcal·mol^{-1}·Å^{-2}; for angle terms, force constants, K_a were randomly assigned in the range of 10 to 100 kcal·mol^{-1}·rad^{-2}. The optimization was repeated five times starting from different random force constants; initial equilibrium bond lengths and angles were taken from QM structures. Relative and absolute standard deviation (SD) for bond and angle terms averaged over terms in individual molecules from the data set are shown in Figure 5 and SD averaged over all molecules in the data set, is given in Table 1.

Table 1. Standard deviation for bonded term parameters averaged over 32 molecules in the data set.

	Force constants, K_a or K_b		Equilibrium parameters, b_0 or θ_0	
Optimized terms	Average RSD	SD	Average RSD	SD
Bonds	0.2 (0.5)%	0.86 (1.94)	0.004 (0.01)%	0.00001 (0.00018)
Angles without the new term[a]	1.20 (2.8)%	0.69 (1.91)	0.06 (0.14)%	0.07 (0.17)
Angles with the new term[b]	0.58 (1.4)%	0.36 (0.88)	0.02 (0.31)%	0.02 (0.36)

[a,b]For angles, force field parametrization was done without and with the additional deformation-based term given by Equation 9 included to the cost function, respectively; the standard deviations for the computed values are shown in parenthesis. K_b and K_a are in kcal·mol^{-1}·Å^{-2} and kcal·mol^{-1}·rad^{-2}, respectively; the equilibrium parameters, b_0 or θ_0, are in Å and °, respectively.

The results demonstrate that the new method produces practically the same force constants starting from very different initial parameters. The relative SD for the force constants for bond terms, K_b averaged over all bond terms (74 total) in 32 molecules is just 0.2%. For angle terms, the relative SD for K_a averaged over a total of 127 angle terms in 32 molecules is also very small, 1.2%. The later value can be further improved to 0.58% by introducing the deformation-based term given by Formula 9 to the cost function. Though, this improvement is small on average, for some angle terms it can present a significant improvement, for example for oxetane and 1,3-dibutene it improves the relative SD from 14% to 7%, and 7% to ~0.5%, respectively. Since PES energies are approximately proportional to force constants, the maximum deviation of 2.5% and 7% observed for bond and angle terms, respectively, corresponds to the energy deviation of just 0.05 and 0.14 kcal·mol^{-1} at the maximum energy deformation considered in this work (2 kcal·mol^{-1}), respectively. The equilibrium parameters deviate only insignificantly in five optimization tests. The relative SD is 0.004% and 0.02% for the equilibrium bond length and angle, respectively. Overall, bond and angle parameters derived by the new method are very stable with respect to the initial values used for the optimization.

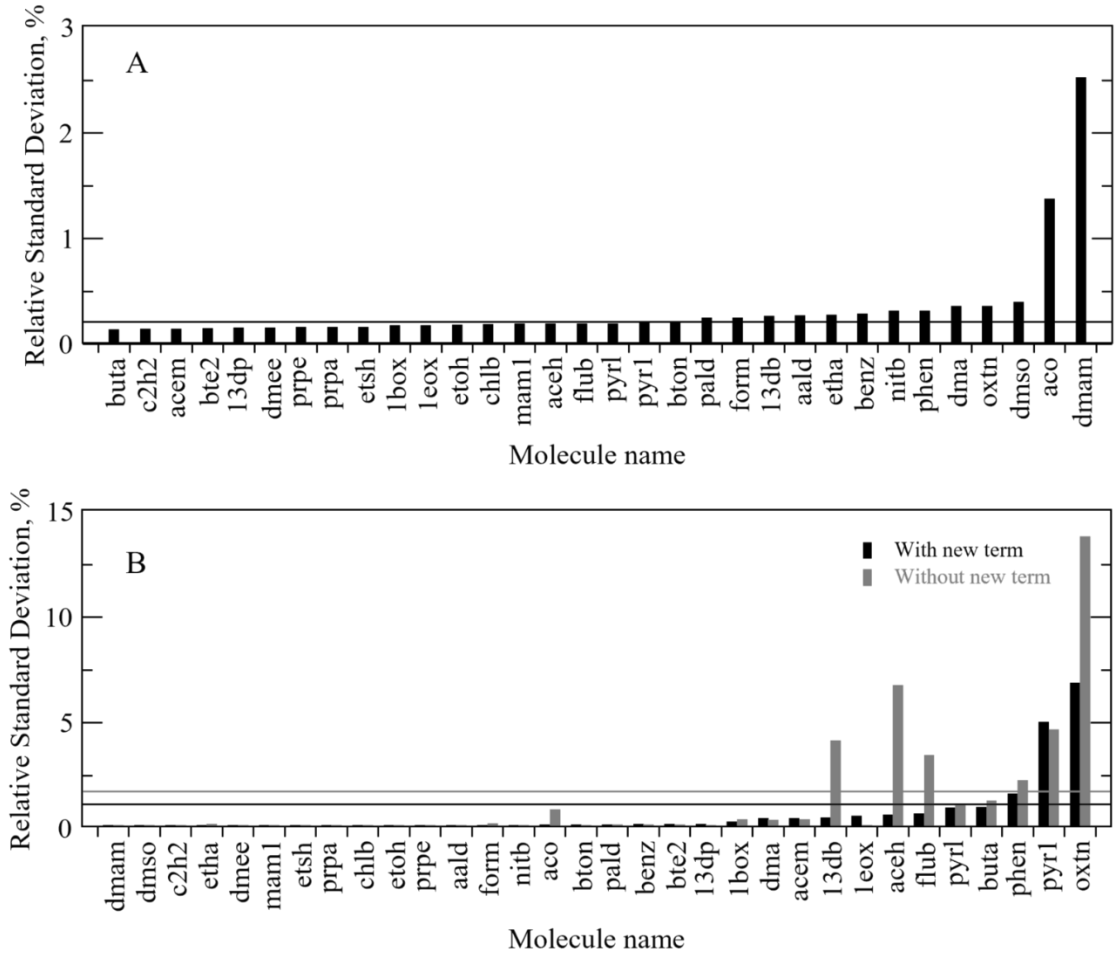


Figure 5. Robustness of optimized FF force constants relative to the initial parameters. The relative standard deviation of force constants for (A) bond and (B) angle terms is shown. Five optimizations started with random force constants were performed to compute the relative SD, which was averaged over bonded terms in individual molecules. The relative SD averaged over all molecules is shown as a horizontal line.

Transferability of optimized parameters

In the previous section we showed that the optimized parameters are robust and do not depend on initial parameters. In this section we will demonstrate that the optimized bonded parameters for bond and angle terms using the new method are transferable. For this, the comparison of the same bonded terms optimized in different molecules is performed. Indeed, 23 bond parameters and 24 angle parameters are shared by 2 to 17 molecules in the data set. For example, the bond term defined between atom types CG331-HGA3, i.e. between methyl hydrogen and carbon, is shared by 17 molecules, while the angle term, CG2R61- CG2R61-CG2R61, i.e. defined for the angle between carbon atoms in an aromatic ring is shared by 6 molecules. For this test, bond and angle terms were optimized using the new method; for angles, the optimization was done with and without the deformation-based term given by Equation 9. Bonds or angles in two or more molecules were identified and the SD and relative SD for the same bonded parameters optimized in different molecules were computed. The SD for bond and angle terms are shown in Figure 6 and given in Table S3 for individual terms. The following analysis does not include

the angle term in two molecules having the three atom ring (1box and 1eox molecules), since by contrast to other molecules, a bending along the angle and stretching along the opposite bond in the three atom ring have two contributions: from the opposite bond and the contribution from the angle, and the bond term can be considered as an Urey-Bradley term for the angle.

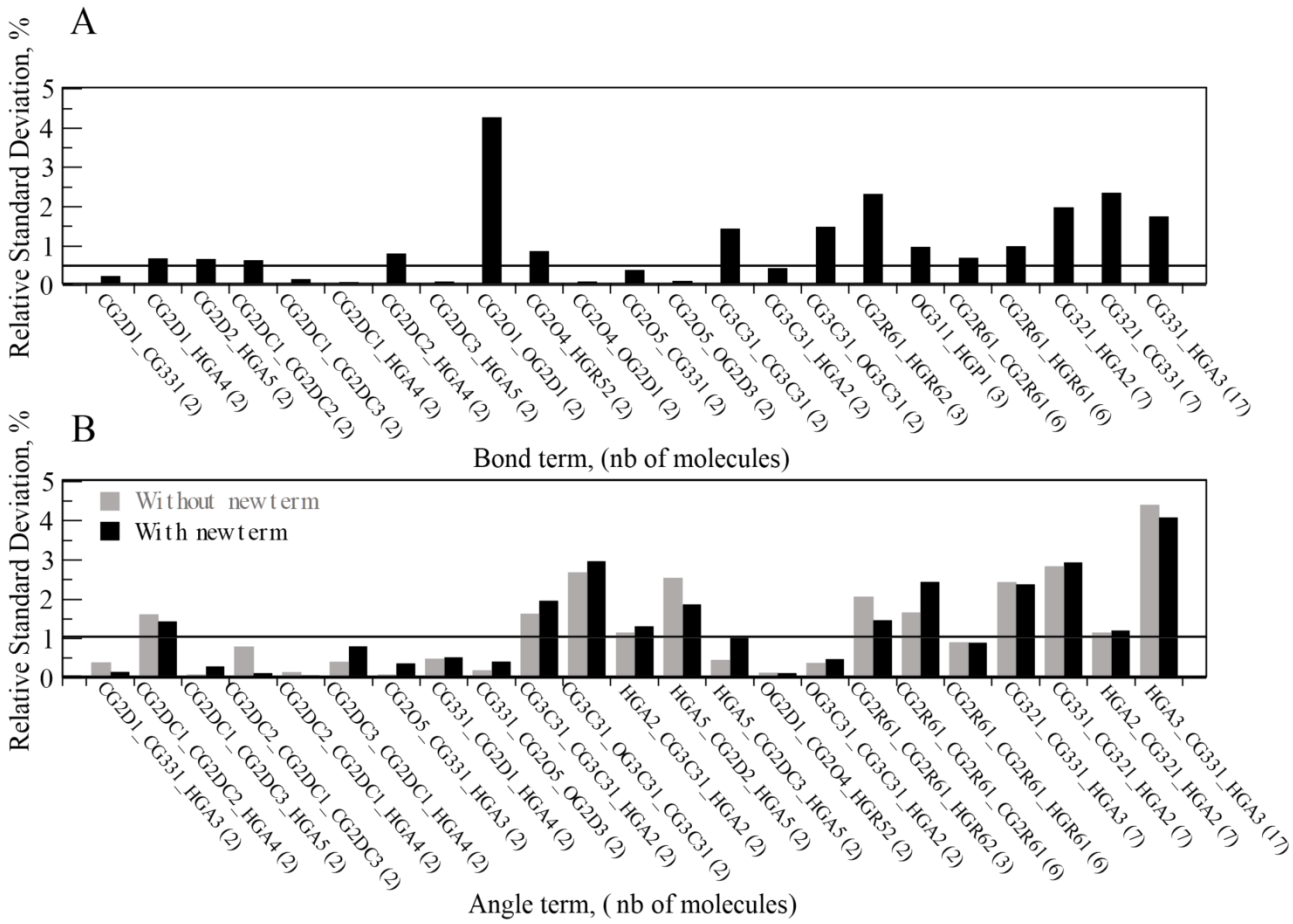


Figure 6. Transferability of optimized FF force constants relative to the initial parameters. The relative standard deviation is shown for bond (A) and angle (B) force constants optimized in different molecules. On panel (B) black and gray bars correspond to the results with and without the new angle term given by Equation 9, respectively. The horizontal lines show the average relative SD. The number of molecules sharing the term is given in parenthesis.

The SD averaged over all bond terms for the force constant, K_b is just 4.3 kcal·mol^{-1}·Å^{-2}, with the maximum relative value of 4.5% for the bond term defined for atom types CG2O1-OG2D1. This bond term is found in two molecules, formamide and dimethylacetamide, where a hydrogen atom on the oxygen-adjacent carbon in the former molecule is replaced with a methyl group, suggesting that this bond may require different force constants in the two molecules, which can be provided by introducing additional atom types. All bond terms except this bond have the relative SD less than 3%. The energy difference due to these deviations (at the maximum energy deformation of 2.0 kcal·mol^{-1}) is lower than 0.1 kcal·mol^{-1}. The SD for the equilibrium bond length averaged over all bond terms is just 0.0016 Å. For angle terms, the SD for force constants K_a optimized in different molecules is again very small 0.6 kcal·mol^{-1}·rad^{-2} (the relative SD is 2%). The largest relative standard deviation

of 4% for K_a was observed for the valence angle H-C-H in the methyl group (atom types HGA3-CG331-HGA3). However, the optimized force constant for the valence angle term involving hydrogens of the methyl group is 37.4±2.0 kcal·mol^{-1}·rad^{-2}, very small in comparison to other angle terms, and other energy terms (interactions with other parts of the molecule) can have comparable contributions to bending along H-C-H, making it difficult to fit K_a. After optimization in different molecules, the equilibrium angles θ_0 are practically identical with SD of 0.8º. Since the force constants and equilibrium parameters deviate insignificantly, the force field parameters optimized by the new method are largely transferable, i.e. can be optimized in one molecule and used for the other molecules, at least as demonstrated for the molecules in the data set.

Normal modes

We further tested the quality of the optimized parameters using normal mode analysis. We note that, typically, for comparison, QM and MM normal modes are sorted based on the magnitude of frequencies.[18,52] However, as shown in table S4, around ~50% of normal modes do not match , i.e. pairs of QM and MM normal modes with the absolute dot product (d) between normalized NMs lower than $d < 0.5$, if they are sorted based on the magnitude of frequencies. Figure S1 shows dot product between QM and MM normalized normal modes sorted based on the frequency magnitude for benzene, ethylene, dimethylsulfoxide and butanol. As it can be seen, many corresponding MM normal modes are not in the same order as QM normal modes, and also QM normal modes may have several contributions from MM normal modes. Thus, if QM and MM modes are sorted only based on their frequencies, one may compare QM and MM normal modes which can be even orthogonal. To solve this problem, in this work we establish the correspondence between QM and MM normalized normal modes based on the dot product before the comparison as described in the methods section. With matching normal modes based on the dot product, 100% of QM and MM normal modes have $d > 0.5$, and ~90% have $d > 0.75$ as indicated in table S4. However, if normal modes are sorted based on their frequencies, only ~51% of pairs of QM and MM normal modes have $d > 0.5$, and ~45% have $d > 0.75$.

To characterize the fit between QM and MM frequencies, Mean Percentage Absolute Error was calculated, after matching QM and MM normal modes based on the dot product. Figure 7 demonstrates the improvement of normal modes relative to QM NM for each individual molecule. Table 2 gives the mean MPAE for the 32-molecule set. We note that the initial CHARMM parameters were derived to reproduce normal modes and thus are expected to give very good results relative to the QM normal modes. A mean MPAE is 9.5% with the initial CHARMM parameters, while with the bond term parameters optimized using the term given by Equation 9, the mean error is lower 8.3%. With both bond and angle terms optimized the MPAE is getting even lower to 6.8%. It should be noted that without matching normal modes based on the dot product, the MPAE is consistently lower with an average of 6.9% with initial CHARMM parameters and with 5.4% for optimized bond and angle terms. However, the optimized parameters still give better results than with the initial parameters.

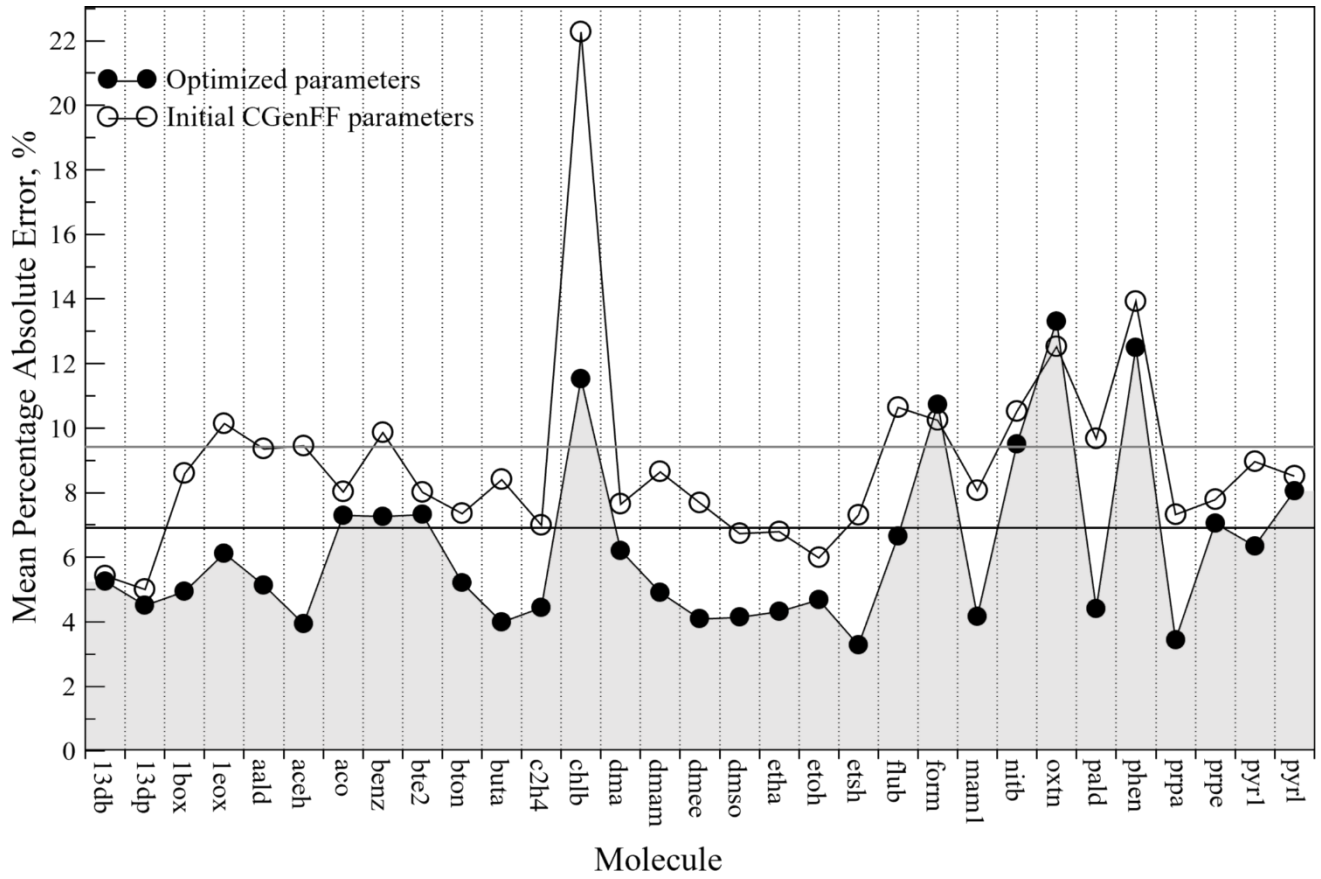


Figure 7. Mean Percentage Absolute Error between QM and MM normal modes computed with initial and optimized bonded term parameters. QM and MM normal modes were matched based on the dot product. Average values are represented as horizontal lines. The data are shown for all molecules in the data set, except acetamide (acem) due to its large value, 26.19% and 25.73% with the initial and optimized parameters, respectively.

The largest MPA error of 22.2% was computed for chlorobenzene with the initial parameters, which was improved to 11.5% with the optimized parameters. The high MPAE between QM and MM normal modes for chlorobenzene with the initial CHARMM parameters is mainly due to two bond terms, as demonstrated by PES scans. The RMS deviation for the PES scan along the C-C bond is 0.28 kcal·mol^{-1} and 0.10 kcal·mol^{-1} with the initial and optimized force constants, respectively. The RMS deviation for the PES scan along the C-Cl bond is 0.43 kcal·mol^{-1} and 0.13 kcal·mol^{-1} with the initial and optimized force constants, respectively. Thus, in in the initial parameters, the C-C bond term in chlorobenzene has the force constant 30% smaller than needed to reproduce the corresponding PE surface, and the C-Cl bond term has a force constant 40% stiffer compared to the optimized value.

Table 2. Mean Percentage Absolute Error between QM and MM normal mode frequencies averaged over 32 molecules in the test set. The pairs of QM and MM NMs were matched based on the dot product between normalized QM and MM NMs.

	[a]initial	[b]bond terms	[c]angle terms	[d]both bond/angle terms
MPAE[a] %	9.46	8.30	8.04	6.84

[a]MM NMs were obtained with the initial parameters, [b]with the bond term parameters optimized, [c]with the angle

term parameters optimized, and with both the bond and angle parameters optimized;

Figure 8 compares initial and optimized force constants. For bond terms, as expected, the optimized parameters are in good agreement with the standard CHARMM force constants with the linear correlation coefficient of 90% and RMS deviation of 68.4 kcal·mol^{-1}·Å^{-2} (18%). This is due to the fact that contributions of bond terms to normal modes are typically well separated from contributions from other degrees of freedom and normally account for high-frequency vibrations. By contrast, for angles the correlation is smaller 69% and the RMS deviation is 14.7 kcal·mol^{-1}·rad^{-2} (31%), which is agreement with the conclusion that angle term parameters may lack robustness as discussed above as well as in ref. 40.

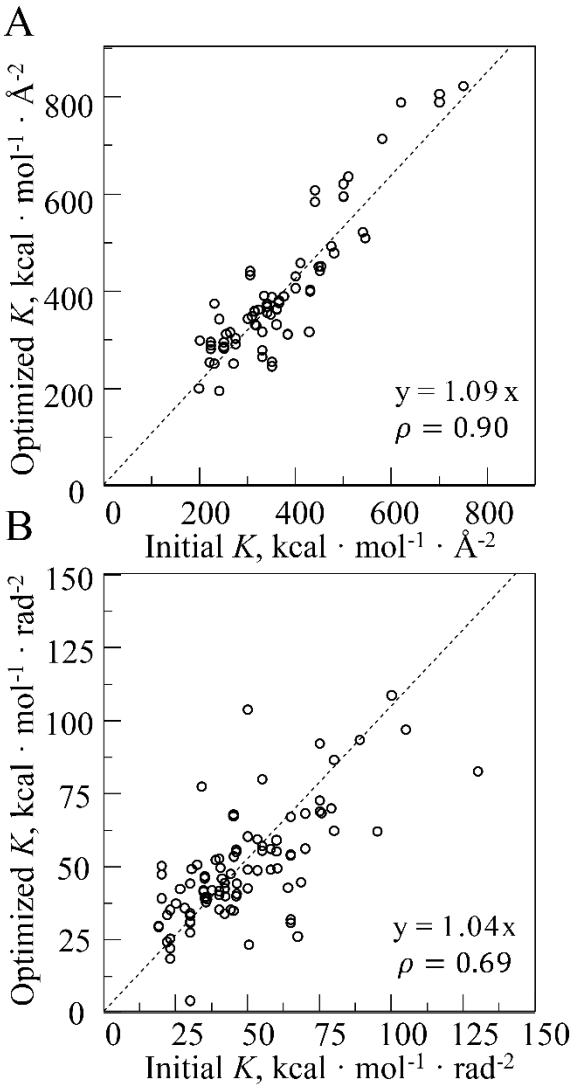


Figure 8. Initial versus optimized force constants for (A) bond and (B) angle terms. Force constants are shown for a total of 72 bond terms and 115 angle terms. The linear fit is shown by a dashed line.

Conclusion

In the present study we demonstrate that optimization of force field parameters based on comparing QM and MM energies and/or forces of structures leads to suboptimal force constants for bond and angle terms, if structural deviations between QM and MM equilibrium structures are present. The presented results show that conventional parametrization methods based on fitting QM energies and forces are largely balanced toward accuracies in force field equilibrium bond lengths and angles, while the accuracy in force constants is sacrificed. With the structural deviation present, the optimized force constants cannot adequately describe the QM flexibility of the molecule, as exemplified on several test molecules. Structural deviations always lead to force constants smaller in comparison to those in the absence of such structural deviations, and to a softer MM model relative to the QM model.

To solve this problem, we developed and implemented a new method to derive force field parameters for bond and angle terms. The new method derives force field parameters based on PES scans where a structural deviation between QM and MM structures is allowed. We tested the method on a set of 32 molecules, and the results show that the optimized force field parameters are robust relative to random initial force constants. Starting from five sets of random force constants, we obtained relative SDs of just 0.3% and 1.2% for the bond and angle force constants, respectively. FF parameters derived by the new method are largely transferable, as demonstrated by the low relative SD (< 2%) for equilibrium bond and angle values and force constants for the same terms optimized in different molecules. We further tested the method to reproduce QM normal modes. The results indicate that there is only 45% correspondence between MM and QM normal modes, if they are sorted based on the frequency magnitude for the comparison, underlying the importance of establishing the correct correspondence based on the dot product. Furthermore, without correctly matching QM and MM normal modes, the agreement for normal modes defined by MPAE may appear better (6.9%) than after matching normal modes based on the dot product (9.5%).

Overall, the new method will allow to parametrize molecules with structural deviations present between QM and MM equilibrium structures, common for force fields for small molecules, producing robust and transferable parameters. In future, the method will be extended to derive parameters for dihedral angle terms.

Data and Software Availability

The data for the transferability and stability tests, and the normal modes computed with QM and CHARMM force field are freely accessible and available for download from: https://github.com/Alexey-AleksandrovCNRS/FF-Terms-under-Structural-Inconsistency.

Supporting Information

Figures comparing QM and MM normal modes; tables with molecules in the test data set; weights for the cost function; average values and standard deviations for bond and angle term parameters optimized in different molecules.

Acknowledgements

This work was supported by grants ANR-18-CE44-0002 to AA. Some calculations were performed using HPC resources from GENCI-CINES (Grant 2018-A0040710436).

Conflict of Interest

No conflict of interest to declare.

References

(1) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell Jr, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain X1 and X2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8* (9), 3257–3273. DOI: 10.1021/ct300400x

(2) MacKerell Jr, A. D.; Feig, M.; Brooks, C. L. Improved Treatment of the Protein Backbone in Empirical Force Fields. *J. Am. Chem. Soc.* **2004**, *126* (3), 698–699. DOI: 10.1021/ja036959e

(3) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586–3616. DOI: 10.1021/jp973084f

(4) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* **2008**, *29* (4), 622–655. DOI: 10.1002/jcc.20820

(5) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7* (9), 2886–2902. DOI: 10.1021/ct200162x

(6) Dickson, C. J.; Madej, B. D.; Skjevik, Å. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C. Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **2014**, *10* (2), 865–879. DOI: 10.1021/ct4010307

(7) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. DOI: 10.1021/acs.jctc.5b00255

(8) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpí, J. L.; González, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M. Parmbsc1: A Refined Force Field for DNA Simulations. *Nat. Methods* **2016**, *13* (1), 55–58. DOI: 10.1038/nmeth.3658

(9) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and Testing of the GROMOS Force-Field Versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40* (7), 843. DOI: 10.1007/s00249-011-0700-9

(10) Robertson, M. J.; Qian, Y.; Robinson, M. C.; Tirado-Rives, J.; Jorgensen, W. L. Development and Testing

of the OPLS-AA/M Force Field for RNA. *J. Chem. Theory Comput.* **2019**, *15* (4), 2734–2742. DOI: 10.1021/acs.jctc.9b00054
(11) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* **2015**, *11* (7), 3499–3509. DOI: 10.1021/acs.jctc.5b00356
(12) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12* (1), 281–296. DOI: 10.1021/acs.jctc.5b00864
(13) Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**, *52* (12), 3144–3154. DOI: 10.1021/ci300363c
(14) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell Jr, A. D. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31* (4), 671–690. DOI: 10.1002/jcc.21367
(15) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52* (12), 3155–3168. DOI: 10.1021/ci3003649
(16) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174. DOI: 10.1002/jcc.20035
(17) Horta, B. A. C.; Merz, P. T.; Fuchs, P. F. J.; Dolenc, J.; Riniker, S.; Hünenberger, P. H. A GROMOS-Compatible Force Field for Small Organic Molecules in the Condensed Phase: The 2016H66 Parameter Set. *J. Chem. Theory Comput.* **2016**, *12* (8), 3825–3850. DOI: 10.1021/acs.jctc.6b00187
(18) Qiu, Y.; Smith, D.; Boothroyd, S.; Jang, H.; Wagner, J.; Bannan, C. C.; Gokey, T.; Lim, V. T.; Stern, C.; Rizzi, A.; Lucas, X.; Tjanaka, B.; Shirts, M. R.; Gilson, M.; Chodera, J.; Bayly, C. I.; Mobley, D.; Wang, L.-P. Development and Benchmarking of Open Force Field v1.0.0, the Parsley Small Molecule Force Field. **2021**. DOI: 10.26434/chemrxiv-2021-l070l-v4
(19) Aleksandrov, A. A Molecular Mechanics Model for Flavins. *J. Comput. Chem.* **2019**, *40* (32), 2834–2842. DOI: 10.1002/jcc.26061
(20) Croitoru, A.; Park, S.-J.; Kumar, A.; Lee, J.; Im, W.; MacKerell Jr, A. D.; Aleksandrov, A. Additive CHARMM36 Force Field for Nonstandard Amino Acids. *J. Chem. Theory Comput.* **2021**, *17* (6), 3554–3570. DOI: 10.1021/acs.jctc.1c00254
(21) van der Spoel, D. Systematic Design of Biomolecular Force Fields. *Curr. Opin. Struct. Biol.* **2021**, *67*, 18–24. DOI: 10.1016/j.sbi.2020.08.006
(22) Lin, F.-Y.; MacKerell Jr, A. D. Force Fields for Small Molecules. *Biomol. Simul. Methods Protoc.* **2019**, 21–54. DOI: 10.1007/978-1-4939-9608-7_2
(23) Vanommeslaeghe, K.; Guvench, O.; MacKerell Jr, A. D. Molecular Mechanics. *Curr. Pharm. Des.* **2014**, *20* (20), 3281–3292.
(24) Dodda, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen Web Server: An Automatic OPLS-AA Parameter Generator for Organic Ligands. *Nucleic Acids Res.* **2017**, *45* (W1), W331–W336. DOI: 10.1093/nar/gkx312
(25) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25* (2), 247–260. DOI: 10.1016/j.jmgm.2005.12.005
(26) Li, P.; Merz, K. M. MCPB.Py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.* **2016**, *56* (4), 599–604. DOI: 10.1021/acs.jcim.5b00674
(27) Huang, L.; Roux, B. Automated Force Field Parameterization for Nonpolarizable and Polarizable Atomic Models Based on Ab Initio Target Data. *J. Chem. Theory Comput.* **2013**, *9* (8), 3543–3556. DOI: 10.1021/ct4003477
(28) Guvench, O.; MacKerell Jr, A. D. Automated Conformational Energy Fitting for Force-Field Development. *J. Mol. Model.* **2008**, *14* (8), 667–679. DOI: 10.1007/s00894-008-0305-0
(29) Waldher, B.; Kuta, J.; Chen, S.; Henson, N.; Clark, A. E. ForceFit: A Code to Fit Classical Force Fields to

Quantum Mechanical Potential Energy Surfaces. *J. Comput. Chem.* **2010**, *31* (12), 2307–2316. DOI: 10.1002/jcc.21523
(30) Betz, R. M.; Walker, R. C. Paramfit: Automated Optimization of Force Field Parameters for Molecular Dynamics Simulations. *J. Comput. Chem.* **2015**, *36* (2), 79–87. DOI: 10.1002/jcc.23775
(31) Kumar, A.; Yoluk, O.; MacKerell Jr, A. D. FFParam: Standalone Package for CHARMM Additive and Drude Polarizable Force Field Parametrization of Small Molecules. *J. Comput. Chem.* **2020**, *41* (9), 958–970. DOI: 10.1002/jcc.26138
(32) Morado, J.; Mortenson, P. N.; Verdonk, M. L.; Ward, R. A.; Essex, J. W.; Skylaris, C.-K. ParaMol: A Package for Automatic Parameterization of Molecular Mechanics Force Fields. *J. Chem. Inf. Model.* **2021**, *61* (4), 2026–2047. DOI: 10.1021/acs.jcim.0c01444
(33) Hudson, P. S.; Boresch, S.; Rogers, D. M.; Woodcock, H. L. Accelerating QM/MM Free Energy Computations via Intramolecular Force Matching. *J. Chem. Theory Comput.* **2018**, *14* (12), 6327–6335. DOI: 10.1021/acs.jctc.8b00517
(34) Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Von Bargen, C. D.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J. Chem. Theory Comput.* **2021**, *17* (7), 4291–4300. DOI: 10.1021/acs.jctc.1c00302
(35) Seminario, J. M. Calculation of Intramolecular Force Fields from Second-Derivative Tensors. *Int. J. Quantum Chem.* **1996**, *60* (7), 1271–1277. DOI: 10.1002/(SICI)1097-461X(1996)60:7<1271::AID-QUA8>3.0.CO;2-W
(36) Wang, L.-P.; Chen, J.; Van Voorhis, T. Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data. *J. Chem. Theory Comput.* **2013**, *9* (1), 452–460. DOI: 10.1021/ct300826t
(37) Hudson, P. S.; Han, K.; Woodcock, H. L.; Brooks, B. R. Force Matching as a Stepping Stone to QM/MM CB[8] Host/Guest Binding Free Energies: A SAMPL6 Cautionary Tale. *J. Comput. Aided Mol. Des.* **2018**, *32* (10), 983–999. DOI: 10.1007/s10822-018-0165-3
(38) Akin-Ojo, O.; Song, Y.; Wang, F. Developing Ab Initio Quality Force Fields from Condensed Phase Quantum-Mechanics/Molecular-Mechanics Calculations through the Adaptive Force Matching Method. *J. Chem. Phys.* **2008**, *129* (6), 064108. DOI: 10.1063/1.2965882
(39) Wang, L.-P.; McKiernan, K. A.; Gomes, J.; Beauchamp, K. A.; Head-Gordon, T.; Rice, J. E.; Swope, W. C.; Martínez, T. J.; Pande, V. S. Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J. Phys. Chem. B* **2017**, *121* (16), 4023–4039. DOI: 10.1021/acs.jpcb.7b02320
(40) Pulay, P.; Fogarasi, G.; Pang, F.; Boggs, J. E. Systematic Ab Initio Gradient Calculation of Molecular Geometries, Force Constants, and Dipole Moment Derivatives. *J. Am. Chem. Soc.* **1979**, *101* (10), 2550–2560. DOI: 10.1021/ja00504a009
(41) Vanommeslaeghe, K.; Yang, M.; MacKerell Jr, A. D. Robustness in the Fitting of Molecular Mechanics Parameters. *J. Comput. Chem.* **2015**, *36* (14), 1083–1101. DOI: 10.1002/jcc.23897
(42) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38. DOI: 10.1016/0263-7855(96)00018-5
(43) Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. Rapid Parameterization of Small Molecules Using the Force Field Toolkit. *J. Comput. Chem.* **2013**, *34* (32), 2757–2770. DOI: 10.1002/jcc.23422
(44) Hopkins, C. W.; Roitberg, A. E. Fitting of Dihedral Terms in Classical Force Fields as an Analytic Linear Least-Squares Problem. *J. Chem. Inf. Model.* **2014**, *54* (7), 1978–1986. DOI: 10.1021/ci500112w
(45) Urwin, D. J.; Alexandrova, A. N. Regularization of Least Squares Problems in CHARMM Parameter Optimization by Truncated Singular Value Decompositions. *J. Chem. Phys.* **2021**, *154* (18), 184101. DOI: 10.1063/5.0045982
(46) Vanommeslaeghe, K.; MacKerell Jr, A. D. CHARMM Additive and Polarizable Force Fields for Biophysics and Computer-Aided Drug Design. *Biochim. Biophys. Acta* **2015**, *1850* (5), 861–871. DOI: 10.1016/j.bbagen.2014.08.004
(47) Xu, Y.; Vanommeslaeghe, K.; Aleksandrov, A.; MacKerell Jr, A. D.; Nilsson, L. Additive CHARMM Force Field for Naturally Occurring Modified Ribonucleotides. *J. Comput. Chem.* **2016**, *37* (10), 896–912. DOI: 10.1002/jcc.24307

(48) Press, W. H.; Teukolsky, S.; Vetterling, W.; Flannery, B. *Numerical Recipes: The Art of Scientific Computing*; , 3rd ed.Cambridge University Press: Cambridge, UK ; New York, 2007.
(49) Brooks, B. R.; Brooks, C. L.; MacKerell Jr, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30* (10), 1545–1614. DOI: 10.1002/jcc.21287
(50) Bacskay, G. B. A Quadratically Convergent Hartree—Fock (QC-SCF) Method. Application to Closed Shell Systems. *Chem. Phys.* **1981**, *61* (3), 385–404. DOI: 10.1016/0301-0104(81)85156-7
(51) Johnson, R. D. *NIST Computational Chemistry Comparison and Benchmark Database*. http://cccbdb.nist.gov/ (accessed 2022-01-15).
(52) Allen, A. E. A.; Payne, M. C.; Cole, D. J. Harmonic Force Constants for Molecular Mechanics Force Fields via Hessian Matrix Projection. *J. Chem. Theory Comput.* **2018**, *14* (1), 274–281. DOI: 10.1021/acs.jctc.7b00785
(53) MarvinSketch, 2019.
(54) M. J. Frisch,; G. W. Trucks,; H. B. Schlegel,; G. E. Scuseria,; M. A. Robb; J. R. Cheeseman; G. Scalmani; V. Barone; G. A. Petersson; H. Nakatsuji; X. Li, M. Caricato; A. Marenich; J. Bloino; B. G. Janesko; R. Gomperts; B. Mennucci; H. P. Hratchian; J. V. Ortiz; A. F. Izmaylov; J. L. Sonnenberg; D. Williams-Young; F. Ding; F. Lipparini; F. Egidi; J. Goings; B. Peng; A. Petrone; T. Henderson; D. Ranasinghe; V. G. Zakrzewski; J. Gao; N. Rega; G. Zheng; W. Liang; M. Hada; M. Ehara; K. Toyota; R. Fukuda; J. Hasegawa; M. Ishida; T. Nakajima; Y. Honda; O. Kitao; H. Nakai; T. Vreven; K. Throssell; J. A. Montgomery, Jr.; J. E. Peralta; F. Ogliaro; M. Bearpark; J. J. Heyd; E. Brothers; K. N. Kudin; V. N. Staroverov; T. Keith; R. Kobayashi; J. Normand; K. Raghavachari; A. Rendell; J. C. Burant; S. S. Iyengar; J. Tomasi; M. Cossi; J. M. Millam; M. Klene; C. Adamo; R. Cammi; J. W. Ochterski; R. L. Martin; K. Morokuma; O. Farkas; J. B. Foresman; D. J. Fox. Gaussian, 2009.

Graphical abstract