JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

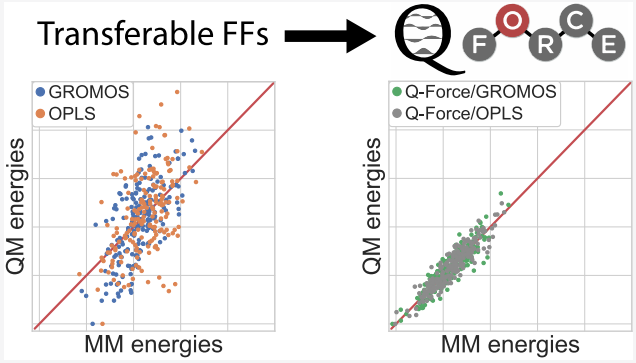
Q-Force: Quantum Mechanically Augmented Molecular Force Fields

Selim Sami,* Maximilian F.S.J Menger, Shirin Faraji, Ria Broer, and Remco W. A. Havenith

Cite This: J. Chem. Theory Comput. 2021, 17, 4946−4960 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The quality of molecular dynamics simulations strongly depends on the accuracy of the underlying force fields (FFs) that determine all intra- and intermolecular interactions of the system. Commonly, transferable FF parameters are determined based on a representative set of small molecules. However, such an approach sacrifices accuracy in favor of generality. In this work, an open-source and automated toolkit named Q-Force is presented, which augments these transferable FFs with molecule-specific bonded parameters and atomic charges that are derived from quantum mechanical (QM) calculations. The molecular fragmentation procedure allows treatment of large molecules (>200 atoms) with a low computational cost. The generated Q-Force FFs can be used at the same computational cost as transferable FFs, but with improved accuracy: We demonstrate this for the vibrational properties on a set of small molecules and for the potential energy surface on a complex molecule (186 atoms) with photovoltaic applications. Overall, the accuracy, user-friendliness, and minimal computational overhead of the Q-Force protocol make it widely applicable for atomistic molecular dynamics simulations.



INTRODUCTION

Computational chemistry often suffers from a dilemma between reaching high accuracy and high applicability, i.e., simulating large systems for a long time. Quantum mechanical (QM) methods that can treat the intricate quantum nature of electrons (and nuclei) are computationally very demanding, even more so when their time evolution is of interest. Therefore, for many materials science and biophysics applications that require simulating thousands/millions of atoms for nano/microseconds, classical molecular mechanics (MM) following Newton's law of motion is used instead, at the cost of computational accuracy. Such MM models make use of a set of parameters, known as force fields (FFs), that are based on the topological features of the molecule (e.g., bonds and angles) to approximate its potential energy surface (PES).[1] In this work, an automated toolkit to match the MM PES of the molecule to the QM one is discussed, which allows MM models to approach the accuracy of QM, without additional computational cost after the initial reparametrization.

The traditional approach to making FFs, as it is done by the commonly used GROMOS,[2−5] OPLS,[6,7] CHARMM,[8−11] and AMBER[12,13] FFs, relies on atom types and their assumed transferability: Parameters for a set of small molecules are fitted rigorously to both experimental (e.g., mass densities, enthalpies of vaporization, and free energies of hydration) and QM (e.g., dipole moments, electrostatic potentials, and torsional barriers) data to generate a set of tabulated parameters for each atom type based on their chemical environment (e.g., alkyl carbon, ether carbon, and ester carbon). In this manner, a significant portion of the chemical space can be covered by only tens to hundreds of atom types. This approach is favorable since it allows nonexpert users to generate FFs for their molecules of study with ease and without FF developers needing to attempt the impossible task of parametrizing the whole chemical space in advance. However, the use of atom types clearly sacrifices accuracy in favor of generalization. The popularity of these transferable FFs is not only a testament to the validity of the transferability assumption to some extent, but also to the importance of having a simple procedure to determine FF parameters for any new molecule. The latter allowed them to be available to a much wider scientific community. Automated tools that assign atom types to given molecular structures and provide an FF based on these atom types, such as ATB[14] for GROMOS, CGenFF[15,16] for CHARMM, and LigParGen[17] for OPLS, have been crucial for the widespread use of these transferable FFs.

An alternative approach to the atom-type-based FF parametrization is to derive molecule-specific FF parameters for the molecule under study. This could be done using both experimental and QM data; however, the former is often not available for novel molecules. The idea of parametrizing FFs

Received: February 24, 2021
Published: July 12, 2021

https://doi.org/10.1021/acs.jctc.1c00195
J. Chem. Theory Comput. 2021, 17, 4946−4960

based on QM is definitely not new and reach back almost 50 years.[18−21] Thereafter, efforts have been made to parametrize both the intramolecular[22−24] and the intermolecular[25,26] FF terms from QM calculations. While such molecule-specific and QM-based FFs have the advantage that for each molecule the best possible FF for a given MM functional form can be created, a disadvantage is that several QM calculations and fitting of every FF parameter must be done for that molecule. This can be a daunting and error-prone task without having automated tools at the disposal. Such tools, however, started to appear in the last decades,[27−45] although their use has not yet come close to rivaling transferable FFs.

An important choice when it comes to designing a tool for molecule-specific automated FFs is the MM functional form. A complex functional form (e.g., QMDFF,[40] xTB-FF,[41] SFAM,[42] QuickFF,[34,35] and MEDFF[36]) yields a highly accurate FF, but it is orders of magnitude more expensive than transferable FFs, which limits their applicability drastically. Moreover, the use of these complex MM functional forms hinders their compatibility to highly optimized and parallelized molecular dynamics (MD) software such as GROMACS,[46] Tinker-HP,[47] and AMBER,[48] which limits their applicability even further. Alternatively, a similar functional form to transferable FFs can be used (e.g., JOYCE/PICKY,[27−32] QUBEKit,[38,39] PHF,[43] ParaMol[45]), which allows MD simulations that are at a comparable computational cost but ideally at a higher accuracy than their transferable FF counterparts.

Aside from the need for an automated toolkit, the main bottleneck for the derivation of molecule-specific FFs is formed by the cost of the QM calculations, in particular of geometry optimizations, Hessian calculations, and relaxed dihedral scans that are needed for the parametrization. As the size of the molecule grows, performing QM calculations on the whole molecule becomes increasingly difficult. This is especially true for dihedral scans since the number of dihedrals to be scanned often grows significantly with the size of the molecule. Therefore, a fragmentation scheme that divides the molecule into chemically meaningful smaller fragments becomes essential when dealing with large molecules (e.g., larger than 50−100 atoms). While such fragmentation schemes have been implemented in some automated toolkits that employ complex functional forms,[40−42] approaches that use standard functional forms have been lacking such schemes. Consequently, most of the applications in the field with such approaches have been limited to systems with less than 100 atoms.

Commonly, molecule-specific FF approaches introduce new nonbonded interaction terms or have reparametrized the coefficients of the standard Lennard-Jones (LJ) and/or Coulomb terms. However, for these altered nonbonded interactions, thermodynamic properties (which strongly depend on these interactions), if at all validated, are often done so for pristine systems and not for mixtures. Therefore, it is still an open question how well the thermodynamic properties are preserved when nonbonded interactions are derived in a molecule-specific manner. In contrast, a considerable amount of effort has been put into ensuring that the cross-molecular interactions yield accurate thermodynamic properties in transferable FFs.[1,4,49]

The aim of this paper is to introduce Q-Force, a toolkit to augment existing transferable FFs with molecule-specific FF parameters that are derived from QM calculations. The augmentation is done by retaining the nonbonded parameters of these transferable FFs and consequently also their rigorously tested thermodynamic properties. In other words, Q-Force can produce a QM-matched FF for a given molecule that can be combined with other molecules in the user's choice of FF family (e.g., OPLS, GROMOS, AMBER, CHARMM). Such an approach has the advantage that novel or nonstandard molecules can easily be augmented with Q-Force, while complex molecules which have already been parametrized rigorously (e.g., DNA, RNA, proteins) do not have to be reparametrized. This compatibility allows the easy application of Q-Force to various fields, such as computer-aided drug design, ligand-binding with biomolecules, and searching for novel organic semiconductors. Several applications of Q-Force for the latter case have recently been published.[50−53] The main characteristics of the Q-Force procedure are outlined below:

- Applicability has been an important consideration for Q-Force: it is open source, easily installable, and fully automated, which allows seamless combination with high-throughput methods.
- It can be combined with any existing transferable (or QM-based) FF through the augmentation approach.
- The automated molecular fragmentation scheme allows the treatment of molecules significantly larger than those commonly treated by similar QM-based approaches (>200 atoms). The fragment database also prevents repeated QM calculations when the same fragment is identified in different molecules.
- Significant improvements to the accuracy of the FF with respect to transferable FFs and competitive performance with respect to other QM-based approaches are demonstrated and discussed in the Results and Discussion section.
- The validation of the FF is done on-the-fly: The fitting quality with respect to the QM reference is provided during the run. Therefore, the user always knows how well the fitting procedure has been performed and whether any further action is necessary.
- It can be applied to any organic molecule and FFs can be created for their ground states, ionic states, and/or excited states.

Two major cases where we foresee the use of Q-Force will show significant improvement over the existing transferable FFs are: (1) Complex/nonstandard molecules that are not covered in the initial parametrization (training set) of the transferable FFs. While most simple functional groups are covered in the training set and perform quite well with the transferable FFs, especially (complex) conjugated ring systems and their combinations with different functional groups often require case-by-case treatment and can benefit significantly by their FFs being augmented by QM data. Several examples of such cases are treated in this paper and significant improvements can be seen for the PES for molecule-specific FFs. (2) A frequently used multiscale approach in computational chemistry is to take snapshots from MD trajectories and perform high-level QM calculations on these snapshots. In these cases, using an FF that has the same minimum energy geometry and a closely matched PES to the QM method that will later be used to perform the QM calculations on the snapshots is highly beneficial to avoid spurious results that are caused by the mismatch in geometries.[54,55] This effect has also been shown in a study using Q-Force FFs.[51]

The structure of the remainder of the paper is as follows: In the Methods section, after an initial description of the

https://doi.org/10.1021/acs.jctc.1c00195
J. Chem. Theory Comput. 2021, 17, 4946−4960

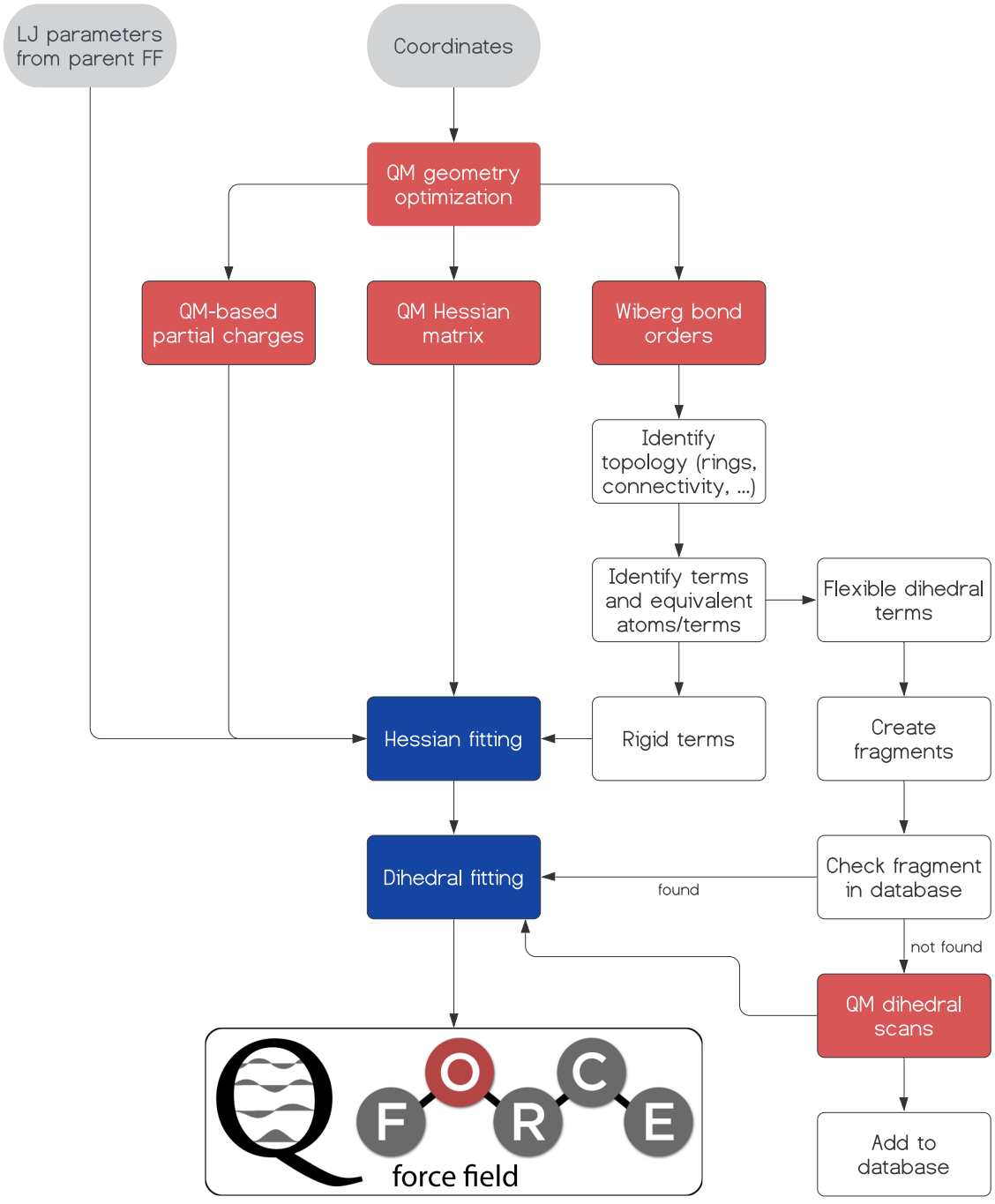


Figure 1. Q-Force flowchart. Gray boxes refer to the input, red ones to the performed QM calculations, and blue ones to the fitting steps. Further description of this flowchart can be found in the text.

functional form of the Q-Force FF, the steps of the Q-Force procedure (as depicted by the flowchart in Figure 1) will be explained, technical and computational details will be shown, and known limitations of the model will be discussed. In the Results and Discussion section, the performance of the Q-Force FF for several small molecules will be benchmarked to a QM reference as a proof-of-concept and further compared to existing transferable and QM-based FFs. The thermodynamic properties of these FFs are also demonstrated. Then, the importance of QM augmentation of the FF will be demonstrated for nonstandard dihedral profiles, where the performance of transferable FFs is shown to quickly deteriorate. Finally, materials science applications of Q-Force will further evidence the impact of augmenting FFs with accurate QM references on obtaining accurate PESs. The Q-Force toolkit, several examples, and tutorials on its usage are freely available on GitHub (https://github.com/selimsami/qforce).

METHODS

Definition of the Q-Force Force Field. FFs contain terms that describe the interactions within the molecule (intramolecular) and between different molecules (intermolecular). The intramolecular part is composed of bonded interactions between closely neighboring atoms (e.g., bonds, angles, dihedrals) and nonbonded interactions within the

https://doi.org/10.1021/acs.jctc.1c00195
J. Chem. Theory Comput. 2021, 17, 4946−4960

molecule. The intermolecular part is composed of nonbonded interactions that occur between different molecules. The total potential energy of the system (V_{pot}) is given by the sum of these interactions

$$V_{pot} = V_{bonds} + V_{angles} + V_{dihedrals} + V_{Coulomb} + V_{LJ} \quad (1)$$

In the following, we describe the functional form of these FF terms that are used in Q-Force. The bond potentials between atom pairs A and B that form a bond AB are treated with the harmonic potential

$$V_{bonds} = \sum_{AB} \frac{1}{2} k_r^{AB} (r^{AB} - r_0^{AB}) \quad (2)$$

where k_r is the force constant and r_0 is the equilibrium bond length. The bond angle potentials for atom triplets of A, B, and C that form an angle ABC are treated by

$$V_{angles} = \sum_{ABC} \frac{1}{2} k_\alpha^{ABC} (\alpha^{ABC} - \alpha_0^{ABC}) + \sum_{AC} \frac{1}{2} k_r^{AC} (r^{AC} - r_0^{AC}) \quad (3)$$

where the first term is a harmonic angle potential for the ABC angle and the second term is a harmonic bond potential between the A and C atoms that is also known as the Urey–Bradley term.[56,57]

For dihedrals, three different functions are employed. For rigid dihedrals, which are proper and improper dihedrals that are constrained to a single minimum (e.g., conjugated rings or double bonds), a harmonic dihedral term given by

$$V_{dihedrals-rigid}(\theta) = \sum_{ABCD} \frac{1}{2} k_\theta^{ABCD} (\theta^{ABCD} - \theta_0^{ABCD}) \quad (4)$$

is used. Inversion dihedrals, which are nonplanar dihedrals (>25°) with a double-well potential (e.g., nonconjugated rings or pyramidal inversions), are represented by

$$V_{dihedrals-inversion}(\gamma) = \sum_{ABCD} k_\gamma^{ABCD} (\cos\gamma^{ABCD} - \cos\gamma_0^{ABCD})^2 \quad (5)$$

For flexible dihedrals, which are dihedrals that have multiple accessible minima (e.g., alkane chain), a Ryckaert–Bellemans type[58] of dihedral function given by

$$V_{dihedrals-flexible}(\phi) = \sum_{ABCD} \sum_{n=0}^{5} C_n^{ABCD} (\cos(\phi - \pi))^n \quad (6)$$

is used, where C_n are the six constants that determine the shape and height of each dihedral profile.

In the case of triplets of colinear atoms (e.g., acetonitrile), identified by an angle larger than 170°, the Urey–Bradley term from eq 3 is skipped to avoid generating linear dependencies. Additionally, the ill-defined dihedral terms containing these atoms are also not generated.

A Coulomb potential given by

$$V_{Coulomb} = \sum_{A<B} \frac{q_A q_B}{r_{AB}} \quad (7)$$

is used to account for the electrostatic interactions between particle pairs A and B that are more than two bonds apart. By default, the CM5 method[59] based on Hirschfeld partitioning[60−62] is used to determine the atomic charges. These charges are scaled by 1.2 in the case of neutral molecules to account for the condensed-phase polarization.[63] CM5 charges have been shown to give consistent[59] and accurate[63,64] atomic charges for MD applications, as also demonstrated in this work. However, users have the option to use a charge method that better matches their parent FF to increase the compatibility (e.g., RESP charges with Hartree–Fock 6-31G* for GAFF).

A 6−12 LJ potential given by

$$V_{LJ} = \sum_{A<B} \frac{C_{AB}^{(12)}}{r_{AB}^{12}} - \frac{C_{AB}^{(6)}}{r_{AB}^{6}} \quad (8)$$

is used to account for the repulsion and dispersion interactions between the particles. As previously described, these terms are inherited from the parent FF. For nonbonded interactions between atom pairs that are three bonds apart (commonly named as 1−4 interactions), the treatment of the parent FF is followed, which can be either scaled-down interactions (e.g., OPLS, AMBER) or special pair interactions (e.g., GROMOS).

Identifying the Topology and FF Terms. The distance matrix of the QM optimized geometry is used to create a graphical representation[65] of the molecule where atoms are registered as nodes and bonds correspond to the edges between the nodes of the graph. Such a graphical representation allows the determination of all of the bonds, angles, dihedrals, and rings in the molecule with ease and is further useful in the fragmentation procedure that is discussed later. Wiberg bond orders[66] are used to determine conjugated, double, and triple bonds of the molecule. This information, along with the previously determined topology of the molecule, is then used for determining the FF terms and their categorization into rigid (bonds, angles, rigid dihedrals, inversion dihedrals) and flexible (flexible dihedrals) terms.

While the determination of bonds, angles, and dihedrals in a molecule is quite straightforward, automatic categorization of dihedrals into rigid, inversion, and flexible dihedrals requires further discussion. A rigid dihedral is recognized in the following cases: (1) the bond order between the central atoms of the dihedral is 1.75 or higher, (2) the dihedral is in a ring and it is planar (the deviation from planarity is smaller than 25°), (3) a special type of rigid dihedral is necessary (improper dihedral) when previously mentioned types of rigid dihedrals are not present near an atom that needs to be kept planar (e.g., for a ketone, an improper dihedral will be necessary, while for a phenol it will not be necessary since the oxygen is kept planar by the rigid dihedrals originating in the phenyl ring). In the first two cases, a rigid dihedral term will be added to all combinations of atoms containing the same central atoms and in the last case, a single improper dihedral will be added. For the last two cases, if the deviation from planarity is larger than 25°, then an inversion dihedral with a double-well potential is added. The remainder of the dihedral terms will be recognized as flexible dihedrals. These dihedrals are expected to have multiple accessible energetical minima. These are located in linear chains where a 360° rotation is technically possible. For these dihedrals, a full 360° QM and MM scan is performed for the fitting.

Equivalence of Atoms and FF Terms. Transferable FFs rely on atom types, which categorize atoms based on their chemical environment (e.g., alkyl carbon, ether carbon, ester carbon), to cover a large portion of the chemical space with a relatively small training set. Such an approach provides generality at the cost of accuracy: Atoms with somewhat

https://doi.org/10.1021/acs.jctc.1c00195
J. Chem. Theory Comput. 2021, 17, 4946−4960

different environments are mapped to the same atom type even though, for example, not all ether carbons are equivalent.

For molecule-specific FFs, since the best set of parameters is derived for each atom, approximating atoms to predetermined atom types is unnecessary. However, it is beneficial to determine the equivalent atoms and FF terms within the molecule for several reasons: (1) it avoids spurious small differences between chemically equivalent atoms and FF terms; (2) for most molecules, considering the equivalence of FF terms greatly reduces the number of fitting parameters for the Hessian matrix which is beneficial for both the stability of the fitting procedure and for avoiding overfitting; and (3) it prevents duplicate QM dihedral scan calculations for chemically equivalent fragments.

The default approach in Q-Force is to generate an identifier for each atom based on their first four neighbors (this number can be changed or the equivalent atom identification can be turned off by the user). The identifier contains the path to each atom that is up to four bonds apart. The elements of the atoms and the bond types (i.e., single, conjugated, double) of the bonds that are within the path are written to that identifier. The equivalent FF terms are then identified based on the equivalent atoms with special considerations that are not discussed here (e.g., bond terms between two atom pairs, with both pairs having identifiers A and B, are not necessarily equivalent).

Hessian Fitting. After the determination of the rigid FF terms, force constants (k) corresponding to these terms (as shown in eqs 2−4) are determined by a linear least-squares fitting where the squared difference between the QM and MM Hessians given by

$$\sum_{ij} (H_{ij}^{\mathrm{QM}} - H_{ij}^{\mathrm{MM}}(k, \mathrm{NB}))^2 \tag{9}$$

is minimized with the constraint that force constants have positive values. Here, H_{ij} refers to the Cartesian Hessian matrix elements. While the force constants (k) of the rigid FF terms are being fitted, the nonbonded (NB) interactions are also accounted for in the MM Hessian matrix but not fitted. Accounting of the nonbonded interactions during the Hessian fitting is essential, especially for molecules with strong intramolecular interactions, as otherwise these nonbonded interactions can affect the parametrization of the bonded terms and the final parameters can be dependent on the minimum at which the Hessian calculation is performed.

A Hessian fitting approach like this has originally been used by Dasgupta et al.[23] and more recently in the QMDFF procedure and has been shown to give accurate harmonic vibrational frequencies. In our implementation, forces are computed analytically and the Hessian is computed numerically with a three-point derivative with displacements of 0.003 Å in both directions. The Hessian terms that are smaller than 10^{-4} kJ/mol/Å^2 are discarded to improve performance and because they have no effect on the results. The calculation and fitting of the MM Hessian is extremely fast, as highly optimized SciPy[67] routines are used: It takes less than a minute for molecules containing 200 atoms on an average laptop.

Fitting of Flexible Dihedrals. While the Hessian matrix contains all of the necessary information for the parametrization of the rigid FF terms at the optimized geometry, for flexible dihedrals that have multiple minima, the Hessian matrix contains insufficient information. Therefore, relaxed dihedral scans (i.e., at each scan interval, the rest of the geometry is optimized, while the scanned dihedral is frozen) must be performed both at QM and MM levels to determine the corresponding dihedral energy profiles. An MM relaxed scan requires an FF, which is available at this stage of the procedure, as seen in Figure 1 (the rigid bonded terms obtained from the Hessian fitting, partial charges obtained from the QM calculations, and the LJ parameters retained from the parent FF). Then, the difference between the two profiles, ΔV

$$\Delta V_{\mathrm{profile}} = V_{\mathrm{profile}}^{\mathrm{QM}} - V_{\mathrm{profile}}^{\mathrm{MM}} \tag{10}$$

can be fitted to a Ryckaert−Bellemans type of dihedral function (see eq 6) using a least-squares fitting. For the fitting, higher weights (w) are given to data points that are lower in energy using

$$w = \exp(-0.2 \times \sqrt{V_{\mathrm{profile}}^{\mathrm{QM}}}) \tag{11}$$

where w scales the residuals during the fitting. This weighting step has minor effects on most of the dihedrals for which a near-perfect fit can be obtained. However, it becomes highly beneficial for more complicated profiles (as exemplified in the results) where a perfect fit cannot be obtained. In these cases, prioritizing the accuracy of the lower energy barriers is sensible as the dynamics are more affected by them (e.g., a 5 kJ/mol error in a 15 kJ/mol barrier is more detrimental than the same error on a 100 kJ/mol barrier—both because of the percent error difference and also because a 100 kJ/mol barrier will be much less frequently visited during a regular MD run). Equation 11, in our experience, has shown to give a good balance between prioritizing lower energies and not neglecting higher energy barriers.

A serial relaxed dihedral scan (i.e., the dihedral angle is scanned consecutively in a single direction), as opposed to single-point calculations at each scan interval, is necessary but not always sufficient to obtain accurate and reliable dihedral energy profiles. A known problem[68,69] with serial relaxed scans is that in some cases, the starting geometry of a given step, which is the optimized geometry of the previous step, can guide the optimization toward a spurious local minimum which makes the profile dependent on the scan direction and causes asymmetries in a dihedral profile that should be symmetrical. Recently, a step-wise wavefront propagation method, called TorsionDrive, in which extra relaxed optimizations are performed in reverse and forward directions, has been shown to avoid these spurious asymmetries, however, at a 3.5−7.5 times increased computational cost.[69] In Q-Force, the user can conveniently request the symmetrization of such spurious asymmetries, allowing us to reach a similar outcome without any additional computational cost. This approach is exemplified in the Results and Discussion section.

Another source of problem is the coupled dihedrals where considering a single scan could lead to the wrong QM dihedral profile as the reference. In Q-Force, this can be partly alleviated by providing multiple QM scans for a specific dihedral (for example, a reverse scan or a scan where the coupled dihedral is at a different angle), which will then automatically take the lowest energy path along the scan and discard higher-energy configurations. A more rigorous treatment of such dihedrals would be the two-dimensional (2D) dihedral scans and the corresponding CMAP potentials,[1,70,71] and this is currently a work in progress.

https://doi.org/10.1021/acs.jctc.1c00195
J. Chem. Theory Comput. 2021, 17, 4946−4960

Fragmentation and the Fragment Database. As the size of the molecule grows, not only the cost of individual calculations grows, but often so does the number of relaxed dihedral scans that needs to be performed. For example, assuming a 200-atom molecule with 25 flexible dihedrals and 15° scan intervals, one would need to do 600 geometry optimizations on a 200-atom molecule. Clearly, this quickly becomes unfeasible. As these relaxed QM dihedral scans were identified as the computational bottleneck of the procedure, a fragmentation scheme was implemented to construct chemically meaningful fragments for each dihedral that needs to be scanned. This resulted in a drastic lowering of the computational cost and made Q-Force applicable to very large (>200 atoms) molecules. Aside from the computational cost, our experience shows that performing dihedral scans on smaller fragments also makes the spurious asymmetries discussed in the previous subsection less likely to happen.

The automated fragment generation works in the following way: When a flexible dihedral is identified, the first three neighbors of the central atoms of that dihedral are by default taken as part of the fragment. Then, the next atoms are removed and the fragment is capped with a hydrogen atom unless: (1) the bond with the next neighbor has a bond order larger than 1.75; (2) the bond with the next neighbor is part of a ring; (3) the current atom has an electronegativity in Pauling scale larger than 3.0 (to prevent unwanted intramolecular hydrogen bonds, e.g., an O−C bond is not replaced by a O−H). If one of these cases holds true, the next atom is included into the fragment and the same check is done for the neighbors of the new atom. The process is repeated in all paths until either breakable bonds are identified or the end of the paths is reached. Two illustrations to the automated fragment generation are given in Figure 2.

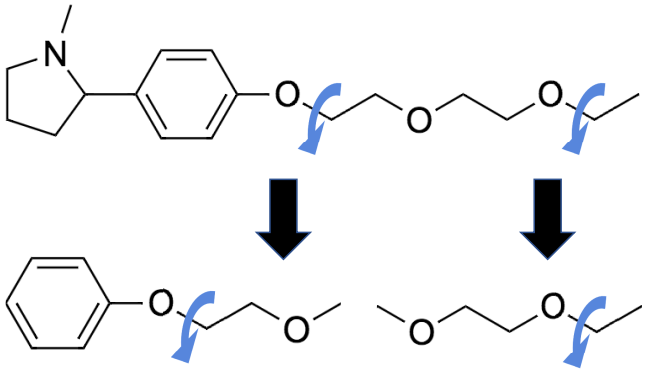


Figure 2. Two examples of the automatic fragmentation procedure. It is shown that fragments are not terminated in the middle of a ring (left) or when a highly electronegative atom has to be capped with a hydrogen (right).

As identical molecular fragments can exist in multiple molecules, we have also implemented a database (currently only stored locally) where previously scanned fragments (together with the employed computational method) are stored. And if these fragments are ever identified on a new molecule again, the scan results are retrieved from the database, preventing additional calculations. The identification of the fragments is done using the previously discussed graphical representation of the molecule and its fragments. Two fragments are considered identical if they have the same connectivities between the same elements, and with matching bond orders and lone pair electrons. Bond orders are obtained from a Wiberg bond order analysis[66] and the lone pair electrons are determined from a natural bond orbitals (NBO) analysis.[72]

Validation on-the-Fly. An important but sometimes lacking ability when using automated and (semi-)black-box procedures is to be aware of when the procedure performs well and when it performs suboptimally. In Q-Force, the validation of the FF is done on-the-fly: In the case of the Hessian fitting, QM and MM vibrational frequencies are plotted and the mean percent error of the MM vibrational frequencies is given. Moreover, MM vibrational modes are also written to a file in a format that can be visualized by VMD[73] and be compared to their QM counterparts. In the case of dihedral fitting, for each flexible dihedral, QM and MM dihedral profiles are plotted together and the R^2 value of the fitting is given as an indication of how well the fitting went. A warning is given if the R^2 is lower than 0.9 and any fitted data point has an error larger than 2 kJ/mol, prompting the user to check if the results have the desired accuracy. These on-the-fly validations allow users to be always aware of the accuracy of the generated FF to decide whether further action is necessary.

Limitations. Some known, and likely not exhaustive, limitations of the Q-Force procedure are discussed below:

- The choice of the functional form of the Q-Force FF enables treatment of large systems, however, at the same time it limits its accuracy compared to more complex functional form FFs (QMDFF, xTB-FF, etc.). Bond breaking, bond forming, and bond order changes cannot occur with the current functional form. Organometallic complexes and metal clusters are outside the scope of this work. Similarly, electrostatics are currently only treated by point charges and charges cannot polarize each other during the simulation. A polarizable version of the Q-Force FF is currently in progress.
- It is well known that some neighboring flexible dihedral profiles are mutually dependent. In such cases, two-dimensional dihedral profiles can be assigned to neighboring dihedrals, often referred to as CMAP correction.[1,70,71] Identification and special treatment of these co-dependent dihedrals, and CMAP potentials are currently not implemented in Q-Force. However, their implementation is currently in progress.
- The fragmentation scheme has only been implemented for the treatment of flexible dihedrals and not for the calculation of the Hessian matrix. Consequently, treatment of extremely large molecules (>∼500 atoms) might be limited by the computational cost of the geometry optimization and Hessian calculation of the whole molecule (unless a cheaper method, such as DFTB, is used for the parametrization). The problem with fragmentation before a geometry optimization is that it relies on accurate input geometries, and in many cases, it can result in wrong bond orders and consequently wrong bonds being cut during the creation of the fragments.

Technical Details. The default choice for the QM reference is density functional theory with the Perdew−Burke−Ernzerhof (PBE) functional,[74] 6-31+G* basis set,[75] and D3-BJ[76] empirical dispersion correction. The PBE functional is chosen because its Hessian matrix does not

https://doi.org/10.1021/acs.jctc.1c00195
J. Chem. Theory Comput. 2021, 17, 4946−4960

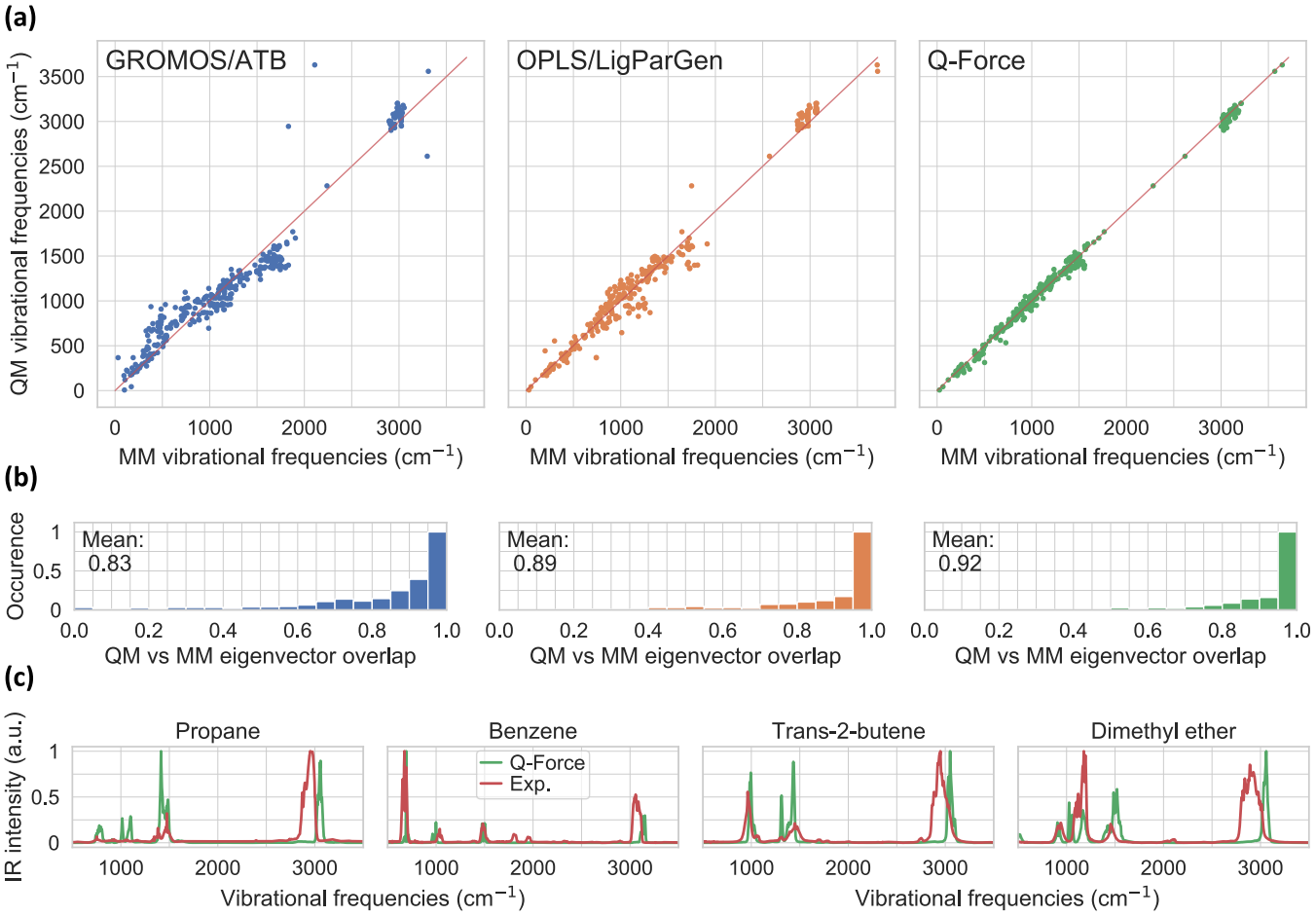


Figure 3. (a) QM vs MM vibrational frequencies for 16 small molecules, with a total of 494 vibrational frequencies for the GROMOS/ATB, OPLS/LigParGen, Q-Force (GROMOS-based) FFs. The red line corresponds to perfect agreement between QM and MM. MADs in vibrational frequencies are 12.9% (134.2 cm^{-1}) for GROMOS/ATB, 8.4% (89.5 cm^{-1}) for OPLS/LigParGen, and 4.2% (37.3 cm^{-1}) for Q-Force. The molecules are propane, isobutane, acetic acid, trans-2-butene, ethanol, acetonitrile, dimethyl ether, methane thiol, pyrazine, thiophene, ethene, benzene, toluene, naphthalene, fluorobenzene, and 1,2-dichloroethane. (b) Normalized occurrence of the overlap between the QM and MM normal modes. A value of 1 corresponds to perfect overlap. The average overlaps between QM and FFs are 0.83, 0.89, and 0.92 for GROMOS, OPLS, and Q-Force, respectively. (c) Experimental (red)[84] and Q-Force (green) gas-phase IR spectra for a selection of molecules. The maximum intensity is normalized to one.

need empirical scaling as with some other functionals to match experimental frequencies.[40] However, the user has the choice to use other functionals and basis sets and scale their Hessian accordingly as the scaling is implemented. For example, for excited states, different functionals may be more preferable.[77,78] Currently, the compatible QM software packages are Gaussian (09 and 16)[79,80] and Q-Chem.[81] Other QM packages can easily be added due to the modular structure of the code. The FF is generated in GROMACS format, which is also readily compatible with OpenMM[82] and can easily be converted to AMBER or CHARMM formats using the ParmEd software.[83]

RESULTS AND DISCUSSION

Vibrational Analysis of Small Molecules. We first look at how Q-Force performs for small molecules with respect to two very commonly used automated topology builders for transferable FFs, namely, ATB for GROMOS FFs and LigParGen for OPLS FFs. The computation of the vibrational frequencies, modes and IR spectra, and the method used for matching QM and MM vibrational frequencies and modes to each other are described in the Supporting Information. Figure 3a shows results comparing the vibrational frequencies of these two transferable FFs and Q-Force to the QM reference for a molecule set comprising 16 small organic molecules (see the figure caption for the list of the molecules). Q-Force outperforms both transferable FFs for every tested molecule with a mean absolute deviation (MAD) in vibrational frequencies of 3.6% (30.5 cm^{-1}) compared to OPLS/LigParGen with 7.3% (78.0 cm^{-1}) and GROMOS/ATB with 12.6% (129.4 cm^{-1}). This is not particularly surprising as Q-Force was parametrized against the QM reference while the transferable FFs were not. However, it is also important to keep in mind that these transferable FFs have been parameterized for this kind of small molecules—the accuracy of these methods is expected to further decrease for more complex molecules that were not covered in the initial parametrization, as also observed in this work.

A closer look at Figure 3a also helps to identify systematic errors in transferable FFs. For example, for the GROMOS/ATB FF, frequencies in the range of 400−700 cm^{-1} have been consistently underestimated. Many of these data points correspond to aromatic ring out-of-plane bending modes, resulting in GROMOS aromatic rings being much less planar

https://doi.org/10.1021/acs.jctc.1c00195
J. Chem. Theory Comput. 2021, 17, 4946−4960

than they are supposed to be. This behavior has been previously noticed and sometimes these parameters are empirically scaled up to make them "stiffer".[51] The use of QM-derived FFs such as Q-Force can also help to identify and correct such systematic errors in existing transferable FFs by means of optimizing the parameters of existing atom types and identifying when extra atom types are necessary.

Along with the vibrational frequencies, the vibrational normal modes corresponding to these frequencies are also important to be in agreement with QM. Figure 3b shows the overlap between QM and MM vibrational modes. While all FFs perform quite well for these small molecules, Q-Force is shown to have the highest mean overlap (0.92). Figure 3c shows gas-phase IR spectra corresponding to a selection of molecules for Q-Force (green) and experiments (red). The results show that the Q-Force-based FF is able to qualitatively reproduce the experimental spectra.

While performed on a different set of small molecules, QMDFF and SFAM, which have much more complicated and computational demanding functional forms of the FF, reported MADs of 44[40] and 66.5 cm^{-1},[42] respectively. QUBEKit[39] and PHF,[43] on the other hand, with a similar functional form to Q-Force, reported a MAD of 6.3% and 73 cm^{-1}, respectively. This shows that the Q-Force FF (with a MAD of 4.2% or 37.3 cm^{-1}) gives, if not better, comparable vibrational frequencies to the existing molecule-specific FFs, even to those with more complex functional forms. A more direct comparison can also be made between the vibrational analysis of thiophene with the Q-Force FF in Figure S1 with another QM-derived FF based on the PICKY protocol present in Figure 3 of ref 32. While both FFs perform very well for this molecule, slightly higher overlaps are observed for the Q-Force FF and the match between the QM and MM vibrational frequencies is higher for Q-Force, particularly at lower frequencies. The better match at lower frequencies hints at the better treatment of the in-ring distortions; however, a more extensive comparison is needed to draw definitive conclusions.

Dihedral Profiles of Small Molecules. After demonstrating the accuracy of the Q-Force vibrational frequencies, the next step toward obtaining accurate MM PESs is having accurate dihedral profiles for the flexible dihedrals. In fact, the performance of the flexible dihedral functions around the global minimum is also evident in the vibrational frequencies shown in Figure 3. The frequencies below ~300 cm^{-1} often correspond to these kinds of low-energy displacements of the flexible dihedrals and they are also shown to be in good agreement with the QM reference. However, since the Hessian contains no information about other possible minima, the complete dihedral profile must be investigated to obtain accurate PESs.

To this end, QM and Q-Force dihedral profiles are compared in Figure 4a for all of the flexible dihedrals of the same 16 small molecules from Figure 3. For all dihedrals, a very good agreement (MAD of 0.21 kJ/mol and a maximum deviation of 1.25 kJ/mol) is obtained between the QM and MD profiles using the automated procedures in Q-Force. This shows that the Q-Force procedure is also able to accurately reproduce the QM dihedral profiles. Moreover, due to the database of dihedral profiles in Q-Force, if these molecules/fragments are identified in another molecule in the future, they would be identified automatically and no new calculations would be necessary.

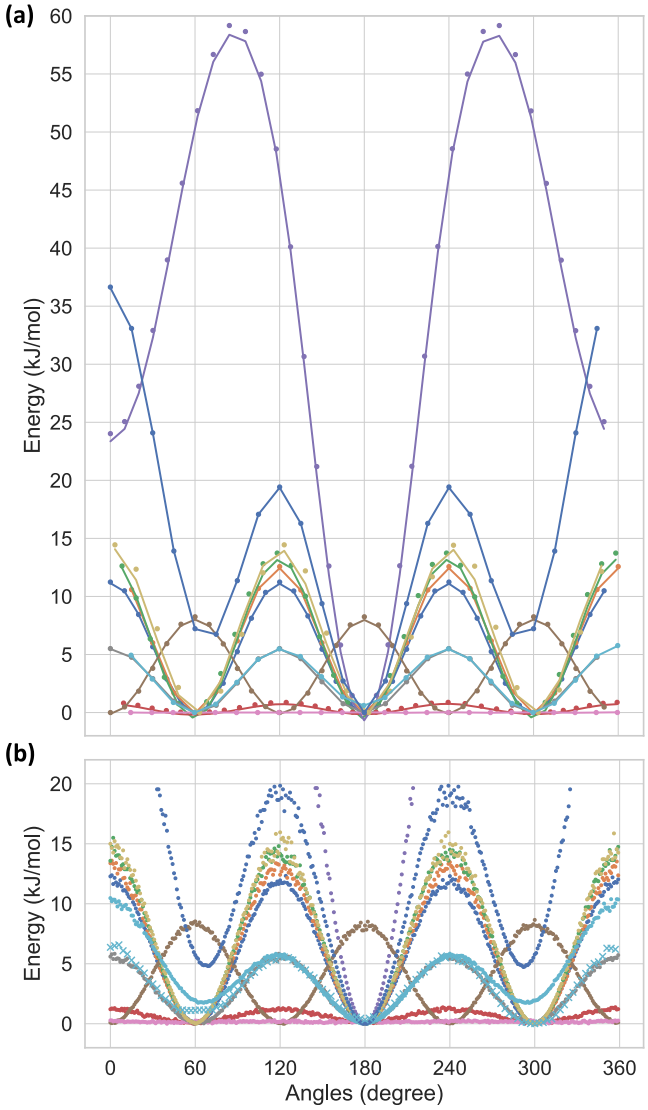


Figure 4. (a) QM (dots) vs Q-Force (lines) dihedral profiles obtained through relaxed scans for all of the flexible dihedrals of the molecules in Figure 3, with a total of 11 dihedrals. MAD is 0.21 kJ/mol, and the maximum deviation of any point in any dihedral profile is 1.25 kJ/mol. (b) Dihedral profiles obtained by the Boltzmann inversion (eq 12) of the dihedral distributions in the condensed phase with the Q-Force FF. The data points beyond a 20 kJ/mol barrier are not presented as these are not properly sampled during the simulations. Uniquely, the cyan profile drawn with crosses instead of dots corresponds to a gas-phase simulation (see text for further discussion).

To investigate the dynamics of these dihedrals, the dihedral distributions in condensed-phase simulations were converted to free energy profiles (Figure 4b) through an inverse Boltzmann analysis given by

$$E_\alpha = RT \log\left(\frac{n_\alpha}{N}\right) \tag{12}$$

where R is the gas constant, T is the temperature, n_α is the number of occurrences of the angle α, and N is the total number of data points. Comparing the scanned QM and Q-Force dihedral profiles (Figure 4a) with the ones from the MD simulations (Figure 4b), a very good match is observed between the two for all dihedrals, indicating consistency

https://doi.org/10.1021/acs.jctc.1c00195
J. Chem. Theory Comput. 2021, 17, 4946−4960

between the fitting procedure and the observed behavior during molecular dynamics. The only noticeable difference between the two is the cyan profile, which corresponds to the rotation of the OH hydrogen in ethanol, and it has a higher barrier in the condensed phase (dots) around 0°. This is a consequence of the hydrogen bonding being less available in the cis configuration (0°) and thus having a high free energy barrier and not a discrepancy due to the Q-Force procedure. This is demonstrated by the agreement of the gas-phase profile for this dihedral (cyan with crosses) with the QM profile.

Thermodynamic Properties. To investigate the effect of the Q-Force procedure on thermodynamic properties, two Q-Force FFs have been generated for each of the 16 molecules discussed earlier. The first set of FFs has inherited both LJ parameters and atomic charges from the parent FF of GROMOS/ATB, having Q-Force only re-parametrize the intramolecular interactions. In the second set, only LJ parameters have been inherited from GROMOS/ATB, while the atomic charges were computed with the CM5 method. CM5 charges were scaled by 1.2 to account for the condensed-phase polarization.[63] Similar calculations were also done with two reference FFs, namely, GROMOS/ATB and OPLS/LigParGen.

Table 1 shows MAD of mass density and enthalpy of vaporization with respect to experiments for the four sets of FFs. Additionally, for the two Q-Force FFs, their MAD to the parent GROMOS/ATB FF is also shown. As demonstrated, retaining both the LJ parameters and the atomic charges results in a minimal difference between Q-Force and its parent FF, strongly indicating that the compatibility to the rest of the GROMOS FF library is retained. Switching to CM5 charges, on the other hand, drastically increases this difference, especially for the enthalpy of vaporization. However, as the comparison to the experiment indicates, the increased difference between Q-Force (1.2 × CM5 charges) and GROMOS/ATB seems to be toward a better agreement with experiments, as MAD in both density and enthalpy of vaporization decreases appreciably for the Q-Force (1.2 × CM5 charges) FF.

Table 1. MAD of Density (ρ, g/cm^3) and Enthalpy of Vaporization (ΔH_{vap}, kJ/mol) to Experiment for the Same 16 Small Molecules[a]

force field	ρ	ΔH_{vap}
	MAD to experiment	
OPLS/LigParGen	0.029	2.9
GROMOS/ATB	0.052	4.5
Q-Force (q = GROMOS)	0.046	4.7
Q-Force (q = 1.2 × CM5)	0.032	2.7
	MAD to GROMOS	
Q-Force (q = GROMOS)	0.012	0.7
Q-Force (q = 1.2 × CM5)	0.031	3.8

[a]For the Q-Force FFs, which uses GROMOS/ATB atom types for the LJ interactions in this study, MAD to GROMOS/ATB is also shown. Data points corresponding to individual molecules are presented in the Supporting Information.

These results show that using the Q-Force procedure with the exactly same nonbonded interactions as the parent FF (Q-Force with GROMOS/ATB charges) will result in very similar thermodynamic properties. This reaffirms that the thermodynamic properties are mostly governed by nonbonded interactions and that FFs augmented with Q-Force can be expected to be compatible in this respect with the parent FF family. Additionally, it has been shown that using QM-based and molecule-specific atomic charges (Q-Force with 1.2 × CM5 charges) can improve the thermodynamic properties relative to the parent FF, as it has done for the set of molecules in this study.

Treatment of Problematic Dihedral Scans. As discussed in the Methods section, serial relaxed dihedral scans can sometimes result in spuriously asymmetric dihedral profiles. This is exemplified in Figure 5 for the 4- and 2-pyridiniophenolate molecules, where the QM profiles (in gray) suffer such asymmetry, especially around the transition region of 90 and 270° even though 0−90, 90−180, 180−270, 270−360° quadrants are expected to be equivalent or mirrored based on the molecular symmetry. The reason behind this asymmetry is that during the scan, the improper dihedrals between the two rings deviate from planarity into a pyramidal structure and are stuck at a local minimum due to the high pyramidal inversion energy barrier until the point where the sharp drop in energy occurs. This clearly indicates a hysteresis effect, possibly due to the neglect of multireference effects in the wavefunction. Evidently, using these asymmetric profiles for the subsequent fitting of the FF parameters would introduce errors. The Q-Force procedure allows the automatic symmetrization of the dihedral profile, taking only the lower-energy data points from the equivalent quadrants, as shown in Figure 5 (in red). A similar molecule with a similar asymmetry problem was studied with the TorsionDrive method[69] and following their procedure, they have obtained a very similar symmetrized profile. Note that there is no additional computational cost in the approach implemented in Q-Force, unlike the TorsionDrive method, which, while being a more rigorous approach, resulted, for a similar molecule, in an ∼4 times increase in the number of performed relaxed scans.

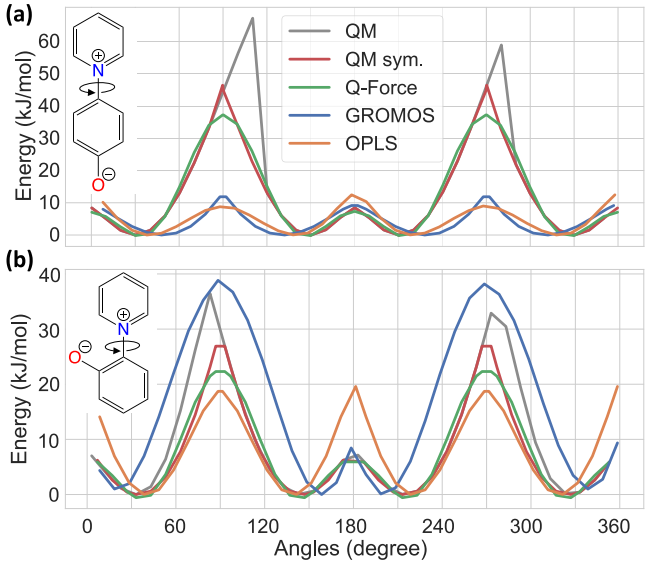


Figure 5. QM, symmetrized QM, Q-Force, GROMOS/ATB, and OPLS/LigParGen dihedral profiles for (a) 4-pyridiniophenolate and (b) 2-pyridiniophenolate. The molecules and the scanned dihedral are indicated in the insets.

The fitting of the Q-Force dihedral profiles then makes use of the symmetrized QM profile and as a result, a profile that is

https://doi.org/10.1021/acs.jctc.1c00195
J. Chem. Theory Comput. 2021, 17, 4946−4960

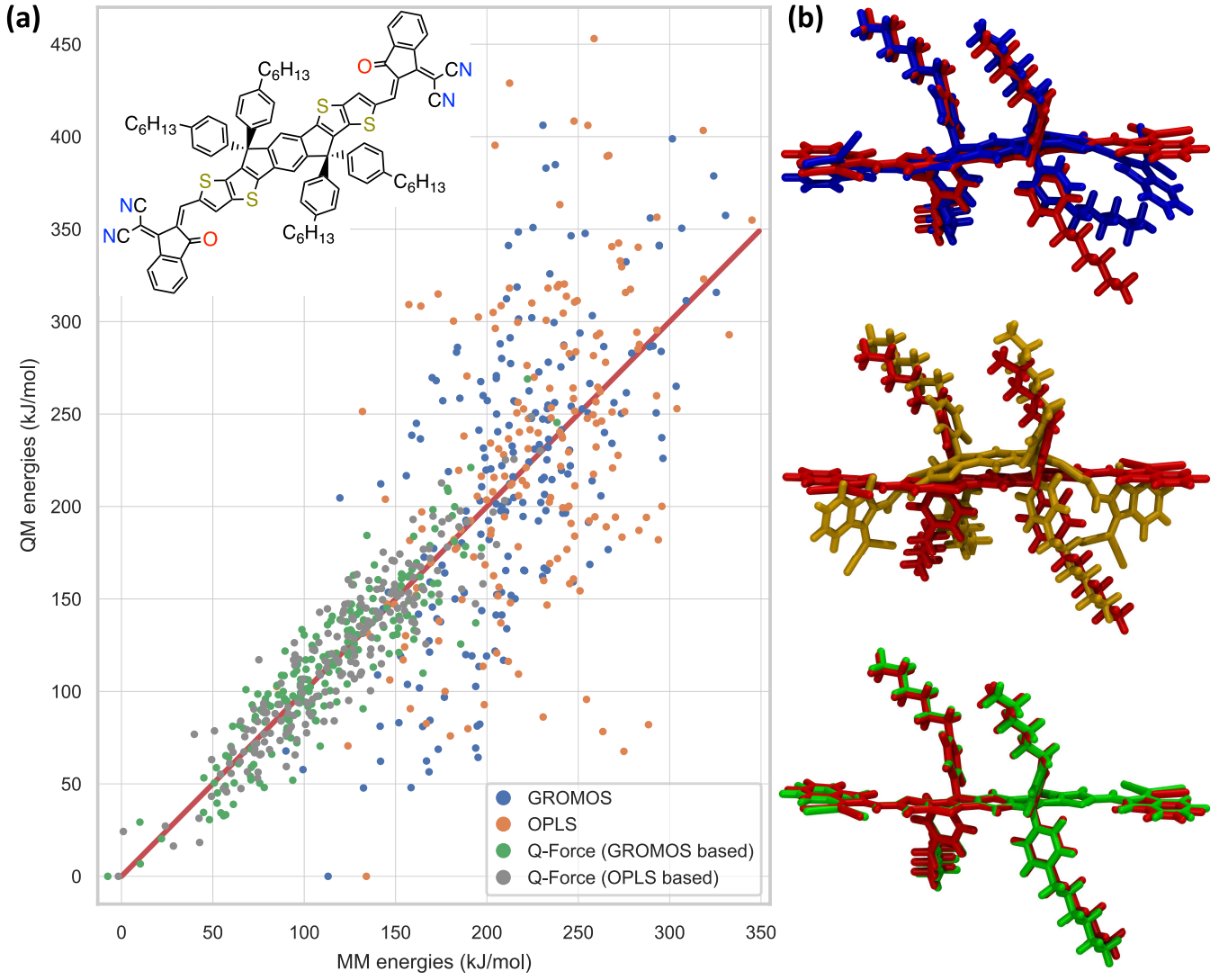


Figure 6. (a) QM vs MM potential energies for the 200 snapshots taken with each FF with the red line indicating a perfect match. The minimum QM energy is set to zero, and the FF energies are shifted to minimize the error with respect to the QM results. MAD between the QM and MD energies are 47.7, 53.6, 14.5, and 13.5 kJ/mol for GROMOS, OPLS, Q-Force (GROMOS-based), and Q-Force (OPLS-based) FFs, respectively. The differences (error) in standard deviation (σ) between QM and FF energies (calculated as $100 \times (\sigma_{QM} - \sigma_{FF})/\sigma_{QM}$) are 82, 85, 7, and 5%, respectively. (b) QM optimized ITIC structures (in red) overlaid with the GROMOS (blue), OPLS (orange), and Q-Force (green) optimized structures. The two Q-Force-based FFs had almost indistinguishable structures; therefore, only one is presented here. The results of the Hessian fitting and dihedral profiles for the ITIC molecule are presented in the Supporting Information (Figures S2 and S4).

symmetrical and in good agreement with the QM profile is obtained for both molecules. The effect of the weighting function from eq 11 can also be seen here: The more important lower-energy points are reproduced with high accuracy at the cost of some level of accuracy for the high-energy points. GROMOS as well as OPLS FFs, both of which clearly were not trained based on these types of molecules, fail drastically to reproduce the dihedral profiles for both molecules. This clearly illustrates the limitation of transferable FFs—that they cannot be expected to work accurately for the type of molecules that were not covered in the original fitting of the FF.

QM vs MM Potential Energy Surfaces. Having shown that the Q-Force procedure leads to both accurate Hessian and dihedral profiles, the next question is whether this accuracy will lead to a good agreement between the QM and MM PESs. To test the PES of the Q-Force FF and the fragmentation procedure on a large molecule, we have chosen the ITIC molecule (186 atoms). This molecule[85,86] and its family of molecules[87,88] have been shown to lead to record power conversion efficiencies for organic photovoltaics applications. For this molecule, 200 snapshots were taken from gas-phase MM simulations of ITIC with two transferable FFs (GROMOS/ATB and OPLS/LigParGen) and the corresponding Q-Force FFs (GROMOS-based, OPLS-based); then, QM single-point energy calculations were performed on these snapshots and the potential energies from the corresponding geometries were compared (Figure 6a). Additionally, Figure 6b shows QM optimized geometries (in red) overlaid with the optimized GROMOS (in blue), OPLS (in orange), and Q-Force (in green) geometries. By visual inspection, a good match can be observed between the Q-Force and QM structures while both transferable FFs show significant deviation, most importantly losing the planarity of the backbone. This is also evident from the root-mean-square deviation (RMSD) between the QM and MM structures,

https://doi.org/10.1021/acs.jctc.1c00195
J. Chem. Theory Comput. 2021, 17, 4946−4960

which are 1.5, 3.2, and 0.3 Å for GROMOS, OPLS, and Q-Force, respectively.

The results of this approach are shown in Figure 6a. MADs between the QM and MM energies are 47.7, 53.6, 14.5. and 13.5 kJ/mol for GROMOS, OPLS, Q-Force (GROMOS-based), and Q-Force (OPLS-based) FFs, respectively. These results show that Q-Force reproduces the QM PES extremely well with no significant outliers. The differences (error) in standard deviation (σ) between QM and FF energies (calculated as $100 \times (\sigma_{QM} - \sigma_{FF})/\sigma_{QM}$) are 82, 85, 7, and 5%, respectively. This indicates that transferable FFs have significantly broader energy distribution, meaning that they visit high-energy configurations that would not have been visited in a QM-based simulation. This is also evident from the QM energies of the snapshots relative to the QM energy of the optimized molecule in Figure S3. While the average relative energies for Q-Force-based FFs (710 and 701 kJ/mol) are very much in line with the available thermal energy for ITIC at 298 K (673 kJ/mol), transferable FFs visit significantly higher-energy configurations (averages: 1036 and 1113 kJ/mol) that are not realistic. Overall, these results demonstrate the ability of Q-Force to sample much more accurate configurations for complex molecules during an MD simulation, compared to the transferable FFs that it has augmented, which is very promising for future applications in both materials science and biophysics.

Importantly, the improvement of the PES with Q-Force comes with no additional computational cost during the MD simulations; as the simulations with three different FFs were completed in comparable amounts of compute time. For the initial parametrization, the geometry optimization and the calculation of the QM Hessian took less than a day on 32 CPUs. Dihedral scans, due to the fragmentation scheme, took significantly less time. Overall, the whole procedure took less than 15 min of manual labor. This demonstrates that much larger systems are within the capabilities of the Q-Force procedure.

CONCLUSIONS

In this paper, we presented the Q-Force procedure that has the aim of augmenting existing transferable FFs with QM-based and molecule-specific bonded parameters and atomic charges. The automation of the procedure enables the derivation of the FF parameters in a transparent and reproducible manner and minimizes both (nonsystematic) human error and manual labor. The on-the-fly validation enables the careful evaluation of the quality of the generated FF. The fragmentation of large molecules allows the parametrization of large systems with significantly reduced computational cost, making the treatment of large molecules (>200 atoms) feasible. The procedure can be applied to any organic molecule not only in their ground state but also in their ionic and excited states; work in this direction is currently in progress in our group.

The accuracy of the Hessian fitting procedure was demonstrated with a set of small molecules where the Q-Force procedure obtained a significantly lower MAD in vibrational frequencies with respect to the QM reference compared to transferable FFs GROMOS/ATB and OPLS/LigParGen. For the same set of molecules, the fitting procedure for the flexible dihedral terms was shown to give high accuracy for all dihedrals. These improvements to the intramolecular properties came at no cost to the accuracy of thermodynamic properties, and in fact, the use of Q-Force with scaled CM5 charges has been shown to improve the thermodynamic properties for our test molecule set with respect to the parent transferable FF. For a large molecule (186 atoms), it was demonstrated that having accurate Hessian and dihedral profiles indeed leads to a significantly more accurate PES: The Q-Force-based PES has shown a significantly lower MAD compared to the transferable FFs and unlike those transferable FFs, it has not suffered from systematic underestimation of the QM energies. Such an underestimation will cause the transferable FFs to access the wrong configurations during the MD simulations, which might have significant consequences on the properties of interest.

The Q-Force toolkit, with an intuitive and easy-to-use interface, and together with relevant tutorials, is freely available on GitHub (https://github.com/selimsami/qforce). The functional form of the Q-Force FF bears no additional computational cost after the initial parametrization, making it available to any MD application that was previously treated with transferable FFs. Additionally, as the nonbonded interactions of the parent FF are preserved, compatibility can be expected with molecules already parametrized with the parent FF, such as complex biological molecules (e.g., DNA, RNA). Hence, Q-Force is easily applicable to various materials science and biophysics applications.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c00195.

Supporting figures, MD simulation and analysis details, and the Q-Force documentation (PDF)

Force fields from MD simulations (ZIP)

AUTHOR INFORMATION

Corresponding Author

Selim Sami − *Stratingh Institute for Chemistry, University of Groningen, 9747 AG Groningen, The Netherlands; Zernike Institute for Advanced Materials, University of Groningen, 9747 AG Groningen, The Netherlands;* orcid.org/0000-0002-4484-0322; Email: s.sami@rug.nl

Authors

Maximilian F.S.J Menger − *Zernike Institute for Advanced Materials, University of Groningen, 9747 AG Groningen, The Netherlands;* orcid.org/0000-0003-1442-9601

Shirin Faraji − *Zernike Institute for Advanced Materials, University of Groningen, 9747 AG Groningen, The Netherlands*

Ria Broer − *Zernike Institute for Advanced Materials, University of Groningen, 9747 AG Groningen, The Netherlands;* orcid.org/0000-0002-5437-9509

Remco W. A. Havenith − *Zernike Institute for Advanced Materials, University of Groningen, 9747 AG Groningen, The Netherlands; Stratingh Institute for Chemistry, University of Groningen, 9747 AG Groningen, The Netherlands; Department of Inorganic and Physical Chemistry, Ghent University, B-9000 Ghent, Belgium;* orcid.org/0000-0003-0038-6030

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.1c00195

Notes

The authors declare no competing financial interest.

https://doi.org/10.1021/acs.jctc.1c00195
J. Chem. Theory Comput. 2021, 17, 4946−4960

■ ACKNOWLEDGMENTS

The authors acknowledge Riccardo Alessandri and Siewert Jan Marrink for insights and helpful discussions. They thank SURFSara for giving access to the Dutch national supercomputer Cartesius. This work was sponsored by the Dutch Research Council (NWO) Exact and Natural Sciences for the use of supercomputer facilities. This work is part of the research programme of the Foundation of Fundamental Research on Matter (FOM), which is part of NWO. This is a publication of the FOM-focus Group "Next Generation Organic Photovoltaics", participating in the Dutch Institute for Fundamental Energy Research (DIFFER). M.F.S.J.M. and S.F. acknowledge Innovational Research Incentives Scheme Vidi 2017 with project number 016.Vidi.189.044, which is (partly) financed by NWO.

■ REFERENCES

(1) Mackerell, A. D., Jr. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.* **2004**, *25*, 1584−1604.
(2) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* **1999**, *103*, 3596−3607.
(3) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.* **2001**, *22*, 1205−1218.
(4) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656−1676.
(5) 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*, No. 843.
(6) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.
(7) 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*, 3499−3509.
(8) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(9) 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*, 3586−3616.
(10) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. 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*, 671−690.
(11) Huang, J.; MacKerell, A. D., Jr CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **2013**, *34*, 2135−2145.
(12) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(13) 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*, 1157−1174.
(14) Stroet, M.; Caron, B.; Visscher, K. M.; Geerke, D. P.; Malde, A. K.; Mark, A. E. Automated topology builder version 3.0: Prediction of solvation free enthalpies in water and hexane. *J. Chem. Theory Comput.* **2018**, *14*, 5834−5845.
(15) 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*, 3144−3154.
(16) 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*, 3155−3168.
(17) 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*, W331−W336.
(18) Kistenmacher, H.; Popkie, H.; Clementi, E.; Watts, R. O. Study of the structure of molecular complexes. VII. Effect of correlation energy corrections to the Hartree-Fock water-water potential on Monte Carlo simulations of liquid water. *J. Chem. Phys.* **1974**, *60*, 4455−4465.
(19) Lie, G. C.; Clementi, E. Study of the structure of molecular complexes. XII. Structure of liquid water obtained by Monte Carlo simulation with the Hartree−Fock potential corrected by inclusion of dispersion forces. *J. Chem. Phys.* **1975**, *62*, 2195−2199.
(20) Karlstroem, G.; Linse, P.; Wallqvist, A.; Joensson, B. Intermolecular potentials for the water-benzene and the benzene-benzene systems calculated in an ab initio SCFCI approximation. *J. Am. Chem. Soc.* **1983**, *105*, 3777−3782.
(21) Linse, P.; Engström, S.; Jönsson, B. Molecular dynamics simulation of liquid and solid benzene. *Chem. Phys. Lett.* **1985**, *115*, 95−100.
(22) Dasgupta, S.; Goddard, W. A. Hessian-biased force fields from combining theory and experiment. *J. Chem. Phys.* **1989**, *90*, 7207−7215.
(23) Dasgupta, S.; Yamasaki, T.; Goddard, W. A. The Hessian biased singular value decomposition method for optimization and analysis of force fields. *J. Chem. Phys.* **1996**, *104*, 2898−2920.
(24) Seminario, J. M. Calculation of intramolecular force fields from second-derivative tensors. *Int. J. Quantum Chem.* **1996**, *60*, 1271−1277.
(25) Cacelli, I.; Cinacchi, G.; Prampolini, G.; Tani, A. Computer simulation of solid and liquid benzene with an atomistic interaction potential derived from ab initio calculations. *J. Am. Chem. Soc.* **2004**, *126*, 14278−14286.
(26) Bizzarri, M.; Cacelli, I.; Prampolini, G.; Tani, A. Intermolecular force fields of large molecules by the fragmentation reconstruction method (FRM): Application to a nematic liquid crystal. *J. Phys. Chem. A* **2004**, *108*, 10336−10341.
(27) Cacelli, I.; Prampolini, G. Parametrization and validation of intramolecular force fields derived from DFT calculations. *J. Chem. Theory Comput.* **2007**, *3*, 1803−1817.
(28) Cacelli, I.; Cimoli, A.; Livotto, P. R.; Prampolini, G. An automated approach for the parameterization of accurate intermolecular force-fields: Pyridine as a case study. *J. Comput. Chem.* **2012**, *33*, 1055−1067.
(29) Barone, V.; Cacelli, I.; De Mitri, N.; Licari, D.; Monti, S.; Prampolini, G. Joyce and Ulysses: integrated and user-friendly tools for the parameterization of intramolecular force fields from quantum mechanical data. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3736−3751.
(30) Prampolini, G.; Livotto, P. R.; Cacelli, I. Accuracy of quantum mechanically derived force-fields parameterized from dispersion-corrected DFT data: The benzene dimer as a prototype for aromatic interactions. *J. Chem. Theory Comput.* **2015**, *11*, 5182−5196.
(31) Cerezo, J.; Prampolini, G.; Cacelli, I. Developing accurate intramolecular force fields for conjugated systems through explicit coupling terms. *Theor. Chem. Acc.* **2018**, *137*, No. 80.

https://doi.org/10.1021/acs.jctc.1c00195
J. Chem. Theory Comput. 2021, 17, 4946−4960

(32) Greff da Silveira, L.; Jacobs, M.; Prampolini, G.; Livotto, P. R.; Cacelli, I. Development and validation of quantum mechanically derived force-fields: Thermodynamic, structural, and vibrational properties of aromatic heterocycles. *J. Chem. Theory Comput.* **2018**, *14*, 4884−4900.
(33) 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*, 2307−2316.
(34) Vanduyfhuys, L.; Vandenbrande, S.; Verstraelen, T.; Schmid, R.; Waroquier, M.; Van Speybroeck, V. QuickFF: A program for a quick and easy derivation of force fields for metal-organic frameworks from ab initio input. *J. Comput. Chem.* **2015**, *36*, 1015−1027.
(35) Vanduyfhuys, L.; Vandenbrande, S.; Wieme, J.; Waroquier, M.; Verstraelen, T.; Van Speybroeck, V. Extension of the QuickFF force field protocol for an improved accuracy of structural, vibrational, mechanical and thermal properties of metal−organic frameworks. *J. Comput. Chem.* **2018**, *39*, 999−1011.
(36) Vandenbrande, S.; Waroquier, M.; Speybroeck, V. V.; Verstraelen, T. The monomer electron density force field (MEDFF): A physically inspired model for noncovalent interactions. *J. Chem. Theory Comput.* **2017**, *13*, 161−179.
(37) Zahariev, F.; De Silva, N.; Gordon, M. S.; Windus, T. L.; Dick-Perez, M. ParFit: A Python-based object-oriented program for fitting molecular mechanics parameters to ab Initio data. *J. Chem. Inf. Model.* **2017**, *57*, 391−396.
(38) 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*, 274−281.
(39) Horton, J. T.; Allen, A. E. A.; Dodda, L. S.; Cole, D. J. QUBEKit: Automating the derivation of force field parameters from quantum mechanics. *J. Chem. Inf. Model.* **2019**, *59*, 1366−1381.
(40) Grimme, S. A general quantum mechanically derived force field (QMDFF) for molecules and condensed phase simulations. *J. Chem. Theory Comput.* **2014**, *10*, 4497−4514.
(41) Spicher, S.; Grimme, S. Robust atomistic modeling of materials, organometallic, and biochemical systems. *Angew. Chem., Int. Ed.* **2020**, *59*, 15665−15673.
(42) Brunken, C.; Reiher, M. Self-parametrizing system-focused atomistic models. *J. Chem. Theory Comput.* **2020**, *16*, 1646−1665.
(43) Wang, R.; Ozhgibesov, M.; Hirao, H. Partial hessian fitting for determining force constant parameters in molecular mechanics. *J. Comput. Chem.* **2016**, *37*, 2349−2359.
(44) Wang, L.-P.; Chen, J.; Van Voorhis, T. Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data. *J. Chem. Theory Comput.* **2013**, *9*, 452−460.
(45) 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*, 2026−2047.
(46) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*−*2*, 19−25.
(47) Lagardère, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. Y.; Ponder, J. W.; Piquemal, J.-P. Tinker-HP: A massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chem. Sci.* **2018**, *9*, 956−972.
(48) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *WIREs Comput. Mol. Sci.* **2013**, *3*, 198−210.
(49) Caleman, C.; van Maaren, P. J.; Hong, M. Y.; Hub, J. S.; Costa, L. T.; van der Spoel, D. Force field benchmark of organic liquids: Density, enthalpy of vaporization, heat capacities, surface tension, isothermal compressibility, volumetric expansion coefficient, and dielectric constant. *J. Chem. Theory Comput.* **2012**, *8*, 61−74.
(50) Sami, S.; Alessandri, R.; Broer, R.; Havenith, R. W. A. How ethylene glycol chains enhance the dielectric constant of organic semiconductors: Molecular origin and frequency dependence. *ACS Appl. Mater. Interfaces* **2020**, *12*, 17783−17789.
(51) Alessandri, R.; Sami, S.; Barnoud, J.; de Vries, A. H.; Marrink, S. J.; Havenith, R. W. A. Resolving donor−acceptor interfaces and charge carrier energy levels of organic semiconductors with polar side chains. *Adv. Funct. Mater.* **2020**, *30*, No. 2004799.
(52) Liu, J.; van der Zee, B.; Alessandri, R.; Sami, S.; Dong, J.; Nugraha, M. I.; Barker, A. J.; Rousseva, S.; Qiu, L.; Qiu, X.; Klasen, N.; Chiechi, R. C.; Baran, D.; Caironi, M.; Anthopoulos, T. D.; Portale, G.; Havenith, R. W. A.; Marrink, S. J.; Hummelen, J. C.; Koster, L. J. A. N-type organic thermoelectrics: Demonstration of ZT > 0.3. *Nat. Commun.* **2020**, *11*, No. 5694.
(53) Liu, J.; Ye, G.; Potgieser, H. G. O.; Koopmans, M.; Sami, S.; Nugraha, M. I.; Villalva, D. R.; Sun, H.; Dong, J.; Yang, X.; Qiu, X.; Yao, C.; Portale, G.; Fabiano, S.; Anthopoulos, T. D.; Baran, D.; Havenith, R. W. A.; Chiechi, R. C.; Koster, L. J. A. Amphipathic side chain of a conjugated polymer optimizes dopant location toward efficient n-type organic thermoelectrics. *Adv. Mater.* **2021**, *33*, No. 2006694.
(54) Andreussi, O.; Prandi, I. G.; Campetella, M.; Prampolini, G.; Mennucci, B. Classical force fields tailored for QM applications: Is it really a feasible strategy? *J. Chem. Theory Comput.* **2017**, *13*, 4636−4648.
(55) Claridge, K.; Troisi, A. Developing consistent molecular dynamics force fields for biological chromophores via force matching. *J. Phys. Chem. B* **2019**, *123*, 428−438.
(56) Urey, H. C.; Bradley, C. A. The vibrations of pentatonic tetrahedral molecules. *Phys. Rev.* **1931**, *38*, 1969−1978.
(57) Pettitt, B. M.; Karplus, M. Role of electrostatics in the structure, energy and dynamics of biomolecules: A model study of N-methylalanylacetamide. *J. Am. Chem. Soc.* **1985**, *107*, 1166−1173.
(58) Ryckaert, J.-P.; Bellemans, A. Molecular dynamics of liquid alkanes. *Faraday Discuss. Chem. Soc.* **1978**, *66*, 95−106.
(59) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge model 5: An extension of Hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases. *J. Chem. Theory Comput.* **2012**, *8*, 527−541.
(60) Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta* **1977**, *44*, 129−138.
(61) Ritchie, J. P. Electron density distribution analysis for nitromethane, nitromethide, and nitramide. *J. Am. Chem. Soc.* **1985**, *107*, 1829−1837.
(62) Ritchie, J. P.; Bachrach, S. M. Some methods and applications of electron density distribution analysis. *J. Comput. Chem.* **1987**, *8*, 499−509.
(63) Dodda, L. S.; Vilseck, J. Z.; Cutrona, K. J.; Jorgensen, W. L. Evaluation of CM5 charges for nonaqueous condensed-phase modeling. *J. Chem. Theory Comput.* **2015**, *11*, 4273−4282.
(64) Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation of CM5 charges for condensed-phase modeling. *J. Chem. Theory Comput.* **2014**, *10*, 2802−2812.
(65) Hagberg, A. A.; Schult, D. A.; Swart, P. J. In *Exploring Network Structure, Dynamics, and Function using NetworkX*, Proceedings of the 7th Python in Science Conference, Pasadena, CA, 2008; pp 11−15.
(66) Wiberg, K. Application of the Pople-Santry-Segal CNDO method to the cyclopropylcarbinyl and cyclobutyl cation and to bicyclobutane. *Tetrahedron* **1968**, *24*, 1083−1096.
(67) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; Vander- Plas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; Vijaykumar, A.; Bardelli, A. P.; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.; Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. V.; Olivetti, E.; Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.; Ingold, G.-L.; Allen,

https://doi.org/10.1021/acs.jctc.1c00195
J. Chem. Theory Comput. 2021, 17, 4946−4960

G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Slavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; de Miranda Cardoso, J.; Reimer, J.; Harrington, J.; Rodríguez, J. L. C.; Nunez-Iglesias, J.; Kuczynski, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmerer, M.; Bolingbroke, M.; Tartre, M.; Pak, M.; Smith, N. J.; Nowaczyk, N.; Shebanov, N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.; Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261−272.

(68) Cioslowski, J.; Scott, A. P.; Radom, L. Catastrophes, bifurcations and hysteretic loops in torsional potentials of internal rotations in molecules. *Mol. Phys.* **1997**, *91*, 413−420.

(69) Qiu, Y.; Smith, D. G. A.; Stern, C. D.; Feng, M.; Jang, H.; Wang, L. P. Driving torsion scans with wavefront propagation. *J. Chem. Phys.* **2020**, *152*, No. 244116.

(70) MacKerell, A. D.; Feig, M.; Brooks, C. L. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **2004**, *126*, 698−699.

(71) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257−3273.

(72) Glendening, E. D.; Badenhoop, J. K.; Reed, A. E.; Carpenter, J. E.; Bohmann, J. A.; Morales, C. M.; Karafiloglou, P.; Landis, C. R.; Weinhold, F. *NBO 7.0*; Theoretical Chemistry Institute, University of Wisconsin: Madison, 2018.

(73) Humphrey, W.; Dalke, A.; Schulten, K. VMD − Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.

(74) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.

(75) Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-consistent molecular-orbital methods. I. Use of Gaussian expansions of Slater-type atomic orbitals. *J. Chem. Phys.* **1969**, *51*, 2657−2664.

(76) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.

(77) Mewes, S. A.; Plasser, F.; Dreuw, A. Universal exciton size in organic polymers is determined by nonlocal orbital exchange in time-dependent density functional theory. *J. Phys. Chem. Lett.* **2017**, *8*, 1205−1210.

(78) Mewes, S. A.; Plasser, F.; Dreuw, A. Communication: Exciton analysis in time-dependent density functional theory: How functionals shape excited-state characters. *J. Chem. Phys.* **2015**, *143*, No. 171101.

(79) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.. *Gaussian 09*. Gaussian Inc.: Wallingford, CT, 2009.

(80) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J.. *Gaussian 16*. Gaussian Inc.: Wallingford, CT, 2016.

(81) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuś, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J., III; H L W; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M., Jr.; R A D; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W.; Harbach, P. H.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K.; Chipman, D. M.; Cramer, C. J., III; W A G; Gordon, M. S.; Hehre, W. J.; Klamt, A., III; H F S; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Voorhis, T. V.; Herbert, J. M.; Krylov, A. I.; Gill, P. M.; Head-Gordon, M. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015**, *113*, 184−215.

(82) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, 1−17.

(83) Swails, J.; Hernandez, C.; Mobley, D.; Nguyen, H.; Wang, L.; Janowski, P. ParmEd: Cross-program Parameter and Topology File Editor and Molecular MechanicalSimulator Engine. https://github.com/ParmEd/ParmEd.

(84) National Institute of Standards and Technology, NIST/EPA Gas-Phase Infrared Database. https://webbook.nist.gov/ (accessed May 11, 2021).

(85) Lin, Y.; Wang, J.; Zhang, Z.-G.; Bai, H.; Li, Y.; Zhu, D.; Zhan, X. An electron acceptor challenging fullerenes for efficient polymer solar cells. *Adv. Mater.* **2015**, *27*, 1170−1174.

(86) Zheng, Z.; Awartani, O. M.; Gautam, B.; Liu, D.; Qin, Y.; Li, W.; Bataller, A.; Gundogdu, K.; Ade, H.; Hou, J. Efficient charge transfer and fine-tuned energy level alignment in a THF-processed fullerene-free organic solar cell with 11.3% efficiency. *Adv. Mater.* **2017**, *29*, No. 1604241.

(87) Yang, Y.; Zhang, Z.-G.; Bin, H.; Chen, S.; Gao, L.; Xue, L.; Yang, C.; Li, Y. Side-chain isomerization on an n-type organic semiconductor ITIC acceptor makes 11.77% high efficiency polymer solar cells. *J. Am. Chem. Soc.* **2016**, *138*, 15011−15018.

https://doi.org/10.1021/acs.jctc.1c00195
J. Chem. Theory Comput. 2021, 17, 4946−4960

(88) Wadsworth, A.; Moser, M.; Marks, A.; Little, M. S.; Gasparini, N.; Brabec, C. J.; Baran, D.; McCulloch, I. Critical review of the molecular design progress in non-fullerene electron acceptors towards commercially viable organic solar cells. *Chem. Soc. Rev.* **2019**, *48*, 1596−1625.

https://doi.org/10.1021/acs.jctc.1c00195
J. Chem. Theory Comput. 2021, 17, 4946−4960