JCTC Journal of Chemical Theory and Computation

Article

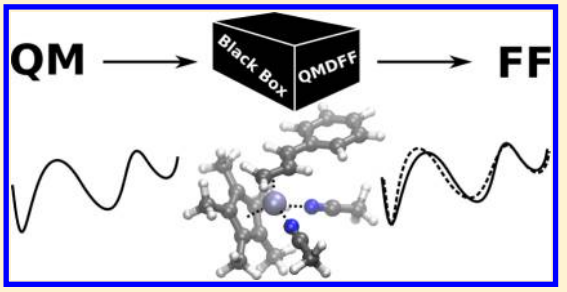
pubs.acs.org/JCTC

A General Quantum Mechanically Derived Force Field (QMDFF) for Molecules and Condensed Phase Simulations

Stefan Grimme*

Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie der Universität Bonn, Beringstr. 4, D-53115 Bonn, Germany

ABSTRACT: A black-box type procedure is presented for the generation of molecule-specific, classical potential energy functions (force-field, FF) solely from quantum mechanically (QM) computed input data. The approach can treat covalently bound molecules and noncovalent complexes with almost arbitrary structure. The necessary QM information consists of the equilibrium structure and the corresponding Hessian matrix, atomic partial charges, and covalent bond orders. The FF fit is performed automatically without any further input and yields a specific (nontransferable) potential which very closely resembles the QM reference potential near the equilibrium. The resulting atomistic, fully flexible FF is anharmonic and allows smooth dissociation of covalent bonds into atoms. A newly proposed force-constant−bond-energy relation with little empiricism provides reasonably accurate (about 5−10% error) atomization energies for almost arbitrary diatomic and polyatomic molecules. Intra- and intermolecular noncovalent interactions are treated by using well established and accurate D3 dispersion coefficients, CM5 charges from small basis set QM calculations, and a new interatomic repulsion potential. Particular attention has been paid to the construction of the torsion potentials which are partially obtained from automatic QM-tight-binding calculations for model systems. Detailed benchmarks are presented for conformational energies, atomization energies, vibrational frequencies, gas phase structures of organic molecules, and transition metal complexes. Comparisons to results from standard FF or semiempirical methods reveal very good accuracy of the new potential. While further studies are necessary to validate the approach, the initial results suggest QMDFF as a routine tool for the computation of a wide range of properties and systems (e.g., for molecular dynamics of isolated molecules, explicit solvation, self-solvation (melting) or even for molecular crystals) in particular when standard parametrizations are unavailable.



1. INTRODUCTION

For many molecular and condensed phase properties, long-time molecular dynamics (MD) simulations for large particle numbers are required. Although these could in principle be conducted at an *ab initio* or first-principles level (Born−Oppenheimer or Carr−Parrnello MD[1]), i.e., computing the forces on-the-fly from electronic structure theory, such treatments are often computationally too demanding. Therefore, classical potential functions, called molecular mechanics (MM) or force fields (FFs), which are based on interatomic distance and angle dependent terms have been used for decades to approximate molecular potential energy surfaces (PES).[2−5] If both quantum mechanical (QM) and MM potentials are combined (QM/MM), chemical reactions in proteins can be modeled as well.[6,7] One of the problems in such multiscale (also called multiresolution) approaches (for which the Nobel Prize in Chemistry 2013 was awarded[8]) is that a hierarchy of very different methods up to nonatomistic, coarse-grained FF have to be combined and connected. New ideas for the information flow from the QM level, which here mostly refers to dispersion-corrected Kohn−Sham density functional theory (DFT-D3[9,10]), into a rather general FF are presented in this work.

The results of MD simulations are sensitive to the functional form of the FF and the parametrization of the individual terms involved (for recent discussions and examples, see refs 5, 11−13). Conventional FF assume transferability of the potentials between different molecules within a class of chemically similar compounds (e.g., organic molecules, peptides, nucleobases). Whether the parametrization is based on experimental data or employs QM computed properties, such FFs describe different molecules within the specific class in an averaged and often rather crude way. The problem is alleviated by introducing various atom types for the same element (or even bond types) which complicates the parametrization process (for recent attempts to QM based, partially automatic FF generation, and fitting procedures, see refs 14−19). For instance, the results of biomolecular simulations depend strongly on the description of torsion terms (conformational sampling) and the level of electrostatic interactions (e.g., atomic partial charges, multipole-moments), and a vast variety of biomolecular FFs exists for different compound classes or functional groups in standard packages like GROMACS,[20] AMBER,[21] CHARMM,[22] or TINKER.[23]

Even more important for the practical use of FFs is that only a very small part of the possible chemical space of compounds is covered by existing parametrizations. Already for relatively

Received: July 3, 2014
Published: August 25, 2014

   dx.doi.org/10.1021/ct500573f | J. Chem. Theory Comput. 2014, 10, 4497−4514

simple organometallic or transition metal complexes, practically no accurate potentials exist (for some special purpose FFs in this area, see e.g. refs 16, 24−26). The same holds for more unusual but chemically very interesting main group systems like frustrated Lewis pairs (FLP), which have been investigated in our group for quite some time.[27−29] In this context, the universal force-field (UFF) approach of Goddard et al.[30] should be mentioned, which is at least in principle able to deal with systems of arbitrary structure. However, its accuracy is insufficient for many practical purposes and although implemented in big software packages like TURBOMOLE,[31,32] the method is not in widespread use. Consequently, the dynamical and structural properties of many interesting systems, such as metalo-enzymes, with extremely high biological relevance cannot be sampled with reasonable accuracy. The same holds for electronically excited states or other geometrically or electronically more complicated situations. Addressing this situation is the major aim of this work.

The current proposal aims at an accurate and specific FF for a single molecule in a particular electronic state, which is in stark contrast to most existing FF approaches. The FF is constructed using QM data from a single equilibrium structure; i.e., it is made specifically for one conformation. However, the description remains reasonably accurate also away from the equilibrium of the PES by construction of the potentials and extensive use of precomputed QM data. Thus, the FF not only has interpolation but also good extrapolation capabilities. As QM input, only the equilibrium structure, the Hessian matrix (second geometrical energy derivatives), the atomic partial charges, and the covalent bond orders are required. With these data, a molecule-specific FF is automatically generated without any further user input. The procedure/FF has the following characteristics:

1. It can straightforwardly be applied to any molecule with a reasonable covalent structure consisting of any elements up to $Z = 86$ and their noncovalent complexes or condensed phases. Molecules without clear directional bonding and metallic character (e.g., clusters) can be treated in principle, but such cases have not been tested extensively yet.

2. Inter- and intramolecular interactions are treated consistently. The FF can be generated for a solute and a solvent and used for explicit solvation studies, but with the same FF a solute can also be "self-solvated" (e.g., in a melt) or even used to describe a molecular crystal. The accuracy for intermolecular (noncovalent) interactions is competitive with semiempirical QM methods and close to that of DFT-D3 for uncharged complexes.

3. All potentials are fully anharmonic. In regions of the PES near the equilibrium, the FF closely reproduces the QM input PES.

4. All bonds properly dissociate into separated atoms. Bond and atomization energies are accurate to about 5−10% of D_e. This allows investigations of the thermal stability of molecules or their behavior under mechanical stress. As a byproduct of the FF generation, bond strengths for individual pairwise interactions are obtained which reasonably reflect common chemical intuition. This property is related to the recently proposed JEDI analysis method.[33] The FF is dissociative but not reactive like ReaxFF;[34] i.e., new bonds in a molecule can *not* be formed.

5. The FF can be generated for any physically reasonable QM input, and the resulting potential reflects and resembles the quality of the input PES. The latter does not strictly hold for most torsional potentials, which are generated independently from automatically performed tight-binding (TB) based QM calculations for model structures.

6. The FF is based on the idea of minimal empirical parametrization; i.e., in total only 47 global (and mostly very weakly coupled) parameters were adjusted to describe any combination of the first 86 elements in molecules. This number is similar to what is required for a chemically global description in modern density functionals.[35]

7. The computational bottleneck of the procedure is the computation of the QM input data (basically the Hessian) and not the FF generation. Because of the use of sophisticated, asymptotically correct bonding and nonbonding potentials, the FF is computationally more expensive than standard forms[36]

Because the potential is completely determined and derived by QM input (and with the aid of a few global and fixed empirical parameters), the force field will be termed Quantum Mechanically Derived Force Field (QMDFF). Effective experimentally derived parameters from, e.g., bulk properties are avoided throughout. The general idea is that QMDFF as closely as possible reproduces the PES of a good-quality QM method for more or less arbitrary molecules. This construction strategy as well as part of the potentials used in QMDFF are to the best of our knowledge new in this research field. Related work has been presented in the MOF-FF of Schmid et al.[16] where, however, a compound class FF is constructed from molecular model systems.

The main areas of application, which we currently have in mind, are studies of conformational flexibility of large supramolecular systems and inorganic coordination compounds, as well as investigations of protein−ligand interactions, explicit solvation, mechano-chemistry, and thermal stability. It should be again emphasized here that as a unique characteristic QMDFF allows straightforward treatment of large metal complexes or coordination aggregates. Furthermore, the FF can be made specifically also for molecular ions as generated by electron impact, and hence its application in the framework of our QCEIMS[37] procedure for computing mass spectra can easily be implemented.

Before the theory and computational details of QMDFF are discussed below, the main downsides of the approach are mentioned already at this point. First of all, the method cannot be applied out of the box to a target molecule as with an ordinary FF because, at first, the QM input data for the equilibrium structure of interest have to be generated. Second, because it is extremely difficult to reflect and fulfill the permutational (nuclear) symmetries of the PES for arbitrary molecules by conventional, atom-wise assigned functions, QMDFF is slightly symmetry breaking in some situations (mainly internal fragment rotations). This means that symmetry equivalent structures do not necessarily have the same energy. However, in the cases observed so far, the errors were found to be much smaller than the inherent limitations of such FF. For ways to circumvent this problem in the special case of only a few elements (atom types) in the system, compare the neural network potentials of Behler.[38] Lastly, the accuracy of QMDFF is biased toward nuclear coordinates close to the input structure where it is clearly superior to any existing FF while a conventional FF is equally good (or bad) in a larger PES region. Whether this is a pro or a con in an actual application has to be judged by the user.

The basic idea of the current approach to use the QM computed molecular Hessian, i.e., the $3N \times 3N$ (N being the

number of atoms in the molecule) second energy derivatives $H_{ij} = \delta^2 E/\delta X_i \delta X_j$ where X are atomic Cartesian coordinates, for FF generation is not new. To our knowledge, the first more general attempt in this direction was made by Goddard et al.[39] Recently, a general scheme for reactive FF generation termed GARFfield[19] was proposed by the same group. Related to this and the present work is the older approach of Hagler et al.[40,41] in which *ab initio* data (gradient and Hessian) obtained for distorted, nonequilibrium geometries were employed to parametrize a standard (transferable) FF.

After an outline of the theory, several key properties of QMDFF are evaluated by benchmarking to common molecular test sets. However, this evaluation is necessarily incomplete because the possible fields of application for QMDFF are rather broad, and it is impossible to cover even a small part of the possible chemical space in one paper. In particular, we test noncovalent interaction energies (S22[42] and an extended version of S12L[43]), conformational energies (e.g., alkanes, peptides, and sugars taken from our GMTKN30 database[44]), atomization energies (G2 set[45] and 18 amino acids[46]), gas phase structures of organic molecules (34 rotational constants[47] and a few larger drug-like molecules), and 10 common transition metal complexes. Comparisons are made with accurate reference data (both, of theoretical and experimental origin), as well as computed data from standard FF or simple semiempirical methods like PM6-DH2[48,49] or DFTB3-D3.[50−52] The performance of QMDFF for bulk properties of liquids and molecular crystals and the treatment of explicit solvation in a QM/MM framework is beyond the scope of the present work and will be published separately.

2. THEORY

2.1. General. The QMDFF total energy E consists of three parts

$$E = E_{e,\mathrm{QM}} + E_{\mathrm{intra}} + E_{\mathrm{NCI}} \tag{1}$$

where $E_{e,\mathrm{QM}}$ is the QM energy of the reference equilibrium (index e) structure, E_{intra} refers to the bonded FF energy (often called strain energy), and E_{NCI} describes the intra- and intermolecular noncovalent interactions (NCI, also called nonbonding terms). In the following, atomic units are used throughout. The $E_{e,\mathrm{QM}}$ is a constant for a given FF and hence does not contribute to the force or other derived properties. However, if the energies of different chemical species, for which different FFs have been generated, are compared, this quantity must be taken into account.

The E_{NCI} terms are applied to all atom pairs which either belong to different molecules or are intramolecular if the two atoms are topologically separated by more than four covalent bonds. For any intermolecular NCI, the same FF is applied; i.e., for the NCI part transferability is assumed. The covalent bonding pattern is automatically determined from a combination of interatomic distance criteria with standard covalent atomic radii[53] and Wiberg−Mayer covalent bond orders (BO).[54−56] The functional form for E_{NCI} is described below.

The E_{intra} part contains a sum of stretch and compression (str), bending angle (bend), and other angle (tors, inv) dependent terms according to

$$E_{\mathrm{intra}} = \sum_{\mathrm{bonds}} V_{\mathrm{str}}^{12} + \sum_{1,3} V_{\mathrm{str}}^{13} + \sum_{\mathrm{bend}} V_{\mathrm{bend}} + \sum_{\mathrm{torsion}} V_{\mathrm{tors}} + \sum_{\mathrm{inversion}} V_{\mathrm{inv}} \tag{2}$$

where the sums run over all bonds (one covalent bond between the atoms), 1,3-interactions (two bonds in between), bending (three-body), torsion (four-body), and inversion angles (four-body, also termed out-of-plane) and V's are the corresponding potential functions. In contrast to a standard FF, the equilibrium structure ($\boldsymbol{R}_e$) of the molecule under consideration is already known and hence definite equilibrium bond distances r_e, bending angles θ_e, and torsion or inversion angles ψ_e are used. In QMDFF the V terms in eq 2 always depend on differences or ratios of the actual value for r, θ, or ψ and the corresponding equilibrium value. Because the potentials vanish by construction for $r = r_e$, $\theta = \theta_e$, and $\psi = \psi_e$, E_{intra} and the corresponding forces are zero at the QM input structure, QMDFF in this preliminary form exactly reproduces the input structure. Exceptions to this important property are discussed below.

2.2. Bond and Stretch Terms. For the bond and 1,3-stretching terms, a novel generalization of the Lennard-Jones (GLJ) potential is proposed

$$V_{\mathrm{str}} = k_{\mathrm{str,AB}}\left(1 + \left(\frac{r_{e,\mathrm{AB}}}{r_{\mathrm{AB}}}\right)^{a} - 2\left(\frac{r_{e,\mathrm{AB}}}{r_{\mathrm{AB}}}\right)^{a/2}\right) \tag{3}$$

which allows smooth dissociation of the bond between atoms A and B into the separated atoms. Here, r_{AB} is the interatomic distance, and the force constant (FC) k_{str} is determined from the fit to the QM computed molecular Hessian (see below). The exponent a is determined from

$$a = k_a(\mathrm{A})k_a(\mathrm{B}) + k_{\mathrm{EN}}|\Delta\mathrm{EN(AB)}|^2 \tag{4}$$

where ΔEN(AB) is the electronegativity difference of atoms A and B scaled by the empirical parameter k_{EN}, and k_a represents element specific parameters. They have been obtained for the elements H, B, C, N, O, and F from a fit to G2 atomization energies. For the elements Al−Cl, the values of the corresponding second row elements B−F are taken and shifted by a constant value k_{a2}. For all other elements, the k_a values depend only on the row of the element in the periodic table, and the corresponding values have been fitted to the D_e reference values for diatomics or small molecules. For rows larger than three, the parameters have been obtained by downward extrapolation. Larger values of a correspond in general to weaker (softer) bonds, and the general bond strengthening tendency with increasing bond polarity is reflected in the value of the parameter $k_{\mathrm{EN}} < 0$. In the above equation, the Allen electronegativity scale[57] is employed, but taking Mulliken's values yields very similar results.

Equation 3 is also used for 1,3-stretch terms except that the exponent is $a = k_{a13} + k_{b13}k_a(\mathrm{A})k_a(\mathrm{B})$, where k_{a13} and k_{b13} are global parameters. If one of the two atoms in the 1,3-interaction belongs to a ring, a constant value $k_{13,r}$ is added to a.

Equations 3 and 4 together with the fit to the Hessian define a sophisticated FC-bond energy relationship of which many exist in the literature (for a detailed discussion and a review of previous work, see e.g. ref 58). A Morse-type function as an alternative to eq 3 has been tested but leads to severe instabilities in the FC fitting procedure. For the simple case of a

diatomic molecule, eq 3 allows the computation of the dissociation energy D_e from the knowledge of the equilibrium stretch FC and the chemical composition. One of the striking new results of this work is that this relation works very well across the whole periodic table from weak metallic interactions up to strong bonds like N_2.

Figure 1 shows examples of computed potential curves together with reference D_e values for a few diatomic molecules

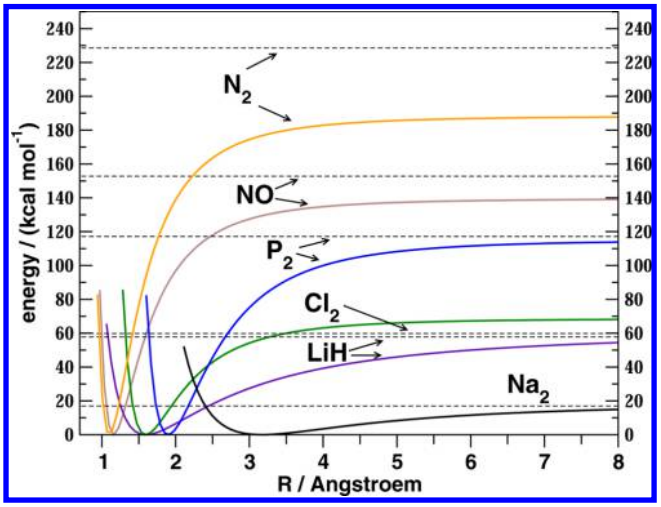


Figure 1. Computed QMDFF potential energy curves and reference dissociation energies (from ref 45, asymptotic value indicated by the dashed line) for six diatomic molecules.

which clearly demonstrate the accuracy of this relation. Note that it works without any specific adjustments relatively well for weak metallic bonds (Na_2, D_e = 17 kcal mol^{-1}), ionic ones as in LiH, third-row species as in P_2 or Cl_2, and very strong multiple bonds as in N_2 (D_e = 229 kcal mol^{-1}), although for this very challenging system, the absolute deviation is large, about 40 kcal mol^{-1} (which, however, corresponds to merely 17% relative error). Further discussion of the performance for atomization energies of polyatomic molecules is presented in section 4.3. It is important to emphasize that the stretch potentials in QMDFF (and the 1,3-coupling terms) are anharmonic and in addition allow reasonable dissociation into free atoms. At intermediate distances, the stretch potentials are, however, not very accurate because of their polynomial (instead of exponential) character. Note that the final relations in eqs 3 and 4 require only 16 global parameters for the first 86 elements of the periodic table which allow the computation of the atomization energy for molecules of arbitrary composition.

2.3. Bending Terms. For the bending terms, which involve three atoms ABC, where A is the central atom, the following potential is employed for equilibrium angles $\theta_{e,\mathrm{ABC}}$ close to linearity ($\theta_e \approx \pi$)

$$V_{\mathrm{bend}}(\mathrm{ABC}, \theta) = k_{\mathrm{bnd}} f_{\mathrm{dmp}} (\theta_{e,\mathrm{ABC}} - \theta)^2 \tag{5}$$

while in all other cases a double-minimum form that allows inversion

$$V_{\mathrm{bend}}(\mathrm{ABC}, \theta) = k_{\mathrm{bnd}} f_{\mathrm{dmp}} (\cos\theta_{e,\mathrm{ABC}} - \cos\theta)^2 \tag{6}$$

is applied. Here, k_{bnd} is the bending FC and $f_{\mathrm{dmp}} = f_{\mathrm{dmp}}(\mathrm{AB})$ $f_{\mathrm{dmp}}(\mathrm{AC})$ is the product of the distance dependent damping functions (modified from ref 59) for AB and AC which is given by

$$f_{\mathrm{dmp}}(\mathrm{AB}, r) = \frac{1}{1 + k_{\mathrm{dmp}}\left(\frac{r_{\mathrm{AB}}}{r_{\mathrm{cov,AB}}}\right)^4} \tag{7}$$

with the covalent distance for the pair, $r_{\mathrm{cov,AB}}$, and k_{dmp} is a global parameter. Its value is chosen such that the potential vanishes approximately at 3 times the typical covalent distance. The same damping scheme is applied to torsion and inversion terms so that the remaining potential in QMDFF for infinitely separated atoms solely results from eq 3 and no angle dependent terms remain.

2.4. Torsion and Inversion Terms. The torsion potential is the most complicated but also the most important part of a general FF (for recent parametrization attempts see, e.g., ref 60). For rotation around a bond AB with atom C connected to A and atom D to B, we have

$$\begin{aligned} V_{\mathrm{tors}}(\mathrm{CABD}, \psi) = f_{\mathrm{dmp}} \sum_n k^n_{\mathrm{tors}} (f_{\mathrm{chiral}}(\psi) \\ [1 + \cos(n(\psi - \psi_e) + \pi)] \\ + (1 - f_{\mathrm{chiral}}(\psi)) \\ [1 + \cos(n(\psi + \psi_e - 2\pi) + \pi)]) \end{aligned} \tag{8}$$

where f_{dmp} is the triple damping product in analogy with the bending term (for CA, AB, BD), ψ_e is the equilibrium torsion angle, and the linear combination using f_{chiral}

$$f_{\mathrm{chiral}}(\psi) = \frac{1}{2}[1 - \mathrm{erf}(\psi - \pi)] \tag{9}$$

ensures the correct mirror symmetry of the potential for molecular enantiomer pairs M and M′ for which $\psi_e(\mathrm{M}') = 2\pi - \psi_e(\mathrm{M})$ holds. This complication appears because as opposed to standard FF where ψ_e is usually restricted to "nonchiral" values, the equilibrium dihedral angles ψ_e in QMDFF can take any value at the QM equilibrium structure.

For nonrotatable bonds (in rings or for BO(AB) > 1.3), only one term in the sum in eq 8 with the appropriate multiplicity number n (determined by the atoms AB and their coordination numbers) is taken into account and k^n_{tors} is obtained from the fit to the Hessian. For example this procedure applies to the CC bond torsion in ethene or a benzene ring which have $n = 2$ and yields qualitatively reasonable torsion barriers without empirical modifications (e.g., 41 kcal mol^{-1} instead of about 60 kcal mol^{-1} for ethene).

The FC for rotatable bonds can *not* be obtained from the Hessian because the information about the rotational barrier cannot be derived from the equilibrium structure, and furthermore, in the general case, the appropriate value of n is unknown. This fundamental problem is to the best of the authors' knowledge treated here for the first time by the following procedure: First, a model geometry containing the CABD fragment with the first shell of substituents on atoms C and D is automatically constructed from the molecular coordinates. Dangling covalent bonds are saturated with hydrogen atoms. A full 2π rotational potential energy curve is computed by a minimal valence basis set TB (also called extended Hückel) Hamiltonian. This method is known to provide rather accurate torsional potentials for a wide variety of compounds.[61,62] The total energy is computed in the standard way as a sum over the occupied orbital energies, i.e., $E_{\mathrm{TB}} = 2\sum_i^{\mathrm{occ}} \varepsilon_i$. To avoid double-counting effects, intersubstituent overlap integrals are neglected or (for 1,4-interactions) scaled

dx.doi.org/10.1021/ct500573f | J. Chem. Theory Comput. 2014, 10, 4497−4514

by a parameter k_{ovlp}. Four terms in the Fourier sum with $n = 1-4$ are included, and k^n_{tors} are obtained from a fit to the TB potential curve. If the fit is too bad, or $r_{e,AB}$ >2.2 Å, or the TB barrier is >25 kcal mol^{-1}, this potential is discarded and replaced by the Hessian fit procedure with one n-dependent term in eq 8. In the following, the two different torsion potentials (Hessian fit or TB derived) are distinguished by using V_{tors}' for the TB derived one. Note that this QM based procedure is absolutely essential to describe reasonably the conformational flexibility of biomolecular structures with many single bonds. This will be discussed below for conformational benchmarks.

The inversion type potential V_{inv} for 3-fold coordinated atoms is constructed in the same manner as the bending angles by using the out-of-plane angle ϕ instead of θ in eqs 5 and 6. This choice yields physically correct double-minimum potentials and proper inversion barriers for small XH_3 (X=N,O$^+$,P) model systems when the corresponding FC k_{inv} are fitted to the Hessian. Special treatments of the various angle dependent terms apply to highly coordinated (mostly metal) atoms, topologies involving linear (or close to linear) units, and rings and are not discussed in detail here.

2.5. Noncovalent (nonbonding) Terms. For the noncovalent inter- and intramolecular energy E_{NCI}, a modified version of a recently developed standard intermolecular FF is used.[63] It consists of Pauli-repulsion (rep), electrostatic (ES), hydrogen-bonding (hbnd), halogen-bonding (xbnd), London dispersion (disp), and (optionally) polarization (pol) terms

$$E_{inter} = E_{rep} + E_{ES} + E_{hbnd} + E_{xbnd} + E_{disp} + E_{pol} \quad (10)$$

For the London dispersion energy, the standard D3 scheme[64] with Becke–Johnson (BJ) rational damping[65,66]

$$E_{disp} = -\sum_{AB} \varepsilon_{rep/disp}(AB)\left[\frac{C_6^{AB}}{r_{AB}^6 + f(R_{AB}^0)^6} + s_8\frac{C_8^{AB}}{r_{AB}^8 + f(R_{AB}^0)^8}\right] \quad (11)$$

with three fitted damping and scaling parameters a_1, a_2, and s_8 is employed. To the best of our knowledge, this study represents the first use of the D3 dispersion scheme in the framework of a general intra- and intermolecular FF. Here, $f(R_{AB}^0) = a_1R_{AB}^0 + a_2$ is the BJ damping function with appropriate radii R_{AB}^0[66] and $\varepsilon_{rep/disp}$ is a topological screening parameter which depends on the number of covalent bonds between atoms A and B (see below). The C_6 coefficients (and derived C_8) in the D3 method are obtained from a modified form of the Casimir–Polder relation[64] where the required frequency-dependent dipole polarizability is computed nonempirically by time-dependent DFT.[67] To reduce the computational effort in particular for large condensed phase systems, the geometry dependence of the C_6 and C_8 dispersion coefficients is not considered here, and the pair-specific values computed at the equilibrium input structure are taken. Although this slightly diminishes the accuracy for dissociating systems, this approach seems appropriate for an inherently not very accurate FF. It is also consistent with the treatment of the electrostatics which employs fixed charges computed for the input structure. In general, the D3 dispersion energy has a good 5–10% accuracy[10,64,68] mainly because atom-in-molecule, pair-specific, *ab initio* computed, isolated molecule dispersion coefficients are employed. Standard FFs usually work with atom-type specific values and simple pair combination rules and derives dispersion terms often from condensed phase data. Furthermore, eq 11 shows an improved (asymptotically correct) short-range behavior and is consistent with dispersion-corrected DFT computed long-range QM interactions which are present in the input data. For a more general discussion on the importance of London dispersion effects for chemical bonding, see e.g. refs 10 and 69.

For the Pauli repulsion, we employ a new atom pairwise formula

$$E_{rep} = \sum_{AB} \varepsilon_{rep/disp}(AB)\frac{Z_{eff}(A)Z_{eff}(B)}{r_{AB}} e^{-\beta_{rep}r_{AB}/R_{0,D3}^{3/2}} \quad (12)$$

where Z^{eff} represents effective nuclear charges, $R_{0,D3}$ represents the standard D3 pair cutoff radii,[64] and β_{rep} is a global fit parameter. The Z_{eff} values are obtained from atomic valence electron numbers Z_{val} which are scaled by parameter k_Z depending on the row of the atom in the periodic table. Special values are used for group 1 and 2 elements while values for rows larger than three are obtained by extrapolation. Nine global empirical parameters k_Z for the elements up to $Z = 86$ are required.

For intermolecular solute–solvent terms, we have also tested a polarization contribution according to the standard Drude model (see e.g. ref 70). For the atomic dipole–dipole polarizability α, we use the empirical relationship

$$\alpha^{Drude}(A) = \sqrt{C_6(AA)} \quad (13)$$

where $C_6(AA)$ is the D3 self-dispersion coefficient for atom A. In the Drude model, the floating charge q_D is obtained by

$$q_D(A) = \sqrt{k_D\alpha^{Drude}(A)} \quad (14)$$

with the Drude spring constant $k_D = 0.3$. Note that E_{pol} is optionally computed for intermolecular interactions but is not used for the noncovalent part of the intramolecular FF.

For atom pairs involving the electronegative elements N, O, F, S, and Cl and if one of them is bonded to hydrogen, an additional hydrogen bonding potential is applied. For a recent overview on related hydrogen-bond corrections in semiempirical methods, see ref 71. Here, the H-bonding potential is represented by a sum over specific atom triples AHB (donor/acceptor-hydrogen-donor/acceptor) as

$$E_{hbnd} = -\sum_{AHB} f_{dmp}^{AHB} f_{dmp}^{hbnd} \frac{c_{AB}^{hbnd}}{r_{AB}^3} \quad (15)$$

where c^{hbnd} determines the interaction strength, f_{dmp}^{hbnd} damps the interaction to zero for large distances between donor and acceptor (similar to eq 7 but with an exponent of 12), and f_{dmp}^{AHB} introduces an AHB angle (θ) dependence by

$$f_{dmp}^{AHB} = \left(\frac{1}{2}(\cos\theta_{AHB} + 1)\right)^6 \quad (16)$$

so that the contribution vanishes for nonlinear arrangements. The dependence on the inverse third power of the AB distance in eq 15 is motivated by the strong polarization and dipole–dipole character of hydrogen bonding which is notably different compared to the ES part which decays as $1/r_{AB}$. To smoothly interpolate between donor/acceptor character, the interaction strength is modified by

dx.doi.org/10.1021/ct500573f | J. Chem. Theory Comput. 2014, 10, 4497–4514

$$c_{AB}^{hbnd} = \frac{c_{hbnd}(A) r_{AH}^2 + c_{hnbd}(B) r_{BH}^2}{r_{AH}^2 + r_{BH}^2} \quad (17)$$

Equation 17 employs the squared distances to the hydrogen atom weighted by atomic factors c_{hbnd}. They are composed of element specific values k_{hbnd} and charge dependence according to

$$c_{hbnd}(A) = k_{hbnd}(A) \frac{\exp(-k_{q1} q)}{\exp(-k_{q1} q) + k_{q2}} \quad (18)$$

This function ensures that the hydrogen bonding correction vanishes for weak donor basicity character as measured by the atomic charge q on atom A (e.g., for nitrogen in NO_2 or CN functional groups). The global parameters in this function are k_{q1} and k_{q2}.

Similar potential is used for the so-called halogen bonding situation[72,73] D−X−Y where atom D is the donor (N or O), X is the halogen (Cl, Br, I), and Y can be any atom. The interaction is treated similar to hydrogen bonding, i.e.,

$$E_{xbnd} = -\sum_{DXY} f_{dmp}^{DXY} f_{dmp}^{xbnd} \frac{c_X^{xbnd}}{r_{DX}^2} \quad (19)$$

where f_{dmp}^{DXY} is the angle dependent damping function in eq 16 but with the bending angle θ_{DXY} as an argument and an analogous D−X long distance dependent damping. The interaction strength c_X in the current very simple version depends only on the halogen acceptor atom. Similar to the HB case, it is composed of an atomic parameter k_X and by applying eq 18 with different parameters k_{q1} and k_{q2} which have been determined for Cl, Br, and I, by a fit to reference data in the XB18 set.[72] The sum in eq 19 runs over all atom triples with any donor atom from group 15−17. The other modifications compared to the HB case are that only a charge dependence on the atom X is involved (i.e., describing the so-called σ-hole[74]), a less steep distance dependence (r^{-2} vs r^{-3}) is applied, and the switching function (eq 17) is omitted. For another recent attempt to model halogen bonding, see ref 75.

The electrostatic part of the NCI is treated conventionally within a damped, fixed-charge model, i.e.,

$$E_{ES} = \sum_{AB} \varepsilon_{ES}(AB) \frac{q(A) q(B)}{r_{AB}} \quad (20)$$

where q are the QM atomic charges computed for the input equilibrium structure and ε_{ES} are topological screening parameters. They are zero if A and B are separated by one or two covalent bonds, i.e., 1,2- and 1,3-ES interactions are neglected to avoid double-counting with E_{intra} terms. For 1,4-interactions, an optimum value for ε_{ES} is obtained from a fit to conformational energies (the "conf" subsets from our GMTKN30 database[44]), and for 1,5- as well as all larger "distances," the screening factor is unity. Consistently, a similar procedure is applied for the dispersion/repulsion where $\varepsilon_{rep/disp}$ scales the 1,4- and 1,5-interactions, and it is zero and unity for smaller or larger distances, respectively. Topological screening factors are common in FF[16] and can be replaced for the ES part by damped (smeared) Coulomb interactions. Such an approach has been tested but not found to be generally superior. It has been abandoned because it complicates the interface of methods in QM/MM treatments. The same holds for the incorporation of charge-penetration effects which were recently found to be of some importance in intermolecular FF.[76,77]

The results of any fixed-charge FF are strongly dependent on the choice of the charges, and in larger, polar systems this is likely the most critical issue. After careful testing of many options, the CM5[78] charges recently proposed by Cramer and co-workers were taken. They are based on Hirshfeld partitioning[79] of the electron density as obtained from single-point calculations at the PBE[80]-COSMO[81]($\varepsilon = \infty$)/SV(P)[82] level. The resulting values are scaled by $k_q = 1.15$. Similar scaling factors of 1.14−1.27 for condensed phase simulations have been recommended recently by Vilseck et al.[83] The insensitivity of the CM5 charges to the theoretical level employed together with their robustness in complicated electronic structures represents a significant advantage which contributes to the accuracy of QMDFF. According to the authors' experience, these charges provide currently overall the best choice in the context of interaction energy calculations.

The global empirical parameters in the NCI part of QMDFF (a_1, a_2, s_8, k_D, k_Z, k_{hbnd}, k_{xbnd} β_{rep}, k_q, ε) have been adjusted so that QMDFF reproduces B3LYP-D3/def2-QZVP reference intermolecular potential energy curves for small molecules and interaction energies in the S22[42,84] and XB18[72] test sets. They were partially readjusted on conformational energy benchmarks in order to get consistent inter- and intramolecular NCI. The fitting procedure of all global parameters was at least partially conducted manually "by inspection."

2.6. The Hessian Fit. After the topological analysis of the input structure and assignment of the potential functions, the QMDFF generation algorithm continues with the determination of the unknown FC k_{str}, k_{bnd}, and k_{tors}. At this step, the potential stemming from V_{tors}' and the E_{NCI} part is included but kept fixed in the form described above.

The Levenberg−Marquardt (LM) algorithm[85,86] is used to minimize the squared deviations between QM and FF Hessians, i.e.,

$$\sum_{ij} (H_{ij}^{QM} - H_{ij}^{FF}(\mathbf{k}, [\mathbf{k}'\,]))^2 \rightarrow \min \quad (21)$$

where H_{ij} are the Cartesian matrix elements of the molecular Hessian and $\mathbf{k}$ represents the unknown FC. The "frozen" FC in the torsional part and the fixed NCI contribution is denoted by $\mathbf{k}'$. The algorithm requires calculation of the Jacobian matrix $\mathbf{A}$ given by

$$A_{ij,m} = \frac{\partial H_{ij}^{FF}}{\partial k_m} \quad (22)$$

Here, ij is a combined index which refers to the numerically significant matrix elements $>10^{-3}$−10^{-4}. The dimension of $\mathbf{A}$ is roughly 10 times the number of atoms N (= number of fitted parameters, index m) times about $100N$ (= number of significant Hessian elements). A ratio of about 5 to 10 between the number of unknowns and the number of parameters contributes to a stable optimization which is observed in practice. In the course of the present work, hundreds of molecules with up to 800 atoms have been treated, and no significant problems appeared in the fitting scheme. The $\mathbf{A}$ matrix requires storage of about 5 Gb of core memory for 1000 atoms, which is readily available on modern computers. The other computationally demanding step is

$$\mathbf{T} = (\mathbf{A}\mathbf{A}^{\dagger})^{-1} \quad (23)$$

where the matrix $\mathbf{T}$ is then used to update iteratively the parameter vector $\mathbf{k}$. The entire fit procedure usually converges

robustly within a few cycles and takes overall about 1 to 2 orders of magnitude less computation time compared to the calculation of the QM Hessian. In typical runs, about 1−5% of the FC (mainly in the torsional part) becomes negative, which is an indication of strong linear dependencies in the space of potential functions or numerical noise in the Hessian matrix. These terms are removed automatically, and the fit is repeated which is usually sufficient to obtain a FF with entirely positive FC. Note that this only holds if the input Hessian is positive definite and represents a true minimum on the PES. Small imaginary eigenvalues in the input Hessian often related to numerical noise from the DFT calculation, however, do not cause any numerical problems. As an initial test for the quality of the fit, the average deviation between QM computed and QMDFF harmonic vibrational frequencies is used which is discussed below.

In summary, the final FF contains a bonded part which by construction reproduces the input structure (which is supposed to be accurate). This is in contrast to a standard FF for which structure prediction is already a challenge. However, this alone would be relatively useless because already for small movements away from the equilibrium nonbonding terms become very significant. The early work of the Goddard group[39] stops exactly at this point and hence is limited to the equilibrium structure (the other less important difference to our work is that singular-value-decomposition instead of LM is used for the Hessian fit). The inclusion of the E_{NCI} and V' terms remedies this extrapolation problem, and as shown below leads to very good accuracy in regions of the PES far away from the equilibrium. There is, however, a price to pay: the additional potentials do not necessarily fulfill the condition $E_{NCI}(\boldsymbol{R}_e) = V'(\boldsymbol{R}_e) = 0$ (and analogously for the forces), and hence the QMDFF and QM input equilibrium structures are (slightly) different. The magnitude of this difference in the PES, which ideally for a perfect FF should vanish globally, is a measure for the quality of the generated FF and is discussed below for exemplary cases.

3. TECHNICAL DETAILS

We employ dispersion-corrected DFT, i.e., TPSS-D3/def2-TZV(-f),[64,65,87,88] as the default QM level and conduct full optimizations and analytical Hessian (unscaled) calculations. All D3 dispersion corrections used refer to the Becke−Johnson default damping scheme.[65,66] The TURBOMOLE[31,32] or ORCA[89,90] codes are applied. Very similar FFs are obtained at the "cheaper" PBE-D3/SV(P) level[80,82] or even with the semiempirical DFTB3-D3[50,52] method. Although GGAs like PBE or TPSS yield in general slightly worse molecular structures than hybrid or even double-hybrid functionals,[91] their cost/performance ratio for large structures is extremely good, and furthermore the FC (vibrational frequencies) can be directly compared to experimental data without any further scaling. Calculations which employ other QM methods than the default are indicated.

The Hirshfeld[79] partitioned electron density which is required to compute the CM5[78] charges is obtained from single-point calculations at the PBE-COSMO($\varepsilon = \infty$)[81]/SV(P) level and scaled as noted above. The bond orders (BO) are taken from the same PBE-COSMO calculation. As a fall-back procedure (if PBE-COSMO fails, e.g. for self-interaction error (SIE) related SCF convergence problems[92]), the PBE0 hybrid functional[93] or even small basis set Hartree−Fock (i.e., the new HF-3c[94] method) is recommended at this step. The COSMO model is mostly employed here to avoid as far as possible the above-mentioned convergence problems with GGAs in larger systems with localized charges. For a few host−guest complexes, the COSMO calculations could not be conducted due to cavity construction problems, and in these cases the charges refer to the bare PBE level. The typically observed changes in, e.g., noncovalent interaction energies when PBE-COSMO charges are replaced by PBE (or even HF-3c ones) are <5−10%.

The default neglect threshold for the Hessian matrix elements included in the fit is $\tau_{H_{ij}} = 10^{-4}$, but the results were found to be rather insensitive to this value. For larger systems or when sufficient memory is not available, increasing the value to 10^{-3} is recommended.

In the TB procedure to determine V_{tors}', Stewarts' STO-4G expansion[95] combined with Hermans' orbital exponents[96] is used to represent the minimal AO basis set. For the hydrogen s orbital, a scaling factor of 1.25 is used. The atomic valence orbital levels up to $Z = 36$ are also taken from ref 96, while for all other elements, HF/SV(P) values are used. For the off-diagonal matrix elements in the TB Hamiltonian, the simple average of diagonal elements scaled by the overlap integral and the standard factor of 1.75 (average of σ and π values) is taken.[97]

If not stated otherwise, all benchmark studies of energetic properties employ input structures from the respective test set in single-point calculations. The QMDFF, however, also in these cases refers to the fully optimized structure from the respective QM method. This approach is consistent with the treatment with other FF or semiempirical QM methods. All results as well as reference values refer to isolated molecule (gas phase) conditions.

All visualizations of molecules and overlays were produced with USCF Chimera version 1.6.1.[98] The root-mean-square deviation (RMSD) of two geometries was calculated using a quaternion algorithm[99] in order to get an all atom best fit. For the UFF calculations, the TURBOMOLE[31,32] implementation was used. The PM6-DH2 and all DFTB3 calculations employ the MOPAC[100] and dftb+ codes,[51] respectively. The MMX FF computations were conducted as implemented in the PCM code.[101] The Amber, MM3, and OPLS2005[102] calculations were carried out using Maestro[103] and MacroModel[104] with a dielectric constant of unity and no distance-based cutoffs for interactions. Amber was employed as implemented in MacroModel. It is a modified version of Amber94,[105] which includes Kollmans' 6,12-Lennard-Jones treatment[106] for hydrogen bonds and an improved peptide backbone parameter set.[107] The MM3 implementation is a slight variation of the original MM3[108] as documented in the MacroModel manual.

Table 1 gives the optimized values for all necessary empirical parameters in QMDFF.

4. RESULTS AND DISCUSSION

4.1. Vibrational Frequencies. Because the largest part of QMDFF is based on a rather accurately computed QM Hessian matrix, the comparison of the corresponding harmonic vibrational frequencies merely gives some indication for the quality of the fit, and therefore it is briefly discussed here. Because the accuracy of DFT computed frequencies in comparison with experimental data is well documented,[109] the comparison is made here between DFT-D3 and QMDFF computed values derived for the same input structure. If the

Table 1. Global Empirical QMDFF Parameters in Atomic Units[a]

parameter	value (element/group/row)
k_a	1.755 (H, He), 2.287 (B), 2.463 (C),
	2.559 (N), 2.579 (O), 2.465 (F, Ne),
	2.2 (group I), 2.8 (group II), 2.75 (row 3)
	2.95 (row 4), 3.15 (row 5), 3.80 (row 6)
k_Z	2.35 (row 1), 0.95 (row 2), 0.75 (row 3)
	0.65 (row 4), 0.6 ($Z > 54$)
	1.7 (Li), 5.5 (Be), 2.5 (Na), 3.0 (other in group I/II)
k_{EN}	−0.164
k_{a2}	0.221
k_{a13}	2.81
k_{b13}	0.53
k_{13r}	0.7
k_{dmp}	0.11
k_{ovlp}	0.5
a_1	0.45
a_2	4.0
s_8	2.7
$\varepsilon_{ES}(1,4)$	0.85
$\varepsilon_{disp/rep}$	0.5
β_{rep}	16.5
k_{hbnd}	0.8 (N), 0.3 (O), 0.1 (F), 2.0 (P−Cl,Se,Br)
k_X	0.3 (Cl), 0.6 (Br), 0.8 (I), 1.0 (At)
k_{q1}	10 (hbnd), −6.5 (xbnd)
k_{q2}	5 (hbnd), 1 (xbnd)
k_q	1.15

[a]For definition and description see text.

QMDFF frequencies are computed in this way at the QM input structure without relaxation, a few small imaginary modes on the order of 20−50 cm^{-1} could be obtained which usually disappear upon QMDFF optimization. Their magnitude and number provide a good indication for the effect of the added and modified potential terms.

The 12 organic molecules from the ROT34 benchmark set[47,91] (ethynyl-cyclohexane, isoamyl-acetate, diisopropyl-ketone, bicyclo[2.2.2]octadiene, triethylamine, vitamine C, serotonine, aspirin, the natural product cassyrane, limonene, lupinine, and a proline derivative Ac-Pro-NH2) and 10 common transition metal complexes (dubbed TMC10 set, see Figure 2) were used as test cases. The QM level for the organic systems is B3LYP-D3/def2-TZVP,[64,65,88,110,111] and the complexes were optimized at the PBE-D3/def2-SV(P) level. A plot of QM versus QMDFF computed frequencies is given in Figure 3. Only QMDFF frequencies >10 cm^{-1} are included.

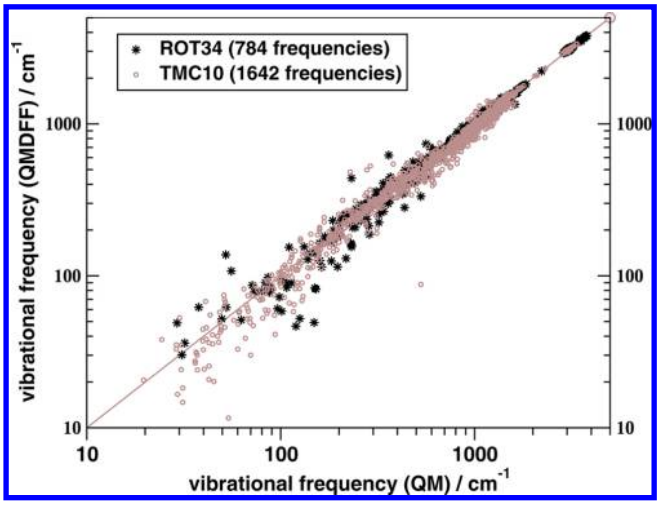


Figure 3. Comparison of QM and QMDFF calculated harmonic vibrational frequencies for 22 organic and organometallic molecules consisting of 18 to 100 atoms (ROT34 and TMC10 sets). The straight line indicates a perfect agreement between both data sets.

The mean absolute deviation (MAD) between the two data sets is only 44 cm^{-1}, the maximum deviation is about 400 cm^{-1}. In general, the largest relative deviations occur for low-lying vibrations <200 cm^{-1}, but also many of the very low-lying ones <50 cm^{-1} are computed accurately, which indicates the quality of the potential, in particular of the torsional part. All in all, QMDFF performs for this property as well as a very good special purpose vibrational FF. Note that the comparison has been made for nonoptimized structures, which results in some deviations attributable to the fact that QM and QMDFF minima are not identical. This in particular holds for the deviations of low-frequency vibrations which correspond to large-amplitude motions in MD simulations. Additional validation for this low-energy part of the QMDFF potential is given in section 4.4 on conformational problems. However, the chosen treatment (QMDFF Hessian//QM geometry) seems to be consistent because a perfect FF would yield vanishing deviations in such a comparison. Part of the differences are also attributed to the inability of the fitted GLJ potential to describe

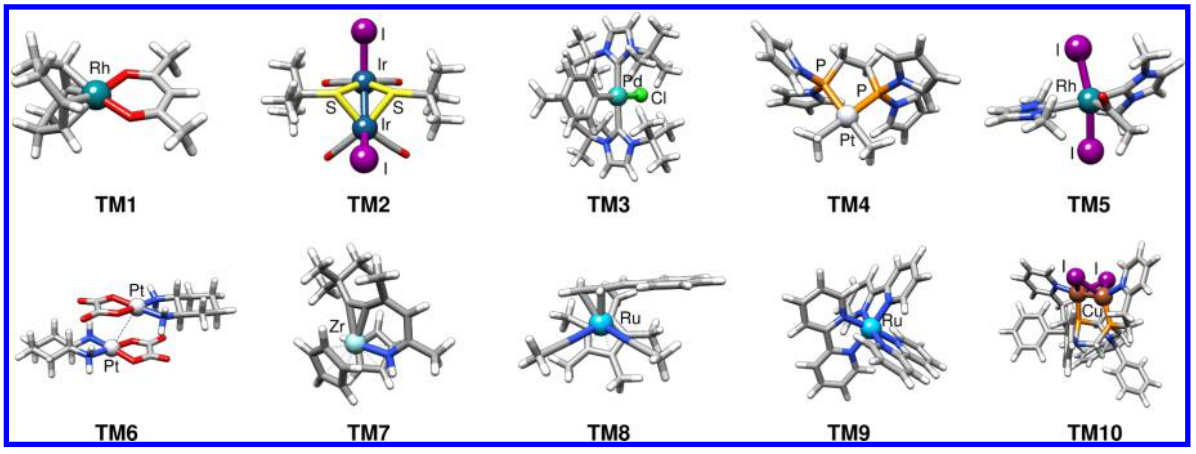


Figure 2. QMDFF optimized structures for the TMC10 test set of 10 larger transition metal complexes.

dx.doi.org/10.1021/ct500573f | J. Chem. Theory Comput. 2014, 10, 4497−4514

a parabola, i.e., already for diatomic molecules, deviations of a few cm^{-1} are obtained. Tests for the quality of fundamental (anharmonic) frequencies which naturally result in QMDFF will be the subject of future work.

4.2. Structures. The equilibrium structures of molecules are fundamental for the computation of many other properties, and it seems natural for an approximate method to test the quality of equilibrium geometries in some detail. As noted above, QMDFF more or less reproduces the QM input structure by construction, and hence this comparison only yields indications for the quality of the V_{tors}' and E_{NCI} parts. Again, ROT34 and TMC10 will be used, but in addition the structures of six more flexible drug-like systems will be tested.

4.2.1. Organic Molecules and the ROT34 Test Set. For the 12 molecules in the ROT34 set, the quality of a structure is assessed indirectly by comparison of the equilibrium rotational constants B_e as discussed recently.[47,91] This property is very sensitive even to fine structural details, and deviations from the accurately measured reference gas phase data provide clear indications for the quality of a theoretical method. Note that most of the considered molecules are flexible and contain important intramolecular interactions of various types. In this and the following test, optimizations were always started from the QM equilibrium structure so that any reliable method could easily converge to the preferably same local mimimum. For ROT34, the considered conformations represent global minima as derived from low-temperature experimental data while for the transition metal test set (see below), reliable model structures (which not necessarily represent the global minimum) are considered. The results are summarized in Table 2.

Table 2. Mean Deviations (MD), Mean Absolute Deviations (MAD) of Calculated and Experimental Rotational Constants for the ROT34 Test Set of Organic Molecules, and Corresponding Relative Standard Deviation (SD) for Different Methods[a]

method	MD (MHz)	MAD (MHz)	SD (%)
PW6B95-D3	−8.4	8.4	0.4
QMDFF (PW6B95-D3)	4.6	14.1	2.5
B3LYP-D3	11.6	11.6	0.8
QMDFF (B3LYP-D3)	23.2	23.3	2.6
DFTB3-D3	19.8	22.0	1.3
UFF	2.6	41.4	3.5
MM3[b]	2.9	43.9	3.0
PM6-DH2[c]	31.3	59.2	8.3

[a]Positive MD values indicate too large (spatially extended) structures. [b]MMX[101] FF for molecules with aromatic rings. [c]One outlier with a different conformation.

Importantly, we observe no fundamental change in any of the conformations in ROT34 compared to the experimentally observed lowest-lying conformer with QMDFF. The largest deviation is found for the aspirine molecule, where the acetyl group is somewhat rotated toward planarity. The PM6-DH2 method yields one outlier; i.e., the optimization converges to a different local minimum, which is an indication of a lower-quality PES. Because QMDFF is derived from specific DFT-D3 input data, at best the MAD and standard deviations (SD) of this particular method can be expected. To test this hypothesis, we employed two functionals which yield systematically different structures: B3LYP-D3, for which too "large" molecules (MD for $B_e > 0$) are obtained, and PW6B95-D3,[64,65,112] which yields too "compact" molecules (MD for $B_e < 0$).

As can be seen in Table 2, the QMDFF structures are slightly less accurate than the input. The SD increases by 1−2%, and the MAD from 12 to 23 and from 8 to 14 MHz, respectively. Almost always, QMDFF increases covalent bond lengths between main group elements compared to the QM reference, which leads to overall spatially more extended structures and larger MD values. Because the B3LYP-D3 structures are already too large, this increases the MAD while for PW6B95-D3, which has an MD < 0, only a small increase of the MAD is observed. On average, this systematic bond lengthening is about 0.003 Å for non-hydrogen distances, which is certainly tolerable for a FF. In general, this means that, regarding structures which should as closely as possible match the experiment, a QM input method which systematically yields too short distances like PW6B95-D3 is preferred.

In comparison with the other tested FFs (UFF and MM3), QMDFF performs much better for both QM inputs. This is notable because in particular MM3 has been designed to reproduce the structures of organic molecules, which is also indicated by its small MD value. With the B3LYP-D3 input, QMDFF is only slightly inferior compared to the semiemprical DFTB3-D3 method (which actually performs much better than PM6-DH2) while the QMDFF based on PW6B95-D3 is very accurate and close to the best performing QM methods.[47]

Figure 4 shows overlays of QM (gray) and QMDFF optimized structures for Penicillin G, a cyclic decapeptide (with +2 total charge), an artificial fat substitute termed Olestra which contains many, very flexible long alkenyl chains (462 atoms in total), two common anticancer drugs (Vinblastin and Etoposid), and a prototype of a frustrated Lewis-pair (FLP). The semiempirical DFTB3-D3 method has been used as QM input except for the FLP, which refers to a PBE-D3/TZVP calculation.

As can been seen from Figure 4, there is a strikingly good agreement of the QM and QMDFF structures, and in particular the close agreement of the very flexible Olestra molecule is encouraging. This clearly demonstrates that the NCI and torsional parts are well described in QMDFF, which is a prerequisite for accurate biosimulations. The observed RMSD values of 0.2−0.3 Å can be considered as small for molecules of this size. Also noteworthy is the good performance for the FLP including important π-stacking interactions and a characteristic P−B interaction that is a borderline case for a covalent bond. The QMDFF procedure automatically takes care of such unusual situations and in this case assigned a bond function (eq 3) with a very realistic bond energy of $k_{str} = 6.5$ kcal mol^{-1} as a result of the fit (cf. ref 28).

4.2.2. The Transition Metal Test Set. Transition metal complexes represent a challenge for any FF. Apart from the parametrization problem (which does not exist in QMDFF), the structural variety is much larger than in (bio)organic systems, and many intramolecular interactions are softer. For example, highly coordinated cyclopentadienyl complexes with very small rotational barriers and many vibrational modes <50 cm^{-1} are difficult to describe.

As can be seen from Figure 2, the QMDFF optimized transition metal complex geometries look very realistic. The systems have been taken more or less randomly from current or previous work[113,114] and are by no means "cherry-picked" to represent only well-behaved cases. Although the set is not extensive, it covers quite different binding motifs with a wide

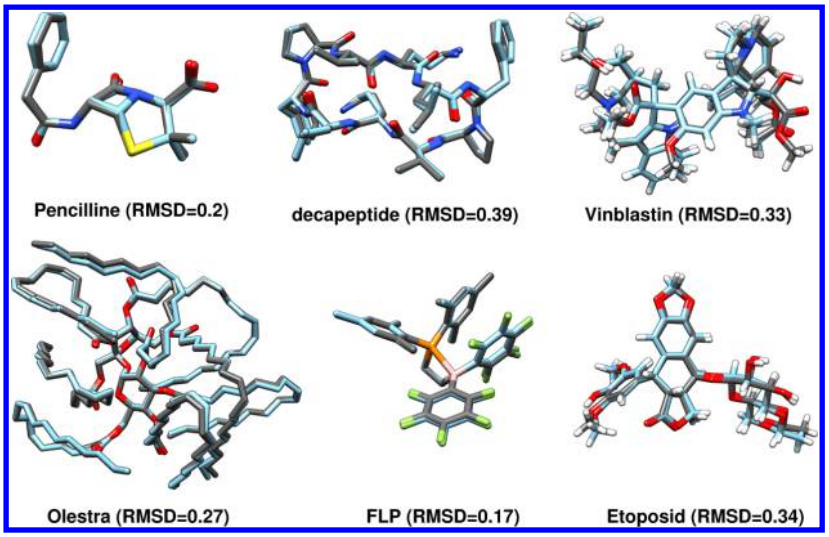


Figure 4. Overlay of QM reference (gray) and QMDFF (cyan) optimized structures for six larger and flexible organic molecules. The root-mean-square-deviation for the all-atom best fit is given in Å. The minimum conformations shown have been obtained randomly from some start structure.

range of metal coordination numbers from four to about 10. An overlay of QM input and QMDFF structures in this case provides little information because the systems are relatively crowded. Therefore, the overall structural RMSD values and mean deviation for metal ligand (ML) distances (both in Å) are discussed (Table 3). The chemically important ML distances are often soft and most strongly influenced by the added E_{NCI}. In this test, the QM input data serve as the reference.

Table 3. Mean and Mean Absolute Deviations (MD and MAD) of DFT (PBE-D3/def2-SV(P)) and QMDFF Calculated Metal-Ligand Distances and All-Atom-Best-Fit RMSD Values for the TMC10 Test Set (all values in Å)

	entire structure	metal−ligand distances	
entry	RMSD	MD	MAD
1	0.36	−0.003	0.003
2	0.14	0.021	0.021
3	0.19	0.075	0.075
4	1.18	0.019	0.019
5	0.54	0.048	0.048
6	0.32	−0.033	0.033
7	0.37	−0.010	0.010
8	0.43	0.031	0.031
9	0.05	−0.023	0.023
10	0.17	−0.003	0.013

The RMSD values for nine out of the 10 complexes are rather small, and this indicates overall good resemblance of QM and QMDFF structures. The TM4 complex with an RMSD value of 1.17 is an exception. Here, the four carbene ligands have very shallow torsional potentials, and the two structures basically differ by about 20−30° rotation of these units while the ML distances are accurate. Because the molecular structure regarding such floppy degrees of freedom is anyway not well-defined at finite temperatures, this case is not considered as a significant failure. The MD and MAD values for the ML distances are mostly small and as opposed to the bio-organic systems are not systematically shifted; i.e., positive as well as negative MD values are found. The overall performance of QMDFF for these difficult cases with MAD values in the range of 0.003−0.08 Å (0.03 Å on average) is very satisfactory.

4.3. Atomization Energies. The literature on r_e−D_e and k_{str}−D_e relationships is extensive, but until now most of the existing ones only work for certain classes of molecules or are highly empirical. Often formulas which connect these fundamental quantities are only used for qualitative purposes but cannot be used practically in large scale computations. As already noted above, QMDFF employs stretch terms which allow the asymptotically correct dissociation of any bond, and an accuracy of about 10% of D_e was already shown for some diatomic molecules (see Figure 1). In this section, it is demonstrated that this also holds for total atomization energies of polyatomic molecules with more or less arbitrary structure, which is a very intriguing and practically useful result. The 1,3-stretch terms are essential for obtaining good accuracy for larger systems. It is emphasized again that only information about the equilibrium structure and chemical composition enters. The D_e values in QMDFF predominantly result from V_{str}^{12} and V_{str}^{13} (and minor V_{tors}' and E_{NCI}) terms and are computed simply by scaling all Cartesian coordinates by a large number and subtracting the resulting total energy from the equilibrium value ($E_{e,QM}$ is constant).

Three test sets are used: the G2 set[45] with 148 mostly small molecules, 18 amino acids[46] with highly accurate W2 reference data, and 50 medium sized main group molecules taken out of the "difficult molecules" section (as indicated by a large B3LYP error for D_e) from Cioslowskis' 500 molecule benchmark set.[115] Because the empirical model parameters have been obtained from a fit to the G2 set, the other two databases are considered as cross-checks. The results are summarized graphically in Figure 5.

The MAD for the atomization energies in the G2 fit set is 22 kcal mol^{-1} (MD = 0.2 kcal mol^{-1}) with a mean relative deviation (MAD%) of only 8%. This is remarkable because many rather difficult systems are in this set including relatively weakly bound diatomics like Li_2 and Na_2. The largest relative deviation of about 100% (D_e of 86 kcal mol^{-1} (QMDFF) compared to 39 kcal mol^{-1} (reference)) occurs not unexpectedly for the F_2 molecule. Larger errors are also obtained for other fluorine containing molecules, which indicates that for this notoriously "difficult" element a general

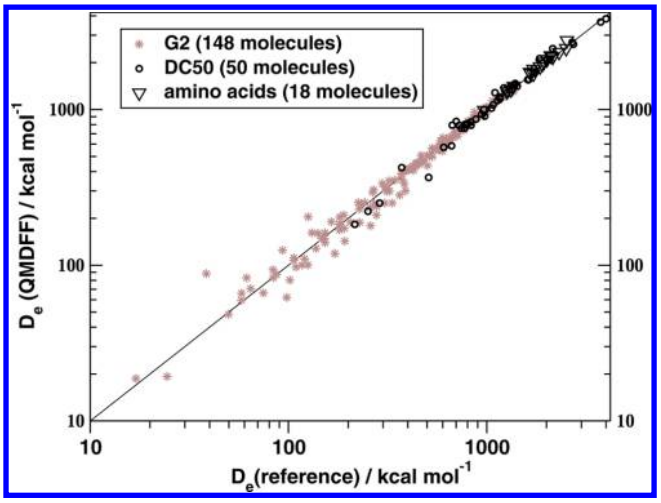


Figure 5. Comparison of QMDFF and reference atomization energies for three test sets: the G2 set used for fitting and the amino acid as well as DC50 sets for cross-checking. The QMDFF is based on TPSS-D3/def-TZVP(-f) QM input data. The straight line indicates a perfect agreement between both data sets.

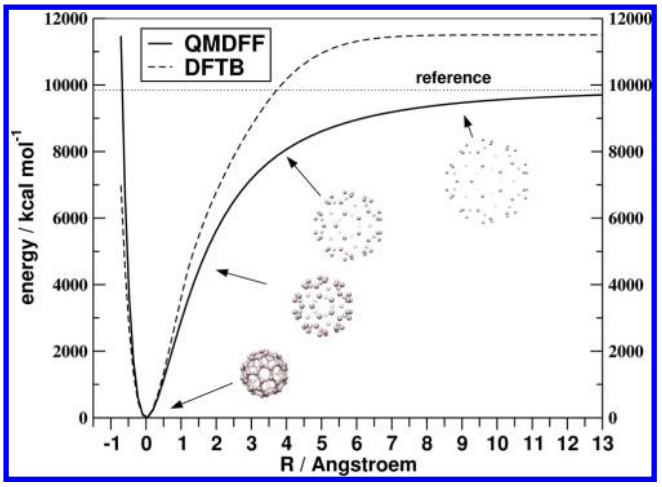


Figure 6. Symmetric full dissociation potential energy curve for $C_{60} \rightarrow$ 60 carbon atoms. The theoretical reference value (which is in reasonable agreement with experimental data) for the atomization energy is taken from ref 117. The QMDFF has been obtained at the PBE-D3/SV(P) level, and DFTB refers to the nonself-consistent charge level (for details see ref 118). The coordinate R describes the elongation or compression of the atomic positions from the equilibrium value.

k_{str}–D_e relationship is problematic. Notably, the MAD% values for the amino acid and DC50 cross-checks are even smaller than for the G2 fit set (3 and 6%, respectively) which clearly indicates the generality of the new relationship for polyatomic molecules. The MAD value for these two sets with 60 and 64 kcal mol^{-1}, respectively, is larger mainly because the absolute atomization energies are also larger (e.g., on average 1840 kcal mol^{-1} for the amino acids).

The achieved accuracy is remarkable for a FF which is not highly parametrized and only uses equilibrium information and is not specially designed for this purpose. However, the typical deviations of a few percent for D_e occur more or less randomly, and as opposed to most QM based electronic structure methods, these nonsystematic deviations preclude the use of the QMDFF D_e values for reaction energy computations. This, however, is also not necessary because $E_{e,QM}$ for thermochemistry applications is already available from the QM treatment. The perspective is that QMDFF can be used to investigate the thermal stability of almost arbitrary molecules because (a) asymptotically it yields good (probably for many applications accurate enough) limiting energies of dissociation and (b) the initial slope of the PES is also described rather well and relative bond strengths are reflected at least semiquantitatively.

For example, the CC bond D_e values for ethyne, ethene, and ethane are reasonably computed to be 189, 141, and 72 kcal mol^{-1}, respectively, with corresponding CH values of 110, 104, and 101 kcal mol^{-1}. These values nicely reflect chemical intuition and correspond well to the results from other energy partitioning schemes.[116] Reasonable trends of metal–ligand bond strengths are not discussed in detail here but support this view. This essentially means that in a high temperature MD run, the various bonds are correctly broken with at least a qualitatively correct rate.

As an example for a large system, the PES for the complete, symmetric dissociation of C_{60} into 60 carbon atoms is shown in Figure 6.

The QMDFF computed atomization energy agrees almost perfectly (within 0.15% or 15 kcal mol^{-1} deviation) with the high-level reference value of 9845 kcal mol^{-1}.[117] As can be seen from the plot, the PES looks quite realistic, and a major part of the dissociation process occurs at about 5 Å elongation from the equilibrium. To the best of the authors knowledge only noniterative tight-binding based QM methods like DFTB can be applied to such process because the single state dissociation to 60 carbon atoms each in the ^3P ground state is currently intractable with all other QM methods. The DFTB curve shown for comparison is very similar to the QMDFF result near the equilibrium but reaches the asymptotic region at smaller elongations due to the more exponential character of the underlying potential. The D_e computed by DFTB, however, is less accurate and about 15% off the reference value. It can be expected that QMDFF is useful for studying the mechanical properties of molecules, e.g. their tensile strength.

4.4. Conformational Energies. The development of the torsional potentials was the most involved part of QMDFF, and these terms represent the "Achilles' heel" of many contemporary FFs.[119] Again various benchmarks will be used to assess the quality of the QMDFF approach. In the fitting process of the global parameters, the well-established ACONF, CYCONF, SCONF, and PCONF subsets from the GMTKN30 database[44] employing accurate CCSD(T) reference values were used. From the SCONF set, only the 15 conformers of the five-membered ring system 3,6-anhydro-4-O-methyl-D-galactitol were included. The five additional sets octanol (21 conformers), melatonine (52 conformers from ref 120), eight diproline conformers,[121] five cys-asn-ser conformers from ref 122, and 10 RNA backbone conformations from ref 119 were investigated as cross-checks. The phe-gly-gly tripeptide in the PCONF set with 37 atoms is the largest system studied here. Note that all results refer to isolated molecule (gas phase) conditions which is different from many other biomolecular FF benchmark studies which employ effective condensed phase FF parametrizations and compare with experimental data.

The results are presented graphically in Figure 7 for the four sets for which consistent results from other FFs could be obtained. Table 4 shows mean absolute deviations for all sets in comparison with semiempirical QM methods and TPSS-D3/def2-TZVP(-f) on which the QMDFF fit is based in all these

dx.doi.org/10.1021/ct500573f | J. Chem. Theory Comput. 2014, 10, 4497–4514

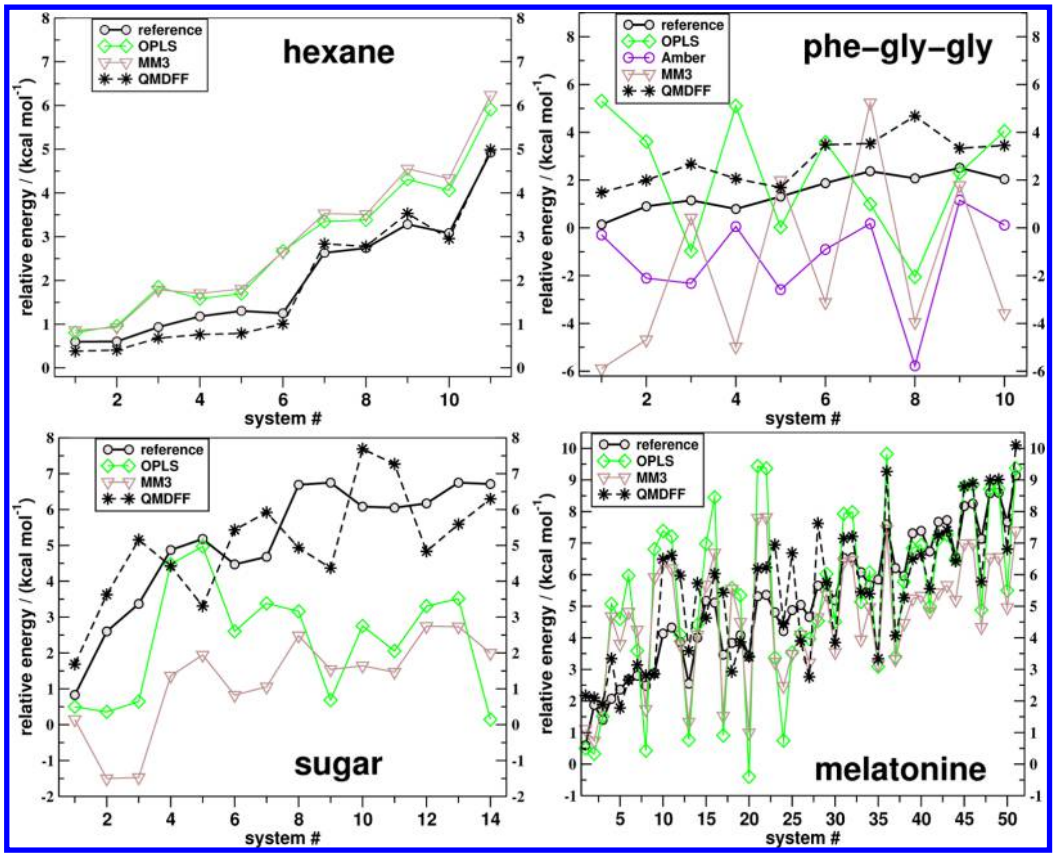


Figure 7. Graphical presentation of conformational energies for four benchmarks sets with a comparison of QMDFF and FF results. The lines connecting data points are just drawn to guide the eye.

Table 4. Mean Absolute Deviation (kcal mol^{-1}) for Conformational (relative) Energies[a]

	N_{conf}	QMDFF	DFT-D3	DFTB3-D3	PM6-DH2
hexane (ACONF)	12	0.23	0.19	0.90	0.80
octanol[b]	21	0.88	0.64	0.93	1.07
melatonine[c]	52	0.97	0.67	1.77	1.30
cytosine (CYCONF)	11	0.96	1.08	1.72	1.44
sugar (SCONF)	15	1.29	1.43	1.19	2.23
tripeptide (PCONF)	11	1.33	1.23	1.16	2.29
RNA backbone[d]	10	1.55	0.54	2.31	1.36
diproline[e]	8	1.94	0.76	1.35	3.19
cys-asn-ser[f]	5	2.09	0.38	0.89	2.77
average MAD		1.3	0.8	1.4	1.8

[a]The sets labeled "CONF" are taken from the GMTKN30 data base, and N_{conf} is the number of conformers in each set. [b]This work. MP2/def2-QZVP//TPSS-D3/def2-TZVP reference values. [c]From ref 120, reference values are est. CCSD(T)/CBS. [d]From ref 119, reference values are est. CCSD(T)/CBS. The molecule has one negative charge. [e]From ref 121, reference values computed in this work at B2PLYP-D3[123]/def2-QZVP level using the original structures. [f]From ref 122, reference values computed in this work at B2PLYP-D3/def2-QZVP level using the original structures.

cases. Again, exactly the same original structures from the respective benchmark set were used in single-point computations. The effect of applying geometries, which were not optimized by the respective method (single-point energy approach), has been estimated to be <0.2 kcal mol^{-1} for various methods as long as the basic structure of the two conformers is the same. The QMDFF was always generated for the lowest-lying conformer, and this FF was used to compute the relative energy of all other conformers. A reasonable comparison for these many structures including optimization is not possible because the conformers do not necessarily represent stable minima for all methods. Note that the chosen DFT-D3 level for comparison in Table 4 represents the typical accuracy which can be achieved routinely nowadays with first-principle QM calculations.

In comparison with other FFs and even computationally much more involved semiempirical methods, QMDFF performs excellently. The MAD for the hexane conformers (as a model for ubiquitous alkyl chains) is extremely good, and also the polar systems including the charged RNA backbone conformations are described rather accurately with MAD values of 0.9−1.5 kcal mol^{-1}. This should be compared to the average of the relative energies in all sets of about 3.5 kcal mol^{-1}. In most cases, DFTB3-D3 and PM6-DH2 are clearly outperformed. Slightly larger deviations (MAD of about 2 kcal mol^{-1}) are obtained for the diproline and cys-asn-ser peptide sets which are, however, still smaller than with PM6-DH2. Because QMDFF performs rather well for the very difficult tripeptide (PCONF) set,[44,124] we do not think that peptides represent a special problem for QMDFF. Furthermore, in some cases (CYCONF, SCONF, PCONF), the QMDFF accuracy (and also the individual relative energies) is close to that of the underlying DFT-D3 method so that at least part of the remaining error may not be due to the FF. For the CYCONF set, this hypothesis was investigated by performing B3LYP-D3/def2-QZVP computations of the QMDFF input data. At this DFT level the MAD is only 0.24 kcal mol^{-1}, but this is not reflected in the QMDFF result, which changes only slightly

from 0.96 kcal mol^{-1} with the TPSS-D3 reference to 1.15 kcal mol^{-1} with B3LYP-D3. More detailed investigations of relations between the QM input and QMDFF errors for a particular property (which was very obvious for structural data) will be the topic of future work.

Figure 7 shows a comparison with the results from standard FF, and already a brief visual inspection of the data reveals a substantial improvement by the new approach. Common FFs which are actually specialized for organic (peptide) systems fail, for example, spectacularly for the PCONF set (data from the Amber FF are not shown but are qualitatively similar to those from OPLS). Also seemingly simple systems like the hexane conformers are not described well by, e.g., the common MM3 method, although here at least the trends are correct. The melatonine conformers also behave more regularly; i.e., all three FFs more or less yield the same error pattern. The sugar conformers which contain many intramolecular hydrogen bonds again represent some kind of worst case scenario for standard FFs, while QMDFF is relatively close to the high-level reference data.

As a final more realistic test, we run an MD for the anticancer drug Taxol (113 atoms) for 10 ps on the QM level (DFTB3-D3) and conducted single point computations on these structures with QMDFF, and OPLS and MM3 for comparison (see Figure 8). The molecule contains a variety of functional groups and typical intramolecular interaction motifs like aromatic stacking as well as hydrogen bonding.

As can be seen from the plot, QMDFF very closely follows the QM reference data while the other FFs (OPLS in particular) show much larger deviations. This confirms the results of the conformational benchmarks for a much larger molecule and under conditions which are usually encountered in FF applications.

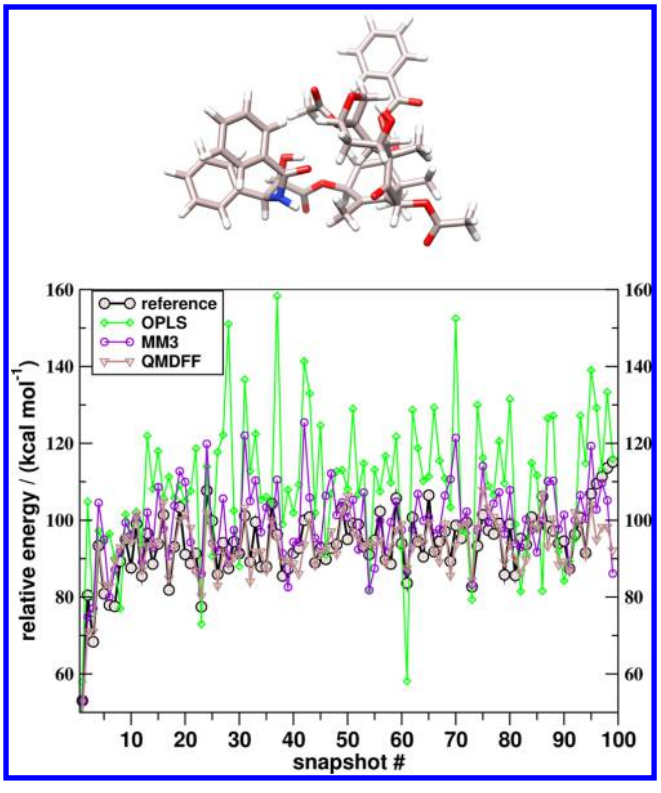


Figure 8. Relative energies for 100 equidistant MD snapshots on a 10 ps QM(DFTB3-D3) trajectory. The NVT run was started from the QM equilibrium structure (shown in the top of the figure) with randomly assigned velocities corresponding to T = 300 K.

In summary, it is found that QMDFF performs extremely well at the fixed-charge FF level, outperforming standard potentials and even semiempirical QM methods. This finding confirms the conclusions from the structural benchmark of the floppy systems that the torsional (and the NCI part, see below) potentials are very reasonable. Before going to the last section on intermolecular interactions, a further note on the conformational problem seems appropriate. Because QMDFF represents an expansion around the lowest-energy conformation of a relatively accurate DFT-D3 PES, its very good performance compared to standard FF (which try to model the PES more globally) is not unexpected, and the comparison is somewhat unfair. In fact, if QMDFF is generated starting from a very high-lying conformer, its performance degrades considerably and even lower-lying conformers are not found at all. This, however, is not a true "failure" but is expected and results from the basic construction principle that the reference structure is the lowest local minimum in a relatively large region of the PES. If this is not fulfilled, the $E_{e,QM}$ term in eq 1 must be taken into account. For practical applications, this means that, first, the lower energetically lying the reference point is and, second, the closer the "distance" from this expansion point on the PES is, the better the performance will be.

4.5. Noncovalent and Supramolecular Interactions. Noncovalent interaction energies can be computed in two different ways with QMDFF: the dispersion coefficients and charges can be taken from QM calculations for the monomers, and these are then employed for the geometry of the complex. This is the common procedure which applies to solvation, QM/MM, and other condensed phase treatments. However, the QM input data can also be generated for the complex which may contain multiple noncovalently interacting fragments. This situation often appears in biomolecular simulations, e.g., when an experimental structure includes counterions or water molecules. QMDFF is constructed such that the two different situations lead to almost identical bonding potentials. The NCI parts are, however, different mostly because the QM charges in the second variant are generated by a supermolecular calculation in which the fragment densities are affected by polarization effects (minor differences from the C_6 coefficients in the D3 dispersion treatment are hardly visible). The differences of the two treatments have been evaluated for various cases and were often found to be small (<5% change in interaction energies for neutral systems). For charged complexes, however, the mentioned polarization effects can be significant, and the two approaches can differ by 10−20%, which should be kept in mind in actual applications. Here, we consistently employ QMDFF generated from monomer computations and use it in single-point calculations with the respective geometries taken from the benchmark set. The S22 and XB18 benchmark sets were used for determining the global empirical parameters, and results for S30L and X40 as cross checks will be discussed. The S30L benchmark is an extension of the S12L set of supramolecular complexes,[43,125] and the structures are shown in Figure 9. The new compilation contains part of the S12L molecules (complexes number 1−4, 9−10, 15−16, 19, 27−28 are from ref 43 and 125), additional binding motifs, as well as more charged and hydrogen bonded complexes (details will be published elsewhere). The back-corrected experimental reference energies for S12L have been validated recently by independent high-level computations[126]

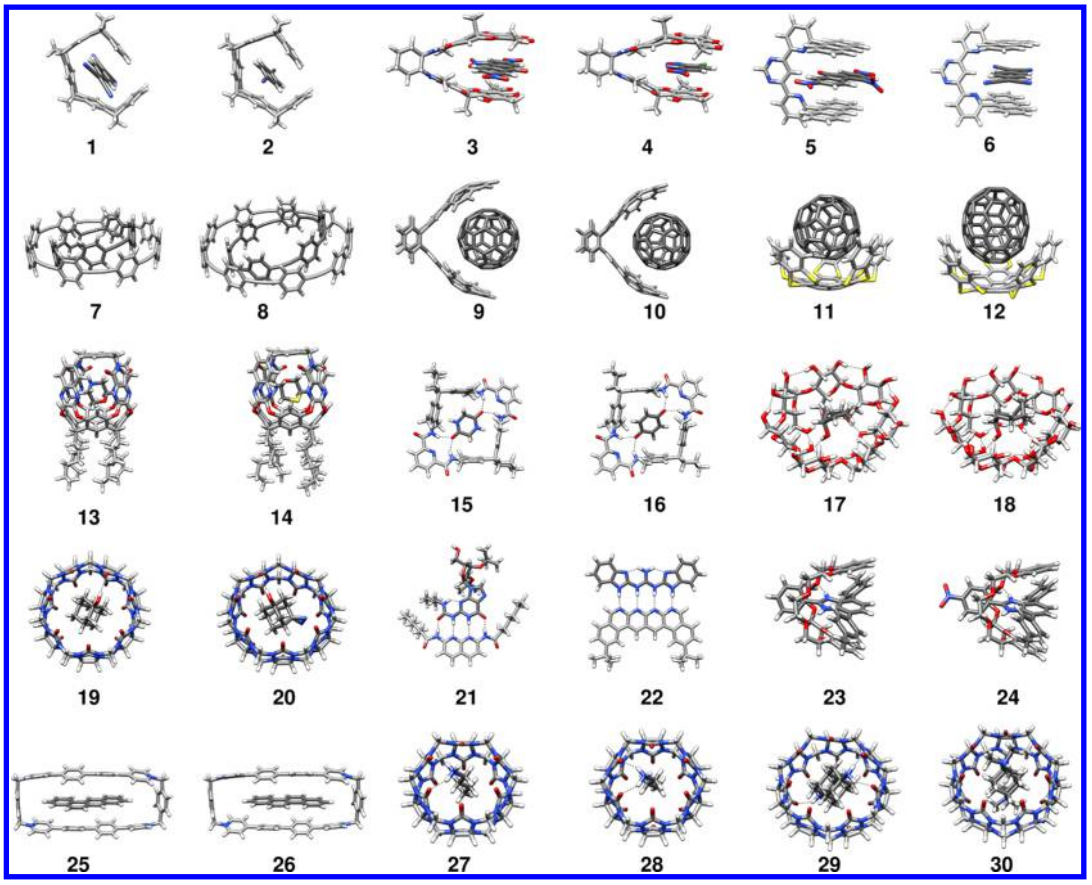


Figure 9. Structures of the investigated supramolecular complexes (S30L set).

but are not available for all S30L complexes so that we employ here consistently rather accurate PW6B95-D3[112]/def2-QZVP(-g,-f)[88,127]//TPSS-D3/def2-TZVP interaction energies with a typical error of about 5% or less.

Before discussing the benchmark sets, three example interaction energy curves for simple dimers (H_2O, benzene (stacked, parallel displaced), and CCl_4) are shown in Figure 10.

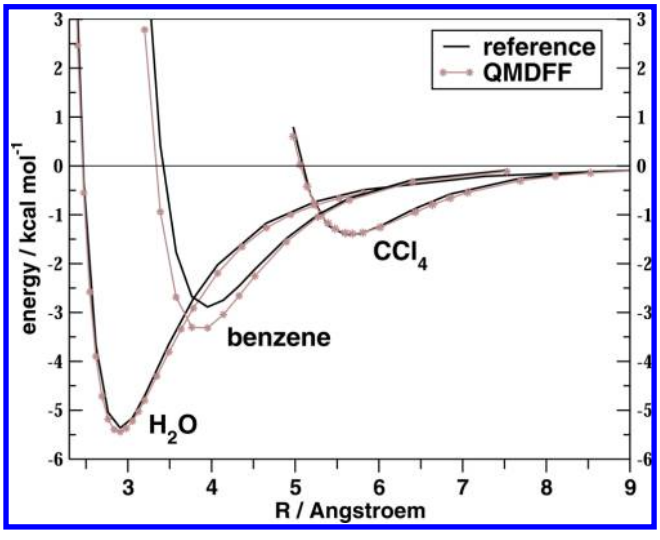


Figure 10. Potential energy curves for noncovalent interactions (frozen monomer geometries, B3LYP-D3/def2-QZVP reference data) for small dimer complexes. The coordinate R refers to intermolecular O−O or C−C distances, respectively.

Excellent agreement between QMDFF and B3LYP-D3/def2-QZVP reference data is noted not only for equilibrium and larger separations but also in the repulsive part of the potentials. In fact, the observed deviations between DFT-D3 and QMDFF are close to the accuracy of the reference data.

The MAD values for S22 and S30L are gathered in Table 5 and graphically shown in Figure 11 in comparison with

Table 5. Mean Absolute Deviation (kcal mol^{-1}) for Noncovalent Interaction Energy Benchmark Sets

	S22	S30L[a]	XB18	X40
QMDFF	0.84	6.3 (3.5)	0.98	0.95
DFTB3-D3	1.01	4.3 (2.3)	–[b]	–[b]
PM6-DH2	0.41	5.2 (4.1)	3.94	1.50

[a]Values in parentheses excluding charged complexes. Including monomer relaxation (binding energy). [b]Parametrization for Br and I atoms not available.

statistical data from semiempirical methods. Reasonable and consistent results for other FFs could not be obtained either due to technical problems or incomplete parametrization. For S22 and X40, the benchmark values do not include the relaxation energy, and for XB18 these values are very small and hence neglected. For S30L, the "true" interaction energies ($-D_e$) using relaxed monomer geometries are significantly different from the bare values, and hence the relaxation energy is included here and taken from the DFT-D3 reference calculations.

In comparison to the two semiempirical methods, QMDFF performs surprisingly well for general noncovalent interactions.

dx.doi.org/10.1021/ct500573f | J. Chem. Theory Comput. 2014, 10, 4497−4514

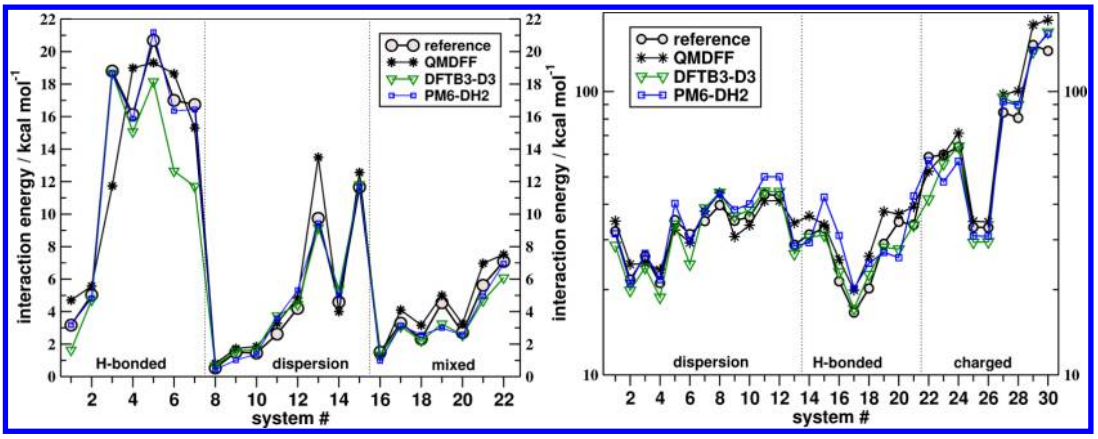


Figure 11. Comparison of interaction energies for the S22 (left) and S30L (right, logarithmic energy scale) sets at various levels of theory. The reference values refer to the CCSD(T)/CBS and PW6B95-D3/def2-QZVP(-g,-f)//TPSS-D3/def2-TZVP levels, respectively. All structures have been used in single-point energy calculations. The lines connecting data points are just drawn to guide the eye.

For S22, only the very strongly (doubly) hydrogen bonded systems 3−6 are as expected somewhat more problematic, but an overall MAD of 0.84 kcal mol^{-1} is quite respectable. While PM6-DH2 yields very good results for this set (on which the method similar to QMDFF and DFTB3-D3 was actually parametrized), its performance for XB18 and XB40 deteriorates while QMDFF again shows a typical 1 kcal mol^{-1} MAD accuracy level. Very good results from QMDFF are also found for the neutral complexes in S30L which are of comparable quality to those from the two semiempirical methods. For the cases 1−13, this is attributed to the accurate account of the dominating London dispersion effects while for complexes 14−21 also the hydrogen-bonding term is relevant and the good results clearly support the potentials employed. These very encouraging results suggest the application of QMDFF in the wide field of supramolecular chemistry as well as for protein−ligand complexes.

Due to their limited (or absent) treatment of short-range electrostatic (penetration) and polarization effects, all tested methods provide larger errors for the charged complexes as expected. Doubly charged systems (complexes 29−30) represent some kind of worst cases scenario for a fixed point charge model like QMDFF and the resulting errors of about 25%, which are regarded as rather typical for such situations, should be kept in mind, e.g., in biomolecular applications when charged residues are involved in the relevant interactions.

5. CONCLUSIONS

The aim of this work is the development of a universal and practical approach for the automatic generation of a classical, atomistic, and fully flexible force-field (FF) from quantum mechanical (QM) input data. As opposed to standard approaches, a molecule specific (nontransferable) potential is constructed which very closely resembles the accurate input potential near the equilibrium nuclear configuration. Particular attention has been paid and new ideas have been introduced to allow accurate and asymptotically correct extrapolation to nonequilibrium regions.

The proposed method termed QMDFF can treat covalently bound molecules and noncovalent complexes with almost arbitrary structure. The QM information from a single equilibrium structure, the corresponding Hessian matrix, atomic partial charges, and covalent bond orders is used to automatically derive the full FF with all force constants. It is therefore very user-friendly, efficient, and less error-prone because basically only one quantum chemical calculation for the target system has to be conducted. The out-of-the-box approach, which covers all elements up to $Z = 86$, is based on about 50 global, system-independent empirical parameters which require practically no further user attention. Most of these parameters are element specific or depend only on the row of the element in the periodic table. About 20 of them were actually fitted in a least-squares sense to accurate reference data which underlines the small degree of empiricism in QMDFF. Subsequent calculations for the target molecule of the QMDFF equilibrium structure, vibrational frequencies, and a short molecular dynamics run are used to internally validate the FF generation process. Molecules with up to 1000 atoms can be treated currently on common desktop computers while larger systems require fragmentation techniques and some kind of merging procedure of the resulting potentials. This will be the topic of further research.

Because QMDFF is a general FF which can be generated for very different systems including also metal complexes, the possible areas of application are enormous and can only be briefly mentioned here. First, the conformational flexibility of isolated molecules could be explored more automatically and systematically than it is possible for example at the DFT level. For molecules with unusual electronic structure, for metal containing systems, or when in general no standard FF is available, QMDFF provides unique and outstanding capabilities in this respect. The second new and intriguing aspect is that QMDFF is fully anharmonic and can relatively accurately dissociate molecules into atoms. This opens the route to average more realistically QM computed properties, e.g. for the spectroscopy of large molecules, but furthermore allows the direct evaluation of their thermal stability or "mechanochemical" properties. Moreover, the inter- and intramolecular noncovalent interactions are treated accurately and consistently, which is important for all condensed phase or biochemical applications.

This works contains comprehensive benchmark data on basic and typical problems. The results indicate that QMDFF not only enlarges the chemical space that can be treated by FF but also significantly improves the accuracy compared to standard and even special purpose potentials.

Structural data and vibrational frequencies are very accurately computed by QMDFF because it is basically constructed from

dx.doi.org/10.1021/ct500573f | J. Chem. Theory Comput. 2014, 10, 4497−4514

these data. In this respect, QMDFF merely reflects the accuracy of the QM input, and the small differences between these QMDFF and QM computed properties are used to evaluate internal consistency. Nevertheless, QMDFF has the important feature that it reproduces as closely as possible the QM potential energy surface and that only small shifts due to systematic errors in bond lengths occur. This is important if QMDFF MD snapshots are used in further QM studies.

The accuracy achieved for the conformational energies of bio-organic systems is about 1 kcal mol^{-1} in the typical energy range of 0−5 kcal mol^{-1}. This is remarkable for a relatively simple fixed-charge model and much better than for all standard FFs tested in this work. While this is certainly good enough for many practical purposes, some important biomolecular structures like RNA are extremely sensitive to details of the torsional potentials where this accuracy is still insufficient. In the authors' opinion, any substantial improvement in this direction requires more or less full QM treatment and hence will substantially increase the computational effort. Note, however, that the simple semiempirical QM methods tested on the very same conformational benchmarks do not perform much better (sometimes even worse) than QMDFF and that the whole problem is in general rather hard to solve.

Similarly, good results were obtained for noncovalent intermolecular interactions. Considering the good performance for conformational energies, this is not unexpected because intra- and intermolecular noncovalent interactions are of the same physical origin and treated on an equal footing in QMDFF. The accurate and robust CM5 charge model and the first-principle D3 dispersion treatment are the main reasons for the good performance. The accuracy achieved for interaction energies in the small molecule test sets (S22, XB18, X40) of about 1 kcal mol^{-1} is good and also typical for semiempirical QM methods like DFTB or PM6. Dispersion-corrected DFT performs much better here with a typical deviation of 0.3−0.5 kcal mol^{-1}, which is clearly attributable to the inherently better account of short-range electrostatic and polarization effects. The 1 kcal mol^{-1} QMDFF accuracy level corresponds to a typical relative error of only 5−15%. By a slight rescaling of the atomic charges for polar systems this could lead to reasonably accurate bulk and solvation properties, a topic that will be addressed in a followup work. It is remarkable that the QMDFF performance does not degrade for the much larger supramolecular complexes for which a good 10−20% relative accuracy (5−10% for neutral cases) is obtained. These "real" host−guest complexes represent a challenge to any theoretical scheme, and in passing it is noted that UFF (which can be straightforwardly applied to this set) yields large errors and totally inconsistent results.

While further studies are certainly necessary to validate the entire QMDFF approach, the results presented herein are encouraging and the method seems to open a bright future for many fields where standard FFs are not existing. One of the critical points is the overall robustness of the method and its behavior in geometrically or chemically complicated bonding situations, a question that, however, can not be fully answered in this work but has to await continuous application in practice.

AUTHOR INFORMATION

Corresponding Author

*E-mail: grimme@thch.uni-bonn.de.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The author thanks Dr. J. Antony, C. Bannwarth, C. Bauer, G. Brandenburg, Dr. S. Ehrlich, Dr. A. Hansen, S. Johanning, Dr. P. Shushkov, Dr. M. Steinmetz, and R. Sure for helpful discussions and for performing some of the calculations. Part of the molecules in the transition metal benchmark set were kindly suggested by Dr. M. P. Checinski (CreativeQuantum GmbH, Berlin).

REFERENCES

(1) Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics*; Cambridge University Press: Cambridge, 2009.
(2) Berendsen, H. C. J. *Simulating the Physical World: Hierachical Modeling from Quantum Mechanics to fluid Dynamics*; Cambridge University Press: Cambridge, 2007.
(3) Reimers, J. R. *Computational Methods for Large Systems*; Wiley: Hoboken, NJ, 2011.
(4) Meier, K.; Choutko, A.; Dolenc, J.; Eichenberger, A. P.; Riniker, S.; van Gunsteren, W. F. *Angew. Chem., Int. Ed.* **2013**, *52*, 2820−2834.
(5) Jorgensen, W. L.; Tirado-Rives, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6665−6670.
(6) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.
(7) Acevedo, O.; Jorgensen, W. L. *Acc. Chem. Res.* **2010**, *43*, 142−151.
(8) The Nobel Prize in Chemistry 2013 was awarded to M. Karplus, M. Levitt, and A. Warshel for the development of multiscale models for complex chemical systems. See http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013/press.html (August 9, 2014).
(9) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133−A1138.
(10) Grimme, S. *WIREs Comput. Mol. Sci.* **2011**, *1*, 211−228.
(11) Zgarbova, M.; Otyepka, M.; Sponer, J.; Hobza, P.; Jurecka, P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 10476−10493.
(12) Li, J.; Lakshminarayanan, R.; Bai, Y.; Liu, S.; Zhou, L.; Pervushin, K.; Verma, C.; Beuerman, R. W. *J. Chem. Phys.* **2012**, *137*, 215101.
(13) Sponer, J.; Mladek, A.; Spackova, N.; Cang, X.; Cheatham, T. E.; Grimme, S. *J. Am. Chem. Soc.* **2013**, *135*, 9785−9796.
(14) Stendardo, E.; Pedone, A.; Cimino, P.; Cristina Menziani, M.; Crescenzi, O.; Barone, V. *Phys. Chem. Chem. Phys.* **2010**, *12*, 11697−11709.
(15) Schwörer, M.; Breitenfeld, B.; Tröster, P.; Bauer, S.; Lorenzen, K.; Tavan, P.; Mathias, G. *J. Chem. Phys.* **2013**, *138*, 244103.
(16) Bureekaew, S.; Amirjalayer, S.; Tafipolsky, M.; Spickermann, C.; Roy, T. K.; Schmid, R. *Phys. Status Solidi B* **2013**, *250*, 1128−1141.
(17) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2013**, *9*, 4046−4063.
(18) McDaniel, J. G.; Schmidt, J. R. *J. Phys. Chem. B* **2014**, *118*, 8042−8053.
(19) Jaramillo-Botero, A.; Naserifar, S.; Goddard, W. A., III *J. Chem. Theory. Comput.* **2014**, *10*, 1426−1439.
(20) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29*, 845−854.
(21) Salomon-Ferrer, R.; Case, D. A.; Walker, R. *WIREs Comput. Mol. Sci.* **2013**, *3*, 198−210.
(22) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr; 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. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(23) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Mobley, D. L.; Schnieders, M. J.; Haque, I.; Lambrecht, D. S.; DiStasio, R. A., Jr.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.

(24) Rappe, A. K.; Colwell, K. S.; Casewit, C. J. *Inorg. Chem.* **1993**, *32*, 3438−3450.
(25) Wiesemann, F.; Teipel, S.; Krebs, B.; Höweler, U. *Inorg. Chem.* **1994**, *33*, 1891−1898.
(26) Xiang, J. Y.; Ponder, J. W. *J. Chem. Theory Comput.* **2014**, *10*, 298−311.
(27) Stephan, D. W.; Erker, G. *Angew. Chem., Int. Ed.* **2010**, *49*, 46−76.
(28) Grimme, S.; Kruse, H.; Goerigk, L.; Erker, G. *Angew. Chem., Int. Ed.* **2010**, *49*, 1402−1405.
(29) Schirmer, B.; Grimme, S. *Top. Curr. Chem.* **2013**, *332*, 213−230.
(30) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.
(31) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165−169.
(32) *TURBOMOLE 6.6*; Universität Karlsruhe and Forschungszentrum Karlsruhe GmbH: Karlsruhe, Germany, 2014. http://www.turbomole.com (August 9, 2014).
(33) Stauch, T.; Dreuw, A. *J. Chem. Phys.* **2014**, *140*, 134107.
(34) van Duin, A. C.; Dasgupta, S.; Lorant, F.; Goddard, W. A., III. *J. Phys. Chem. A* **2001**, *105*, 9396−9409.
(35) Peverati, R.; Truhlar, D. G. *Philos. Trans. R. Soc. A* **2014**, *372*, 20120476.
(36) As suggested by an anonymous reviewer, the computation times for typical QMDFF calculations in comparison to a standard FF (amber03 or charmm27, no long-range interaction cutoffs in both calculations) were estimated by conducting 1 ns MD(NVT) simulations for $(ala)_n$ (n = 25 and 50). According to this test, QMDFF in this first version is a factor of about 2 slower than GROMACS. According to a performance analysis, the majority of this factor is attributed to the more complicated D3 dispersion terms compared to, e.g., a standard 6−12 LJ expression.
(37) Grimme, S. *Angew. Chem., Int. Ed.* **2013**, *52*, 6306−6312.
(38) Behler, J. *J. Phys. Cond. Matter* **2014**, *26*, 183001.
(39) Dasgupta, S.; Yamasaki, T.; Goddard, W. A., III. *J. Chem. Phys.* **1996**, *104*, 2898−2919.
(40) Maple, J.; Dinur, U.; Hagler, A. T. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 5350−5354.
(41) Dinur, U.; Hagler, A. T. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 1991; Vol. *2*, pp 99−164.
(42) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985−1993.
(43) Risthaus, T.; Grimme, S. *J. Chem. Theory Comput.* **2013**, *9*, 1580−1591.
(44) Goerigk, L.; Grimme, S. *J. Chem. Theory Comput.* **2011**, *7*, 291−309.
(45) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063−1079.
(46) Karton, A.; Yu, L.; Kesharwani, M. K.; Martin, J. M. L. *Theor. Chem. Acc.* **2014**, *133*, 1483.
(47) Risthaus, T.; Steinmetz, M.; Grimme, S. *J. Comput. Chem.* **2014**, *35*, 1509−1516.
(48) Stewart, J. J. P. *J. Mol. Mod.* **2007**, *13*, 1173.
(49) Korth, M.; Pitoňák, M.; Rezáč, J.; Hobza, P. *J. Chem. Theory Comput.* **2010**, *6*, 344−352.
(50) Gaus, M.; Goez, A.; Elstner, M. *J. Chem. Theory Comput.* **2013**, *9*, 338−354.
(51) Frauenheim, T. *DFTB+ (Density Functional based Tight Binding)*; DFTB.ORG, Universität Bremen: Bremen, Germany, 2008. http://www.dftb.org (August 9, 2014).
(52) Brandenburg, J. G.; Grimme, S. *J. Phys. Chem. Lett.* **2014**, *5*, 1785−1789.
(53) Mantina, M.; Valero, R.; Cramer, C. J.; Truhlar, D. G. In *CRC Handbook of Chemistry and Physics*, 91st ed.; Haynes, W. M., Ed.; CRC Press: Boca Raton, FL, 2010; pp 9-49−9-50.
(54) Wiberg, K. B. *Tetrahedron* **1968**, *24*, 1083−1096.
(55) Mayer, I. *Chem. Phys. Lett.* **1983**, *97*, 270.
(56) Bridgeman, A. J.; Cavigliasso, G.; Ireland, L. R.; Rothery, J. *J. Chem. Soc., Dalton Trans.* **2001**, 2095−2108.
(57) Allen, L. C. *J. Am. Chem. Soc.* **1989**, *111*, 9003−9014.
(58) Cremer, D.; Wu, A.; Larsson, A.; Kraka, E. *J. Mol. Model.* **2000**, *6*, 396−412.
(59) Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2000**, *10*, 6615−6620.
(60) Burger, S. K.; Ayers, P. W.; Schofield, J. *J. Comput. Chem.* **2014**, DOI: 10.1002/jcc.23636.
(61) Albright, T. A.; Hofmann, P.; Hoffmann, R. *J. Am. Chem. Soc.* **1977**, *99*, 7546−7557.
(62) Yan, J. F.; Momany, F. A.; Hoffmann, R.; Scheraga, H. A. *J. Phys. Chem.* **1970**, *74*, 420−433.
(63) Li, W.; Grimme, S.; Krieg, H.; Möllmann, J.; Zhang, J. *J. Phys. Chem. C* **2012**, *116*, 8865−8871.
(64) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
(65) Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(66) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *123*, 154101.
(67) van Gisbergen, S. J. A.; Snijders, J. G.; Baerends, E. J. *J. Chem. Phys.* **1995**, *103*, 9347−9354.
(68) Goerigk, L.; Kruse, H.; Grimme, S. *ChemPhysChem* **2011**, *12*, 3421−3433.
(69) Grimme, S. In *The Chemical Bond: Chemical Bonding Accross the Periodic Table*; Frenking, G., Shaik, S., Eds.; Wiley-VCH: Weinheim, 2014; pp 477−499.
(70) Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2005**, *1*, 153−168.
(71) Korth, M. *ChemPhysChem* **2011**, *12*, 3131−3142.
(72) Kozuch, S.; Martin, J. M. L. *J. Chem. Theory Comput.* **2013**, *9*, 1918−1931.
(73) Rezac, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2012**, *8*, 4285−4292.
(74) Clark, T. *WIREs Comput. Mol. Sci.* **2013**, *3*, 13−20.
(75) Mu, X.; Wang, Q.; Wang, L.-P.; Fried, S. D.; Piquemal, J.-P.; Dalby, K. N.; Ren, P. *J. Phys. Chem. B* **2014**, *118*, 6456−6465.
(76) Wang, B.; Truhlar, D. G. *J. Chem. Theory Comput.* **2010**, *6*, 3330−3342.
(77) Tafipolsky, M.; Engels, B. *J. Chem. Theory. Comput.* **2011**, *7*, 1791−1803.
(78) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2012**, *8*, 527−541.
(79) Hirshfeld, F. L. *Theor. Chim. Acta.* **1977**, *44*, 129−138.
(80) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868; *Phys. Rev. Lett.* **1997**, *78*, 1396 erratum.
(81) Klamt, A.; Schüürmann, G. *J. Chem. Soc. Perkin Trans. 2* **1993**, 799−805.
(82) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571−2577.
(83) Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2014**, *10*, 2802−2812.
(84) Takatani, T.; Hohenstein, E. G.; Malagoli, M.; Marshall, M. S.; Sherrill, C. D. *J. Chem. Phys.* **2010**, *132*, 144104.
(85) Levenberg, K. *Quart. Appl. Math.* **1944**, *2*, 164−168.
(86) Marquardt, D. *SIAM J. Appl. Math.* **1963**, *11*, 431−441.
(87) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.
(88) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.
(89) Neese, F. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73−78.
(90) Neese, F. *ORCA - An Ab Initio, DFT and Semiempirical electronic structure package*, Ver. 3.0 (Rev 4882); Max Planck Institute for Chemical Energy Conversion: Mühlheim, Germany, 2014.
(91) Grimme, S.; Steinmetz, M. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16031−16042.
(92) Grimme, S.; Hujo, W.; Kirchner, B. *Phys. Chem. Chem. Phys.* **2012**, *14*, 4875−4883.

(93) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158−6170.
(94) Sure, R.; Grimme, S. *J. Comput. Chem.* **2013**, *34*, 1672−1685.
(95) Stewart, R. F. *J. Chem. Phys.* **1970**, *52*, 431−438.
(96) Herman, A. *Modelling Simul. Mater. Sci. Eng.* **2004**, *12*, 21−32.
(97) Wolfsberg, M.; Helmholtz, L. *J. Chem. Phys.* **1952**, *20*, 837.
(98) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(99) Coutsias, E. A.; Seok, C.; Dill, K. A. *J. Comput. Chem.* **2004**, *25*, 1849−1857.
(100) Stewart, J. J. P. *MOPAC2012*; Stewart Computational Chemistry: Colorado Springs, CO, 2012. http://OpenMOPAC.net (August 9, 2014).
(101) Gilbert, K. E. *PCmodel 9.1*; Serena Software: Bloomington, IN, 2006. http://www.serenasoft.com/ (August 9, 2014).
(102) Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; M, L. R. *J. Comput. Chem.* **2005**, *26*, 1752.
(103) *Maestro*, version 9.8; Schrödinger, LLC: New York, NY, 2014. http://www.schrodinger.com/Maestro/ (August 9, 2014).
(104) *MacroModel*, version 10.4; Schrödinger, LLC: New York, NY, 2014. http://www.schrodinger.com/Macromodel/ (August 9, 2014).
(105) 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. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(106) Ferguson, D. M.; Kollman, P. A. *J. Comput. Chem.* **1991**, *12*, 620.
(107) McDonald, D. Q.; Still, W. C. *Tetrahedron Lett.* **1992**, *33*, 7743.
(108) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.
(109) Merrick, J. P.; Moran, D.; Radom, L. *J. Phys. Chem. A* **2007**, *111*, 11683−11700.
(110) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(111) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623−11627.
(112) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 5656−5667.
(113) Petrovic, P. V.; Grimme, S.; Zaric, S. D.; Pfeffer, M.; Djukic, J.-P. *Phys. Chem. Chem. Phys.* **2014**, *16*, 14688−14698.
(114) Moellmann, J.; Grimme, S. *Organometallics* **2013**, *32*, 3784−3787.
(115) Cioslowski, J.; Schimeczek, M.; Liu, G.; Stoyanov, V. *J. Phys. Chem. A* **2000**, *117*, 9377−9389.
(116) Grimme, S. *J. Am. Chem. Soc.* **1996**, *118*, 1529−1534.
(117) Karton, A.; Chan, B.; Raghavachari, K.; Radom, L. *J. Phys. Chem. A* **2013**, *117*, 1834−1842.
(118) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260−7268.
(119) Mládek, A.; Banas, P.; Jurecka, P.; Otyepka, M.; Zgarbova, M.; Sponer, J. *J. Chem. Theory Comput.* **2014**, *10*, 463−480.
(120) Fogueri, U. R.; Kozuch, S.; Karton, A.; Martin, J. M. *J. Phys. Chem. A* **2013**, *117*, 2269−2277.
(121) Zhong, H.; Carlson, H. A. *J. Chem. Theory Comput.* **2006**, *2*, 342−353.
(122) Soriano-Correa, C.; Olivares del Valle, F. J.; Munoz-Losa, A.; F. Galvan, I.; Martin, M. E.; Aguilar, M. A. *J. Phys. Chem. B* **2010**, *114*, 8961−8970.
(123) Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108.
(124) Reha, D.; Valdes, H.; Vondrasek, J.; Hobza, P.; Abu-Riziq, A.; Crews, B.; de Vries, M. S. *Chem.—Eur. J.* **2005**, *11*, 6803−6817.
(125) Grimme, S. *Chem.—Eur. J.* **2012**, *18*, 9955−9964.
(126) Ambrosetti, A.; Alfe, D.; DiStasio, R. A.; Tkatchenko, A. *J. Phys. Chem. Lett.* **2014**, *5*, 849−855.
(127) Weigend, F.; Furche, F.; Ahlrichs, R. *J. Chem. Phys.* **2003**, *119*, 12753−12762.