JCTC Journal of Chemical Theory and Computation

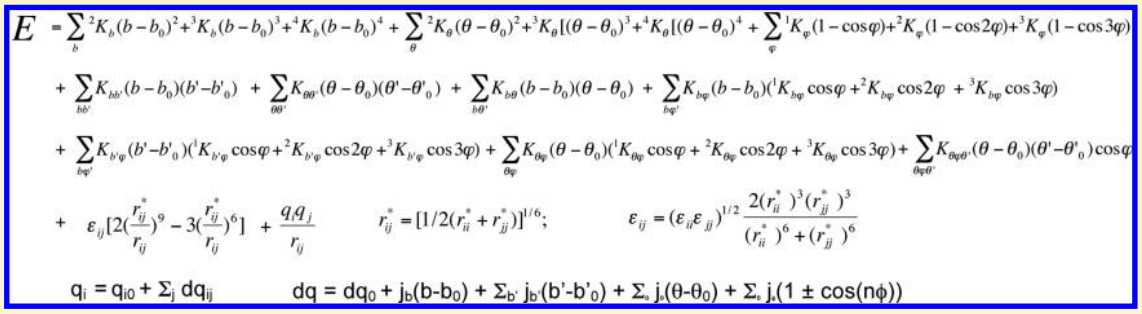
Perspective

pubs.acs.org/JCTC

Quantum Derivative Fitting and Biomolecular Force Fields: Functional Form, Coupling Terms, Charge Flux, Nonbond Anharmonicity, and Individual Dihedral Potentials

A. T. Hagler*

Department of Chemistry University of Massachusetts, Amherst, Massachusetts 01003, United States
Shifa Biopharm, Shifa Biomedical Corporation, Malvern, Pennsylvania 19355, United States



ABSTRACT: Computer simulations are increasingly prevalent, complementing experimental studies in all fields of biophysics, chemistry, and materials. Their utility, however, is critically dependent on the validity of the underlying force fields employed. In this Perspective we review the ability of quantum mechanics, and in particular analytical ab initio derivatives, to inform on the nature of intra- and intermolecular interactions. The power inherent in the exploitation of forces and second derivatives (Hessians) to derive force fields for a variety of compound types, including inorganic, organic, and biomolecules, is explored. We discuss the use of these quantities along with QM energies and geometries to determine force constants, including nonbond and electrostatic parameters, and to assess the functional form of the energy surface. The latter includes the optimal form of out-of-plane interactions and the necessity for anharmonicity, and terms to account for coupling between internals, to adequately represent the energy of intramolecular deformations. In addition, individual second derivatives of the energy with respect to selected interaction coordinates, such as interatomic distances or individual dihedral angles, have been shown to select out for the corresponding interactions, annihilating other interactions in the potential expression. Exploitation of these quantities allows one to probe the individual interaction and explore phenomena such as, for example, anisotropy of atom−atom nonbonded interactions, charge flux, or the functional form of isolated dihedral angles, e.g., a single dihedral X−C−C−Y about a tetrahedral C−C bond.

INTRODUCTION

In this Perspective we discuss the amplification of the power of quantum mechanical (QM) methods that may be achieved through the use of analytical first and second QM derivatives in force field derivation. We explore how these derivatives may be used to probe the validity of the functional form and how selected derivatives can shed light on important interactions, such as, charge flux, individual atom−atom nonbonded interactions and anharmonicity, and internal coupling, among others. Some of the latter are significant but often omitted, or ill determined, in applications.

Monte Carlo and molecular dynamics simulations are becoming an ever more ubiquitous adjunct to experimental studies, providing atomic resolution insight into structural, energetic, and dynamic properties of biomolecular and organic systems, as well as materials. The ultimate utility of these techniques rests on the faithfulness to which they recapitulate the experimental data. This in turn relies on the validity of the force field (FF)—or analytical representation of the energy surface. Quantum mechanics provides a valuable source of data to determine and test these analytical representations; especially as computational resources expand and allow more rigorous methods and basis sets to be exploited.

The applicability of quantum mechanics to biological systems was recognized almost from its inception as Pullman and Pullman,[1] Hoffman and Imamurs,[2] Hehre and Pople,[3] Scheraga et al.,[4] and others began applying semiempirical methods to peptide and nucleotide systems as early as the 1950s and 1960s. Results from these methods were soon incorporated into FF derivation, especially as a source of partial atomic charges[5,6] and also torsion potentials. The method of deriving charges from molecular electrostatic potentials introduced by Williams in 1981[7] based on the work of Scrocco and Tomasi[8] and popularized by the Kollman group[9] became one of the standard methods for derivation of atomic charges for biomolecular systems. In 1976 we

Received: July 10, 2015
Published: October 23, 2015

DOI: 10.1021/acs.jctc.5b00666
J. Chem. Theory Comput. 2015, 11, 5555−5572

invoked Hartree−Fock along with crystallographic structure determination for the first time, to calculate the critical rotational potential about φ and ψ.[10] Clementi was one of the first to derive intermolecular force fields based solely on QM data.[11,12]

In the late 1980s and 1990s advanced QM algorithms such as density functional theory (DFT) were introduced[13,14] and applied to biological and organic systems,[15,16] and computer power continued to increase dramatically. These developments coincided with the ever-growing realization of the need for force field improvement, and QM began playing an ever-increasing role in the derivation of second generation force fields. Thus, Lii and Allinger used ab initio methods to help parametrize a hydrogen-bond function in MM3,[17] and Cornell et al.[18] in their derivation of a second generation Amber force field in 1995 invoked quantum calculations of the φ and ψ map to refine torsion potentials. Jorgensen et al. also exploited QM results to derive torsion potentials in their development of the all-atom OPLS FF in 1996,[19] and in 1998 MacKerell et al. made extensive use of QM calculations of peptide conformational energies to develop their all-atom CHARMM potential.[20] Much of this work has been reviewed by Zhou et al.[21] in 2010 and more recently by Lopes et al.[22]

These advances opened up the possibility of utilizing the wealth information provided by QM about the energy surfaces of the molecules to develop a widely applicable FF. Major efforts in this field led to rigorously derived FFs such as that in eq 1, some of which accounted for anharmonicity and coupling between internals.

$$E = E_b + E_\theta + E_\varphi + E_{Coupling} + E_{Non-bonded} \quad (1)$$

$$E_b = \sum_b {}^2K_b(b-b_0)^2 + {}^3K_b(b-b_0)^3 + {}^4K_b(b-b_0)^4 \quad (a)$$

$$E_\theta = \sum_\theta {}^2K_\theta(\theta-\theta_0)^2 + {}^3K_\theta(\theta-\theta_0)^3 + {}^4K_\theta(\theta-\theta_0)^4 \quad (b)$$

$$E_\varphi = \sum_\varphi {}^1K_\varphi(1-\cos\varphi) + {}^2K_\varphi(1-\cos 2\varphi) + {}^3K_\varphi(1-\cos 3\varphi) \quad (c)$$

$$E_{Coupling} = E_{bb'} + E_{\theta\theta'} + E_{b\theta} + E_{b\varphi} + E_{b'\varphi} + E_{\varphi\theta} + E_{\theta\varphi\theta'} \quad (d)$$

$$E_{bb'} = \sum_{bb'} K_{bb'}(b-b_0)(b'-b'_0) \quad (e)$$

$$E_{\theta\theta'} = \sum_{\theta\theta'} K_{\theta\theta'}(\theta-\theta_0)(\theta'-\theta'_0) \quad (f)$$

$$E_{b\theta} = \sum_{b\theta'} K_{b\theta}(b-b_0)(\theta-\theta_0) \quad (g)$$

$$E_{b\varphi} = \sum_{b\varphi'} K_{b\varphi}(b-b_0)({}^1K_{b\varphi}\cos\varphi + {}^2K_{b\varphi}\cos 2\varphi + {}^3K_{b\varphi}\cos 3\varphi) \quad (h)$$

$$E_{b'\varphi} = \sum_{b\varphi'} K_{b'\varphi}(b'-b'_0)({}^1K_{b'\varphi}\cos\varphi + {}^2K_{b'\varphi}\cos 2\varphi + {}^3K_{b'\varphi}\cos 3\varphi) \quad (i)$$

$$E_{\theta\varphi} = \sum_{\theta\varphi} K_{\theta\varphi}(\theta-\theta_0)({}^1K_{\theta\varphi}\cos\varphi + {}^2K_{\theta\varphi}\cos 2\varphi + {}^3K_{\theta\varphi}\cos 3\varphi) \quad (j)$$

$$E_{\theta\varphi\theta'} = \sum_{\theta\varphi\theta'} K_{\theta\varphi\theta'}(\theta-\theta_0)(\theta'-\theta'_0)\cos\varphi \quad (k)$$

$$E_{Non-bond} = \sum_{i<j} E_{vdW}(r_{ij}) + \sum_{i<j} E_{Coul}(r_{ij}) \quad (l)$$

$$E_{vdW}(r_{ij}) = \varepsilon_{ij}[2(\frac{r^*_{ij}}{r_{ij}})^9 - 3(\frac{r^*_{ij}}{r_{ij}})^6] \quad (m)$$

$$E_{Coul}(r_{ij}) = \frac{q_i q_j}{r_{ij}} \quad (n)$$

$$r^*_{ij} = [1/2(r^*_{ii} + r^*_{jj})]^{1/6}; \quad \varepsilon_{ij} = (\varepsilon_{ii}\varepsilon_{jj})^{1/2}\frac{2(r^*_{ii})^3(r^*_{jj})^3}{(r^*_{ii})^6 + (r^*_{jj})^6} \quad (o)$$

In eq 1, E_b and E_θ are the strain energies associated with bond (b) and valence angle (θ) distortions from their reference values b_0 and θ_0; E_ϕ is the torsion angle (ϕ) energy; $E_{coupling}$ is the coupling energy between b, θ and ϕ; r^* and ε are the van der Waals (vdW) distance and well depth; and q is the partial charge.

While the QM calculations were a primary source of information in deriving FFs, it remained a necessity to verify (and adjust when needed) the derived FF with respect to experimental results, which ultimately are the only measure for success.

In the derivation of the FFs described earlier, high-level QM calculations were used to calculate structures and relative conformational energies, be they rotational barriers, cis−trans energy differences, or $\varphi-\psi$ maps. In some cases vibrational frequencies and intermolecular energies were also computed. These quantities were then used to determine constants in the force field.

PART I. HESSIANS AND FORCES IN THE DERIVATION OF FORCE FIELDS

Quantum Derivative Fitting. In the late 1980s we noted that the QM information on the characteristics of the molecular energy surface was contained not only in relative energies but also in the Cartesian derivatives of the energy. The latter provided a wealth of information and could be calculated cheaply; in fact they were often "free", being calculated in minimizations of energy and essentially discarded. This observation led to the development of "Quantum Derivative Fitting,"[23,24] (QDF) in which first and second derivatives of the energy (the force and Hessian, respectively), along with energy and structure, were incorporated into a robust algorithm to derive and test force fields. A package "Probe" was created which optimized the FF parameters to fit these quantities by least-squares, providing not only optimized parameters, but also statistical analyses of the results.[25]

Deriving Force Fields from QM Hessians, Forces, and Energies: A Novel Paradigm for Force Field Development. Based on these observations a new approach to the derivation of FFs for molecular simulations of biomolecules was developed at Biosym,[23] in a project supported by a Consortium for Research and Development of Potential Energy Functions.[26] We had two overall objectives in developing this methodology:

(1) The first, in common with all endeavors to derive FFs, was developing a technique to rapidly, objectively, and accurately determine reliable force constants for intermolecular and intramolecular interactions.

(2) The second objective was to develop a technique which would allow the rapid comparison of optimal force field *functional forms* for describing the energies of molecular deformations and interactions.

The essence of the method was to exploit the ab initio molecular energy surfaces, as described by the energy and its first and second derivatives with respect to coordinates. For a molecule with n atoms there are $3n$ first derivatives and $3n(3n+1)/2$ second derivatives, which describe the slope and curvature of the molecular energy surface.[23,24,27] Thus, for a compound with 10 atoms, a single point ab initio calculation with analytical derivatives returns, in addition to the energy, 30 first derivatives and 465 second derivatives, or roughly 500 bits of information about the energy surface in the region. This allows us to obtain a large amount of information about the potential energy surface in the sampled region with a relatively minimal computational effort. In practice the force field for a given family is obtained by fitting the distorted configurations of 15−20 model compounds containing a variety of strained conformations. This results in $\sim 10^5$ "quantum observables" which can be used to determine the FF, literally orders of magnitude more than available from only

DOI: 10.1021/acs.jctc.5b00666
J. Chem. Theory Comput. 2015, 11, 5555−5572

Table 1. Formate Ion Calculations[a]

force field	distorted configuration		equilibrium geometry		
	$\sqrt{(s/s_0)}$	ΔE_{rms}	Δr_{rms}	$\Delta\theta_{rms}$	$\lvert\Delta\nu\rvert$, av
diagonal					
FF1, quadratic (9)	0.298	2.18	0.17	0.27	155.6
FF2, cubic (13)	0.137	2.07	0.010	0.41	186.6
FF3, quartic (18)	0.125	2.05	0.009	0.62	172.3
with coupling terms					
FF4, second order (14)	0.278	1.00	0.010	0.15	88.3
FF5, third order (35)	0.032	0.13	0.002	0.42	31.1
FF6, third + quartic diagonals (40)	0.017	0.17	0.001	0.16	7.2
FF6 + different definitions of OOP					
pyramid height	0.017	0.17	0.001	0.16	7.2
improper torsion	0.032	0.53	0.001	0.45	13.4
Wilson et al.	0.018	0.21	0.000	0.11	11.8

[a] S_0 is the sum of squares of ab initio values in eq 2, energy is in kcal/mol, distances are in Å, angles are in degrees, and the frequencies are in cm^{-1}. Values in parentheses are the number of parameters. OOP is the out-of-plane angle.

QM energies and structures. Importantly, It also provides orders of magnitude more observables than FF parameters. The paucity of experimental and even standard quantum mechanical data for some families often leaves the force field parameters underdetermined.

Thus, rather than fitting equilibrium geometries and relative energies as is commonly done, the parameters in a given functional form of a force field were optimized by least-squares to minimize the sum of squared deviations (S), in energy, first and second derivatives, calculated from the FF from those calculated by ab initio, of the distorted structures of the model compounds (eq 2).

$$S = \sum W_E(E_{ff} - E_{qm})^2 + \sum W_D(\partial E_{ff}/\partial x_i - \partial E_{qm}/\partial x_i)^2 + \sum W_H[(\partial^2 E_{ff}/\partial x_i \partial x_j - \partial^2 E_{qm}/\partial x_i \partial x_j)^2 \quad (2)$$

where S is the (weighted) sum of squared deviations between the force field and ab initio values of the relative energies, and the energy first and second derivatives for all structures of all compounds in the training set. E_{ff} is the relative conformational energy as calculated with the force field (relative to the minimum energy); E_{qm} is the corresponding relative energy calculated by ab initio; x_i and x_j are the Cartesian coordinates of atoms i and j; and W_E, W_D, and W_H are weighting factors to balance the number of observables and their values (in the ratio of 150,000:100:1 for the deviations in energy, first derivative (forces), and second derivative, respectively[28]).

We note that if W_E and W_H are set to zero, eq 2 reduces to a fit of the forces sometimes called "force matching".[29] If W_E and W_D are set to zero, the result is a fit to simply the Hessians. The force field resulting from this optimization may then be further scaled to fit experimental data.[30]

The distorted conformations are not necessarily (or even desirably) minimum energy conformations, and thus only single point calculations need be done for each configuration. The distorted configurations are generated by random displacements of the internal coordinates about their reference values or by displacing them along the molecular normal modes (which can be obtained from a rough force field). In some applications, e.g., study of rotational barriers, the coordinates of interest may be systematically varied, while the other coordinates are distorted at random. This algorithm was incorporated into the Probe software package.[23−25]

Because the procedure is automated and parameters are optimized by least-squares, once the first and second derivatives are available it is easy to quickly test the ability of various functional forms to account for the shape of the quantum energy surface and even to test the importance of individual terms, such as, for example, anharmonicity in bond-stretching or particular coupling interaction terms.

We note that the basis sets that were used to demonstrate the basic protocols are quite primitive by current standards. It should be emphasized that both the derivation of FFs by fitting energies, forces, and Hessians and the theory for extracting individual atom−atom nonbonded interactions or dihedral terms from appropriate second derivatives of the QM dimer energies or total molecular energies are independent of the level of theory. We also note that in the former case the derived FF is scaled to fit experimental data. Thus, it may well be that smaller basis sets are desirable in this case for computational efficiency as the final arbiter of the quality of the FF is fit to experimental data. However, if the derived FF is to be used directly, one would exploit a much higher level of theory. This is also the case for the study of the isolated dihedrals and individual atom−atom nonbonded interactions where clearly correlation effects should be included.

Assessing the Importance of Coupling Terms. The method was demonstrated by fitting the ab initio surface of the formate anion ($HCOO^-$), determining optimal potential constants for various functional forms, and testing the accuracy of a variety of analytical representations (force field forms).[23] In addition alternate representations of the out-of-plane deformation for a trigonal center were evaluated. To accomplish this, eight different configurations of this molecular ion were generated by a random walk around a guessed reference point estimated to be near the (a priori unknown) equilibrium geometry. Ab initio 4-31G* Hartree−Fock calculations of the energies and the first and second derivatives were then carried out for the eight structures, yielding 720 Cartesian first and second derivatives and seven relative energies. The ab initio energy was minimized later to obtain the equilibrium geometry of the molecule and its energy, second derivatives, and the harmonic spectrum. As it turned out the energies of the eight distorted configurations spanned an energy range of 49 kcal/mol. Only the energy derivatives of the eight distorted configurations were used in the least-squares optimization of the potential constants. The configurational energies of the eight structures were not included in the optimization. Relative energies, equilibrium geometries, and the

DOI: 10.1021/acs.jctc.5b00666
J. Chem. Theory Comput. 2015, 11, 5555−5572

harmonic spectra were then used to test how well the optimized force fields could predict observables outside the domain of the fitting.[23]

Six force fields were tested with this system of coordinates. The first one, FF1, was a simple harmonic diagonal force field. The second force field, FF2, introduced anharmonicity in the form of cubic terms for all coordinates (except for the out-of-plane, which by symmetry has only even terms), and the third, FF3, extended the anharmonic contribution to quartic terms.

The next three force fields accounted for coupling between internals in the form of cross-terms (see, e.g., eq 1). The fourth, FF4, included quadratic cross-terms added to the quadratic diagonal force field, FF1, and likewise, FF5 and FF6 included cubic coupling terms added to FF2 and FF3 (quartic diagonal FF), respectively.

As expected, and seen in Table 1, more elaborate force fields fit the ab initio data progressively better, but importantly, some terms were more effective than others. The two most significant modifications were the addition of cubic terms to the harmonic diagonal force field, which led to a reduction of more than a factor of 4 in the sum of squares of residuals, S, and the further addition of cross-terms to the cubic diagonal force field, which reduced the sum of squares by an additional factor of 20.

As noted previously, none of the relative energies of the eight distorted configurations, the ab initio equilibrium geometry, or the harmonic spectrum at the equilibrium geometry were included in the fitting procedure and, consequently, the magnitude of the deviations in these quantities directly demonstrates the predictive power of the force fields (Table 1).

Inspection of the table immediately establishes the relative importance of different interactions. Thus, it is seen that inclusion of anharmonicity (through cubic and quartic terms) in a diagonal force field, while improving the fit to the derivatives, does not necessarily yield more accurate relative energies or more accurately predict the minimum energy geometry and, in this case, actually leads to a deterioration in the harmonic frequencies. On the other hand the quality of the force field is uniformly improved when coupling between internals is accounted for (cross-terms; see FF4−FF6). This is consistent with the demonstrated importance of these terms in accounting for energies and structures as well as frequencies in various FFs,[31−36] (see these references also for a discussion of the relative importance of different cross-terms). Thus, for example, on going from a diagonal quadratic FF (FF1) to a full quadratic force field including cross-terms (FF4), the deviation of the calculated energies from the ab initio values is reduced from 2.18 to 1.0 kcal/mol, and the deviation with respect to the harmonic spectrum is reduced from 156 to 88 cm^{-1}. It can be seen that we can get an almost quantitative reproduction of the quantum molecular energy surface over a wide energy range by including cubic coupling terms and quartic diagonal deformations (FF6), with average energy deviations of only 0.17 kcal/mol (in relative energies ranging from 5 to 49 kcal/mol), equilibrium bond lengths and angles predicted to within 0.001 Å and 0.16°, respectively, and frequencies calculated to within 7 cm^{-1}. *Basically we could carry out quantum level calculations of the properties of this compound using this force field.* (This suggests an automated methodology for getting potential functions of (any) quantum level accuracy for an arbitrary ligand with arbitrary functional groups.)

To further drive home the point that it is not simply a matter of adding parameters to achieve a better fit, we note that FF4, the quadratic FF with coupling terms, has 14 parameters but has significantly better predictive abilities and fit to molecular properties than the quartic diagonal force field (FF3), which has 18 parameters. *Furthermore, these results with respect to the cross-terms contradict the common belief that such terms are important for evaluating the harmonic spectrum but not for the energetics or structure of the molecule.* Again this makes sense since if the curvature of the energy surface is incorrect, it implies deviations in the value of the energy surface itself in some nearby regions. Furthermore, the vibrational frequencies represent the dynamics of the system, and thus if these are off, one cannot expect MD trajectories to be true.

Out-of-Plane Representation, Using the ab Initio Energy Surface To Probe the Functional Form of Potential Function Terms. Most force fields used in proteins and other biomolecular systems today define this coordinate as an improper torsion (Figure 1). This definition is somewhat artificial, and two,

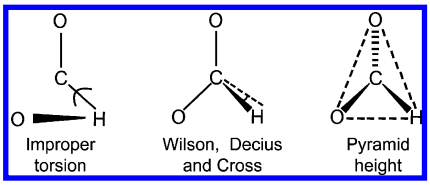


Figure 1. Out-of-plane coordinate.

possibly more realistic, descriptions of this motion are the height of the pyramid formed by the four atoms (i.e., the distance of the central atom to the plane of the other three atoms) or the angle between a bond and the plane formed by the other two bonds as given by Wilson et al.[37] The methodology described earlier gives an efficient and straightforward way to test these different representations.

The results obtained by including the three definitions in FF6 are given in the last three rows of Table 1. (Note FF6 used the pyramid height, so this line is repeated.) As seen in the table, differences are not drastic but the improper torsion definition results in a generally inferior fit and an energy deviation that is 2−3 times larger than those obtained with the other two definitions. One can conclude that the improper torsion angle should be replaced by either of the alternative definitions to describe out-of-plane deformations.[23]

Quantum Derivative Fitting: A Paradigm for Force Field Development. Quantum derivative fitting (QDF) was applied to derive the consistent force field (CFF) series of force fields by fitting sets of compounds from families representing proteins, other biomolecules, and polymers.[30,38−41] The overall strategy was to fit the energy surfaces of distorted conformations of a relatively large sample of compounds (~15−20) including highly strained compounds, three- and four- member rings, and so on.

Aside from the inclusion of many compounds, the procedure followed that described earlier for the formate ion. In a subsequent step, the resulting quantum force field (QMFF) was scaled to fit experimental data.[30] The quantum derivatives yielded enough information on the shape of the surface and relative values of the force constants that it was found that only a few scaling factors and reference values were needed to achieve a good fit to experiment,[30] thus greatly reducing the amount of experimental data required to derive the "experimental FF". This can be important for some of the less common functional groups, where experimental data can be lacking. This two-step methodology for deriving force fields can be considered a logical extension of Lifson's CFF method[31,42,43] for deriving force fields from as wide a

DOI: 10.1021/acs.jctc.5b00666
J. Chem. Theory Comput. 2015, 11, 5555−5572

variety of experimental data, including structures, energies, barriers, and frequencies, etc., as possible, because it first exploits the extensive data available from QM and then uses the same wide variety of experimentally available data, characterizing the CFF methodology.

The first family to be addressed was, not surprisingly, the hydrocarbons.[28,30,44] A total of 16 molecules were chosen to generate the energy surface for the derivation of the hydrocarbon force field. These included straight chain, branched, strained (e.g., neopentane), and cyclic compounds, *as well as six compounds containing three- and four-membered rings*. In total, the energy and derivatives of 196 distorted structures were calculated by ab initio, resulting in a total of 128,376 relative energies and first and second derivatives. (We note in the derivation of common simple diagonal force fields often only two to three compounds from each family have been used in the optimization, and properties of small rings are outside the ability of these force fields to account for.)

To test the force field, structures, frequencies, and energies of molecules selected both from those used in the original training set and from outside this set were calculated, and the values obtained from QM were compared with those derived from the classical analytical form (the force field).[44] *None of these properties were used directly in the original derivation of the force field.* In order to assess the importance of anharmonic and coupling interactions that occur in and contribute to molecular energy surfaces, the results are compared to a diagonal quadratic force field. The fit of the calculated structures, energies, and vibrational frequencies to the corresponding quantum properties is summarized in Table 2.

Table 2. RMS Deviations of Analytical FF Properties from Quantum Mechanical Properties

property	anharmonic, coupled FF	harmonic, diagonal FF
bonds (Å)	0.003	0.006
angles (deg)	0.4	0.8
torsions (deg)	1.2	2.3
energies (kcal/mol)	1.0	3.3
frequencies (cm^{-1})	34	100

The QMFF was shown to reproduce the bond lengths, bond angles, torsion angles, and conformational energy differences to an accuracy of 0.003 Å, 0.4°, 1.2°, and 1.0 kcal/mol, respectively. This compares to the corresponding deviations of 0.006 Å, 0.8°, 2.3°, and 3.3 kcal/mol for the harmonic diagonal force field.[30,44] Again, *of special note is the profound improvement in the fit to relative energies when coupling interactions are accounted for*. As noted, this is contrary to the common dogma in the field, though no evidence has been presented underlying this dogma. We suspect that omission of coupling interactions such as coupling between bonds or angles and torsions[28] is a major defect, in even the most recent biomolecular force fields.

Excluding three- or four-membered rings, the QMFF root mean square (rms) frequency deviations were 24 cm^{-1}, which again is much better than the ∼100 cm^{-1} deviations for the harmonic diagonal force field. Larger average rms frequency deviations of 71 and 36 cm^{-1} were found with QMFF for molecules with three- and four-membered rings, respectively.[44] This may be compared to the deviation in the frequencies calculated with the common diagonal force fields, where typical deviations in frequencies for simple unstrained compounds are also often of the order of 100 cm^{-1} and larger, and energy deviations of 2−3 kcal mol are also encountered.

Transferability. The force constants in this work were found to be totally transferable. That is, they depend on whether the bonded atoms in the internal coordinates are alkyl carbons or hydrogens, but they do not depend on any other details of the substructural environments in which they are found. Thus, they are independent of whether carbon is primary, secondary, or tertiary. For example, unlike essentially all other FFs, including the MM4 and MMFF94 force fields, the torsion parameters for HCCH, HCCC, and CCCC are all the same, depending only on the nature of the central bond. Furthermore, this force field with coupling terms and anharmonicity accounts for the properties of three- and four-membered rings, with no modification of parameters. *It is the only functional form to date able to account for both these strained rings and normal hydrocarbons simultaneously.* In addition to the importance of these small rings in many compounds of interest (see below), the ability of the FF to account for them demonstrates the degree of strain and range of applicability of the FF. Since our goal today is to simulate the molecular dynamics of complex systems including protein−ligand systems, it is important that the FF can accurately account for the strain induced both by dynamic fluctuations and by ligand binding, which may be considerable.

Scaling the QM Energy Surface To Account for Experimental Properties. In the work described previously, it was shown that one could use QDF to determine the detailed functional form required to describe molecular energy landscapes, i.e., to reveal the importance of various quantities such as out-of-plane representations, anharmonicity, and coupling between internals. These have long been the subject of discussion and had generally been difficult to determine in the past. This is especially true in functional groups other than alkanes, because of a dearth of experimental data. In a follow up study[30] it was demonstrated that scaling of the analytical representation, derived from the quantum mechanical energy surface, could be carried out to fit experimental data. This is analogous to the practice of scaling QM vibrational frequencies to account for experimental spectra. Importantly, *it was found that it was sufficient to determine only a single scale factor for each class of internal deformation.* Thus, the angle, torsion, cross-terms, and out-of-plane terms each had one scaling factor associated with them for all types of angles, torsions, and so on. In addition to force constant scale factors some slight adjustments to bond reference values, b_o, were required, resulting in a total of only seven parameters to be determined from the experimental data. *Furthermore, and equally important, work on different functional groups indicated an additional advantage in that the scale factors may be transferable as well.*

To demonstrate the power of the process and the large observable-to-parameter ratio achievable, we review the paradigm as applied to the alkanes. As can be seen in Table 3, only a subset of available experimental observables were needed to determine the scale factors for the quantum force field: 13 bond lengths and 199 (125 nondegenerate) frequencies were chosen. This is because only five scale factors and two reference values needed to be determined (Table 4). Thus, even with this extremely limited data set an observable to parameter ratio of ∼20:1 was achieved, far higher than obtainable in a standard parameter derivation. In addition, no energetic, structural, or conformational observables, with the exception of bond lengths, were included in the training set. This left them as a powerful test of the resulting force field and validity of the overall shape of the quantum energy surface and the functional form representing it.[30]

Validity of Analytical Representation of Experimental Energy Landscape. Is the Shape of the Quantum Energy

DOI: 10.1021/acs.jctc.5b00666
J. Chem. Theory Comput. 2015, 11, 5555−5572

Table 3. Number of Experimental Observables in the Training Set for Determination of Scale Factors

	no. of observables	
compound	bond lengths	frequency (nondegenerate)
methane	1	9 (4)
ethane	2	18 (12)
n-butane	2	36 (24)
isobutane	2	36 (24)
cyclopropane	2	21(14)
cyclobutane	2	30 (22)
cyclohexane	2	41 (25)
total	13	199 (125)

Table 4. Seven Scale Constants and Reference Values for Hydrocarbon Force Field

Scaling Constants	
bond stretching:	
S_b(C−C)	0.88
S_b(C−H)	0.83
angle bending: S_θ	0.81
torsion: S_ϕ	0.84
cross-terms: S_{xt}	0.87
Reference Values	
b_0(C−C)	1.535
b_0(C−H)	1.111

Surface True to the Experimental Surface? The validity of the use of quantum data in force field derivation is an issue for development of essentially all current force fields.[21,22,45,46] Of course, with increasing computational resources and the ability to use larger basis sets, these calculations are continually improving. To test the hypothesis that the quantum energy surface accurately mirrors the topography of the experimental energy surface, the experimental geometries, energy conformational differences, and rotational energy barriers, as well as experimental vibrational frequencies for small simple alkanes, including the small rings, were calculated. This requires a less stringent criterion of the ab initio surface than direct use of the QM results, as it requires relative, rather than absolute, accuracy of the surface as is implied when the QM data are used directly. A subset of the results is given in Table 5.[30] It is important to note that *none* of the energetic properties or structural properties (not shown), with the exception of bond lengths, were used in the fit of the scaling factors, and none of these observables, or even their quantum equivalents, were used in the fit of the underlying quantum FF. Despite this the scaled CFF (with fewer parameters) accounted for these properties as well as or better than common harmonic FFs, derived specifically to fit these very observables.

These results, and other similar comparisons, more than validate the hypothesis that the ab initio surface reflects the properties of the experimental surface. Moreover the FF is unique in its ability to account for the properties of small rings with no additional parameters. Both MMFF[48] and MM4,[49] for example, invoke a different set of parameters for four-membered rings and MMFF also requires different parameters for five-membered rings. Neither treats three-membered rings.

Continuity of Energy Surfaces and Ability To Fit Strained Configurations. Implication of Three- and Four-Membered Rings: Are They Merely Theoretical Curiosities or of Significant Practical Importance? The results on the nature of the quantum energy surface and the ability of the anharmonic FF with coupling terms to account for the strained three- and four- membered rings are of fundamental significance. It implies that the energy surface over the range of these significant distortions is continuous and that the function does not have singularities at some strain points, where there are discontinuous transitions in the value of the force constants. A discontinuity is implied when different force constants are invoked for these small rings. For example on going from a CCC angle of 109.5°, characteristic of normal hydrocarbons, to the 90° in cyclobutane to the 60° angle in cyclopropane, it is implied that at some point the force constant changed abruptly from the normal value to the small ring value.

This lack of different force constants is intuitively attractive and of great importance as it gives us confidence that the FF is valid over the domains of distortion likely to be encountered in MD simulations. These may be induced in substructural components due either to distortion imposed by covalent constraints as in cyclic peptide hormones, by requirements to fulfill tertiary structure requirements in proteins, or in ligands experiencing

Table 5. Comparison of Experimental and Calculated Conformational Energies, Rotational Barriers, and Vibrational Frequencies for Simple Hydrocarbons

compound	property	experiment	CFF93	MM3[32]	AMBER[47]
methane	$\Delta\nu$[a]	(0)	28		73
ethane	barrier	2.88	2.75	2.41	3.01
ethane	$\Delta\nu$	(0)	22	58	93
propane	barrier	3.4	3.11		3.27
cyclopropane	$\Delta\nu$	(0)	51	132	173
n-butane	barrier syn-rot	4.36	4.95	4.83	5.31
n-butane	barrier t-g[b]	3.3	3.47	3.3	3.53
n-butane	ΔE(t-g)	05−0.9	0.72	0.81	0.79
n-butane	$\Delta\nu$	(0)	25	41	78
isobutane	barrier	3.9	3.51		3.48
cyclobutane	$\Delta\nu$	(0)	41		102
pentane	ΔE(a-a/a-g)[c]	0.47−0.57	0.72	0.81	0.76
pentane	ΔE(a-a/g-g)	1.36[d]	1.38	1.62	1.48
neopentane	barrier	4.2−4.8	3.95	3.35	3.94
cyclohexane	$\Delta\nu$	(0)	51	132	173

[a] $\Delta\nu$, rms deviation from experimental values of vibrational frequencies. [b] t-g, trans-gauche. [c] a-g, anti-gauche. [d] Ab Initio (MP3).

DOI: 10.1021/acs.jctc.5b00666
J. Chem. Theory Comput. 2015, 11, 5555−5572

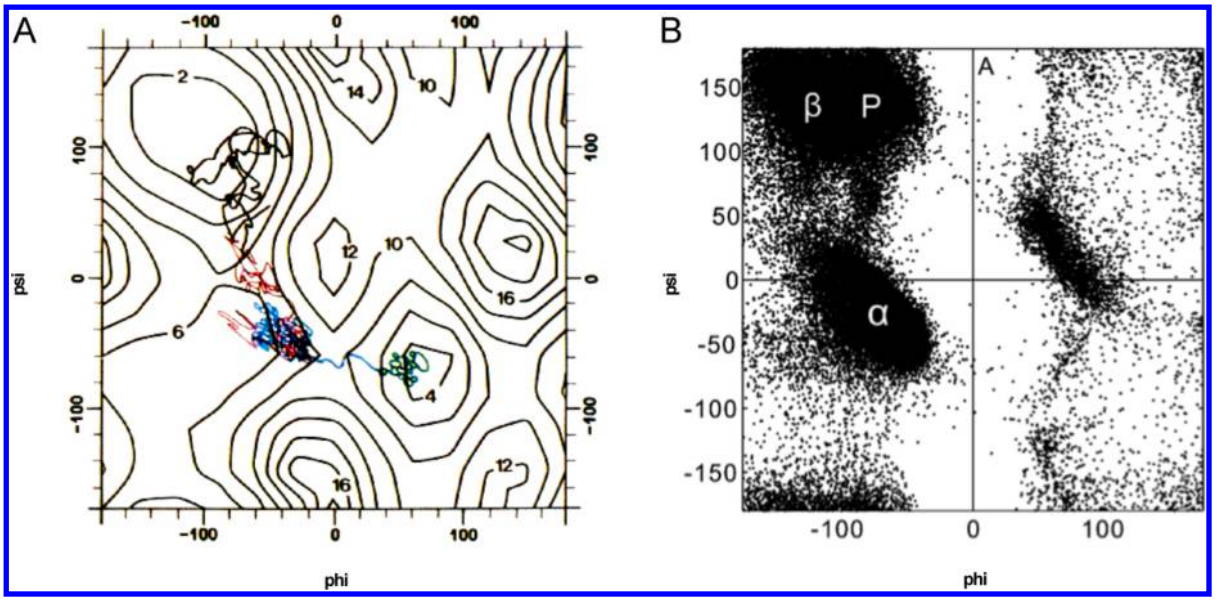


Figure 2. (A) Dynamic trajectory of the Phe residue in lysine vasopressin superposed on a $\varphi-\psi$ energy map of the energy showing the conformations visited (and their energies) during the simulation. As can be seen, strain energies as high as 8 kcal are visited. Time evolution is represented by colors, with black, red, blue, and green lines in the trajectory representing sequential 2 ps intervals. From ref 50. Reprinted with permission from AAAS. (B) $\varphi-\psi$ distribution of residues in 500 nonhomologous proteins, demonstrating the strain induced in individual residues on folding. Reprinted with permission of 51. Copyright 2008 PLOS.

induced strain on binding. Two examples of this behavior are demonstrated in Figure 2. Figure 2A shows the strain energy in the Phe residue in the cyclic hexapeptide portion of the hormone lysine vasopressin, as it takes part in the trajectory traversed by the nonapeptide. Note the energy map corresponds to the isolated residue energy map, while the trajectory of the Phe is determined by the energy of the entire peptide within which it resides. Local strain energies of 8 kcal/mol are encountered in this residue, even over a very short (8 ps) dynamic trajectory.[50] Clearly this is due to both the constraints imposed by the covalent structure as well as the dynamic fluctuations in the entire nonapeptide. (Barriers of this size would not be crossed in the isolated system in this time scale; and in fact, correlated transitions between residues in the full peptide are observed, presumably modulating the total energy.[50])

In Figure 2B, the $\varphi-\psi$ distribution of residues in 500 nonhomologous proteins as compiled by Wu and Zhang is presented.[51] Here we see a similar phenomenon, where in this case the local strain (which can be seen by comparing with the familiar $\varphi-\psi$ energy map) is induced in residues by the fold of the protein, and perhaps ligand binding. Thus, even in equilibrium structures, portions of the protein experience considerable strain.

Of equal practical importance, many drug candidates, natural compounds, and marketed drugs contain three- and four-membered rings often in fused heterocyclic systems. Thus, in both docking and simulations it is mandatory to have a FF that can adequately treat these systems. This can be seen for example in Table 6, where the results of substructure searches of the ZINC database, a large database of millions of purchasable compounds for docking,[52] the MDDR, a database of ∼150 K biologically relevant compounds (http://accelrys.com/products/databases/bioactivity/mddr.html), and the TCM database, containing ∼23,000 compounds found in Chinese Medicinal Herbs,[53] are presented. As can be seen from this table, small rings are common in drugs and libraries used to screen for novel compounds—requiring FFs to be able to account for these systems in virtual screening and simulations.

Table 6. Prevalence of Biologically Relevant Compounds Containing Three- and Four-Membered Rings

ring size	zinc	MDDR	TCM
three	738,365	4,967	1,449
four	90,592	6,379	284
five	8,337,435	55,616	11,419

The need for different sets of parameters is a major disadvantage when it comes to treating fused rings as described later. What parameters does one assign to the common internals shared by the small ring and the "normal" fragments of the molecule?

The difficulty of simulating drugs and other biologically active compounds containing these moieties with potential functions that do not extend to small rings may be brought home by examining some examples, several of which are displayed in Figure 3. Ciprofloxacin (Figure 3A) is one of the most well-known members of the fluoroquinolone family of antibiotics,[54] many of which, like Cipro, contain the cyclopropyl ring. They treat a variety of infections including urinary tract infections and anthrax (and are also being investigated for use in cancer.[55] The mitomycin family of compounds are natural products containing a fused three-membered azridine ring. Mitomycin C displayed in Figure 3B is a toxic antibiotic and antitumor agent used as a chemotherapeutic in a variety of cancers.[56,57] The thiirane derivatives possess mainly antimicrobial and cytotoxic activities[58] [e.g., SB-3CT (Figure 3C)], a potent and selective inhibitor of matrix metalloproteinases 2 and 9 (MMP-2 and MMP-9) geltinases.[59] The examples of four-membered-ring drugs in Figure 2 include the following: The ubiquitous β-lactam antibiotics in Figure 3D, including the penicillins and cephalosporins, contain a variety of fused heterocyclic four-membered rings; the antidepressant Zalospirone,[60] Figure 3E, containing the fused, unsaturated cyclobutene ring; and Stadol,[61]

DOI: 10.1021/acs.jctc.5b00666
J. Chem. Theory Comput. 2015, 11, 5555−5572

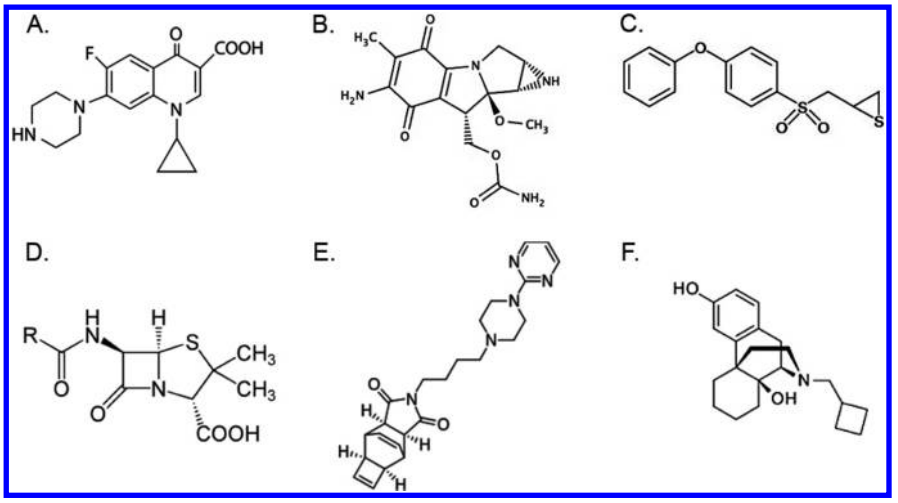


Figure 3. Examples of biologically active compounds and drugs containing three- and four-membered rings: (A) ciprofloxacin, a widely used antibiotic; (B) mitomycin C, a chemotherapeutic and antibiotic agent; (C) SB-3CT, a member of the thiirane family possessing antimicrobial and cytotoxic properties; (D) β-lactam antibiotics including penicillins and cephalosporins; (E) zalospirone, an antidepressant; and (F) stadol, an opioid.

Figure 3F, containing a cyclobutane ring, a morphine-like opioid used for migraines, among other indications.

A cursory examination of such drugs emphasizes the need to be able to account for small rings, especially fused systems, when calculating binding affinities, and especially the contributions of induced strain energies in docking. These fused ring systems emphasize the practical problem in the use of different FFs for different sized rings (Figure 3).

Quantum Derivative Fitting Paradigm Exploited for Diverse Systems. The methodology described previously provides a paradigm for force field development, where the ultimate goal is to determine force fields for a wide range of molecules. Norrby et al.[62] in their derivation of a force field for oxymethylpyridines outlined "four distinct advantages of the method" as,

> *"(1) The number of possible geometries to use in the parametrization is effectively limited only by available computer time; (2) the parameter refinement is well suited for least-squares methods; (3) by judicious choice of low-energy geometries, it is possible to cover the whole range of attainable values for each structural element in the molecule (e.g., bond lengths, bond angles, dihedral angles, nonbonded distances) ...; and (4) no geometry optimization is required in the ab initio calculations. The last point in particular increases the number of accessible data points dramatically if available computer time is a limiting factor."*

In addition, faced with a new functional group in a system where experimental data are sparse, the methodology outlined earlier can be invoked. Either the derived QMFF energy functions and constants can be used directly, yielding a molecular mechanics or dynamics result of comparable quality to that which would result if carried out on the quantum energy surface, or scaling factors can be derived from the limited experimental data, or transferred. Ever more powerful computational resources as well as increasingly more accurate quantum mechanical methods offer the possibility that at some point the derived QMFF itself will approach experimental accuracy.

Following its original formulation[23] the approach was adopted in a wide variety of applications. Several adaptations were introduced as discussed later, some using a subset of the information, for example either the Hessian alone or only the first derivatives ("force matching"), and some combined experimental data with the derivatives in a single optimization of parameters. Again, in one of the quirks of science where often related fields remain insular, the vast majority of applications and exploitation occurred in the polymers and materials field, while aside from Krimm et al.[36] and Halgren,[48] who exploited the Probe software used in derivation of CFF93[23,40] and an earlier work by Dasgupta and Goddard,[63] the robustness of the technique has escaped the notice of the mainstream biomolecular FF community.

In an early application, Dasgupta and Goddard combined the Hessian from QM calculations with experimental structural and spectroscopic data to derive FFs, utilizing formaldehyde, formate anion, thioformaldehyde, carbonyl chloride, and carbonyl fluoride as model systems.[63] Though a complete review of all subsequent applications of derivative methodology is beyond the scope of this discussion, we review some of the key developments and more recent work, and the reader is referred to references therein for further applications.

Norrby and Liljefors[64] briefly reviewed the methodology at the time, and automated the parametrization method, fitting quantum and experimental data simultaneously. They demonstrated the technique by developing parameters for ethane, and compared the use of a variety of ab initio basis sets. Ercolessi and Adams[29] developed a similar algorithm utilizing only first derivatives, which they coined "force-matching", and applied it to determining the intermolecular interactions in aluminum crystals. Hill and Sauer[65] exploited Probe[23] to derive a full CFF for silica and zeolite catalysts from quantum derived energies and first and second energy derivatives.[65]

In a compelling series of studies, Krimm and co-workers included ab initio Hessians along with a variety of other properties to determine their SDFF force field for organic molecules and proteins.[36,66−68] This force field is noteworthy for several reasons. It is the most complete and rigorous force field determined to date for the treatment of biomolecular systems. Krimm et al. not only accounted for multipole interactions but also included a complete set of coupling terms (e.g., eq 1) and accounted for polarizability and charge flux. (Unfortunately it was not extended to a wide range of functional groups.)

Following the application to alkanes,[28] CFF force fields were derived for amides and peptides,[38] and carbohydrates,[39] and then

DOI: 10.1021/acs.jctc.5b00666
J. Chem. Theory Comput. 2015, 11, 5555−5572

extended to a large number (32) of additional functional groups.[40] Energies and first and second derivatives of distorted structures, as well as equilibrium structures and transition states of ∼400 compounds, were examined. This resulted in a total of ∼1.3 M quantum observables, against which to fit the parameters. They demonstrated a good fit to the properties of this extensive set of families, comparable to that achieved with alkanes.

Tafipolsky and Schmid,[69] addressing metal−carboxylate interactions, incorporated the Hessian and structure from high-level ab initio calculations in order to derive force fields for these complexes. More recently Waldher et al.[70] constructed a program and GUI to fit FF parameters to QM energies and gradients. It is a flexible program able to parametrize a variety of functional forms. It does require an associated MM/MD program, which needs to be run in order to obtain the first derivatives of the energy to input to the optimization procedure.

Song and Rhee[71] extended the approach of fitting second derivatives (solely at the minimum energy structure) to derive force field parameters in both ground and excited states. They then used these parameters to study differences in dynamics in the two states of oxyluciferin in aqueous solution. An ab initio analysis of an array of gallium nitride clusters was carried out with a variety of high-level basis sets.[72] The results were used to derive a FF based on Hessian matrices and employed to carry out an MD study of GaN, which was able to account for the heat capacity of the crystal. Biczysko et al.[73] demonstrated the use of quantum mechanical studies, along with continuum solvent approaches and a FF derived from the QM Hessian and structure, to investigate the properties of pyrimidine in the gas phase as well as in aqueous and CCl_4 solutions.

In the past year or so, there appears to be an increase in the applications utilizing these techniques. For example Jorn et al.[74] applied a least-squares technique to determine the parameters for a nonaqueous solution of $LiPF_6$ in ethylene carbonate, by selecting random configurations from an MD simulation of the system and requiring that the first derivatives derived from the potential match to those calculated from QM. Barone et al.[75] have created a flexible algorithm, along with a GUI, to optimize force fields for a single molecule. As in Probe[23,28] both first and second derivatives can be employed, and the technique was also applied to the first excited state of nicotine. Also noteworthy is the inclusion in the force field of polarizability and the complete complement of coupling terms.

An algorithm for deriving specialized FFs for single molecules from a single QM derived Hessian has been presented by Grimme[76] and demonstrated with applications to a variety of organic and organometallic compounds. A noteworthy feature of this application is the use of a "generalized Lennard-Jones potential" for bond stretching, which allows for dissociation of the bonds. In a recent elegant application, the use of first derivatives has been combined with experimental data, to optimize the AMOEBA water potential.[46,77] Though combined methods have been used previously, as discussed earlier, here the group not only used condensed phase properties but also calculated derivatives of the simulated properties with respect to parameters to allow least-squares optimization.[46,77] Finally, very recently, Limé and Norrby adapted the method of fitting Hessians to derive a force field for transition states.[78] They found, among other things, that coupling terms were required to account for these significantly distorted states, and adopted those found in MM3.[32]

PART II. THEORY OF DERIVATIVES: SPECIFIC FORCE CONSTANT ANALYSIS OF INDIVIDUAL FORCE CONSTANTS THROUGH SELECTED QUANTUM ENERGY SECOND DERIVATIVES

In addition to the use of derivatives for sampling of the quantum mechanical energy surface to derive quantum mechanical force fields, we can use the properties of specific second derivatives to probe individual terms within the force field such as nonbonded interactions. This is a powerful methodology, which allows us to isolate *heretofore inaccessible* individual interactions.

Basically we take advantage of the fact that by looking at the force on a particular coordinate (which is simply the negative of the derivative), we effectively probe the local molecular response to a specific perturbation. In general, nuclear positions or other internals that contribute to the energy surface *but are not coupled or related to a given coordinate* do not contribute to the force acting on that coordinate and are thus eliminated from its mathematical expression.

A specific second derivative of the energy then further extracts a specific interaction from the total energy. It in turn basically probes the local response of the force to the specific perturbation and thus removes all interactions not quadratic in the specific coordinate perturbed. For example $\partial^2 E/\partial b_i^2$ extracts the stretching constants $K^n(b_i)$ from eq1 (although, because it is impossible to change a bond length without changing some of the nonbonded distances, the second derivative picks up contributions from the nonbonded interactions as well). To demonstrate this methodology, we will examine the relatively straightforward bond-stretching function as a simple example.

Example with Individual Bond-Stretching Term Probed by Second Derivative. The first derivative of the energy with respect to a given bond length, $\partial E/\partial b_i$, is given in eq 3. It is a well-defined quantity that is a linear combination of Cartesian first derivatives of the energy, and as noted above, *these derivatives are readily accessible from molecular orbital calculations.* As seen from eq 3, the derivative of the energy with respect to bond i, b_i, selects out only those terms containing this bond. Thus, it is independent of all other bonds, angle, torsion, and out-of-plane terms (E_{b_j}, E_θ, E_φ, and E_χ), although this derivative does depend on other coordinates, through the coupling functions such as $E_{bb}(b_i b_j)$, $E_{b\theta}(b_i\theta)$, and so on that contain the bond b_i. In addition nonbond terms survive, because the distance between atoms on either side of the bond b_i, varies as the bond length is changed.

$$\begin{aligned}\partial E/\partial b_i &= \partial(E_b + E_\theta + E_\varphi + E_{\text{coupling}} + E_{\text{nonbonded}})/\partial b_i \\ &= \partial E_b(b_i)/\partial b_i + \partial E_{bb'}(b_i, b_j)/\partial b_i + \partial E_{b\theta}(b_i, \theta)/\partial b_i \\ &\quad + \partial E_{b\varphi}(b_i, \phi)/\partial b_i + \partial E_{\text{nonbonded}}/\partial b_i\end{aligned} \tag{3}$$

If we now go on to take the second derivative $\partial^2 E/\partial b_i^2$, the number of contributing interaction terms is even further reduced because now we probe the change in the force on the ith bond due to a change in its own length (eq 4). Thus, all components of the potential energy that are not at least quadratic in the bond stretch are eliminated.

$$\begin{aligned}\partial^2 E/\partial b_i^2 &= \partial^2 E_b(b_i)/\partial b_i^2 + \partial^2 E_{\text{nonbonded}}/\partial b_i^2 \\ &= {}^2K_{b_i} + {}^3K_{b_i}(b_i - b_{i0}) + {}^4K_{b_i}(b_i - b_{0_i})^2 \\ &\quad + \partial^2 E_{\text{nonbonded}}/\partial b_i^2\end{aligned} \tag{4}$$

DOI: 10.1021/acs.jctc.5b00666
J. Chem. Theory Comput. 2015, 11, 5555−5572

A general pattern emerges. A specific second derivative of the energy extracts a specific interaction from the total energy. Thus, $\partial^2 E/\partial b_i^2$ as shown earlier, extracts the stretching constants from eq 1, although, as noted, because it is impossible to change a bond length without changing nonbonded distances on either side of the bond, the second derivative picks up contributions from the nonbonded interactions as well.

In this case, the nonbonded constants are usually much smaller than the bond-stretching constants, so that the difference may not be that significant. The crucially important point for now, however, is that calculating $\partial^2 E/\partial b_i^2$ at several different bond lengths *yields a curve that, to a very good approximation, is just the second derivative of the bond-stretching energy* E_b. *That is, to second order, the second derivative extracts the bond-stretching function from the total quantum energy surface.*

This is demonstrated in Figure 4, which shows the C−H stretching constant $\partial^2 E/\partial b^2$(C−H) plotted as a function of the

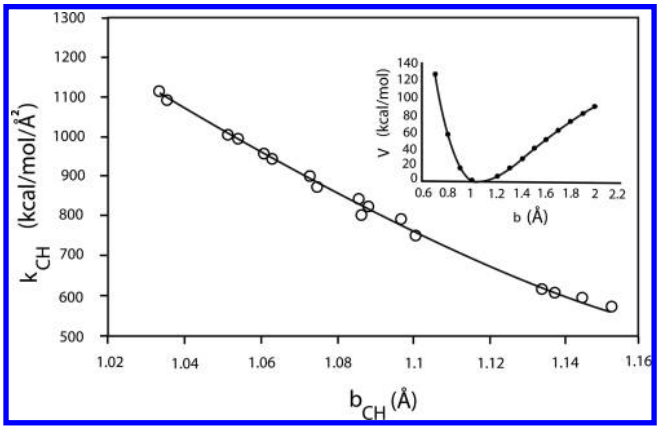


Figure 4. C−H force constants ($\partial^2 E/\partial b_i^2$) in ethane calculated at the 6-31G* level as a function of the C−H bond length in the equilibrium configuration and two distorted configurations. Points are ab initio, and the curve corresponds to the second derivatives calculated from a Morse potential with parameters derived by a least-squares fit to the points. Inset: Bond-stretching energy as derived from the derived Morse potential. Note that whereas the force constant is not constant (*as it would be for a purely harmonic system*), it is changing in a nearly linear manner over this range.

C−H bond length for 13 C−H distance.[24] These are the results of calculations performed on three ethane configurations—the equilibrium structure and two distorted structures taken from the set of QM calculations described previously.[28,30]

We can now fit the constants in the expression for the second derivative given in eq 4 to these points and obtain the values for the force constants for the C−H bond stretch, ${}^{i}K_b$(C−H). (In this case the nonbond potential was omitted, as dispersion is not accounted for by HF and the contribution of exchange repulsion is expected to be small.) The resulting fit is shown as the continuous curve in Figure 4. We note that in this case a Morse function was actually fit rather than the quartic representation, but, of course, the principle is identical. The corresponding C−H bond-stretching potential obtained is displayed in the inset.[24]

Probing the Functional Form. Figure 4 also clearly shows how this analysis can inform as to the functional form of a particular term in the potential energy surface. Thus, in this simple case it is clear that the C−H bonds are strongly anharmonic—the quadratic force constant, $\partial^2 E/\partial b^2_{(C-H)}$, is reduced by 40% as the bond length changes by 0.1 Å. (Of course for the commonly used harmonic spring representation, $E_b = K_b(b - b_0)^2$, $\partial^2 E/\partial b^2$ would be constant, and the curve in Figure 4 would simply be a horizontal line.) Though this is a trivial example, as it is well-known that bond stretching is anharmonic; it demonstrates the power of the second derivative to yield unambiguous information, not only about the value of the force constants but also on the functional form of the energy term. The same technique can be used to gain information on other terms such as torsion and nonbond interactions, which are not well characterized and often inaccessible by other techniques, as shown later.

As seen, the points are slightly scattered about the continuous curve. This is due to the fact that along with the CH bonds all other coordinates were also changed, and both higher order terms starting at third order and perhaps the non bonded terms also contribute slightly.

It should be stressed that *it is the second derivative that allows extraction of the specific interaction*, in this case, the bond-stretching term (here modulated by nonbonded and third and higher order coupling terms). This contrasts to the approach where one attempts to parametrize E_b by changing just one C−H bond length and fitting the energy along this coordinate; contributions from other interactions would be picked up, for example bond−bond, bond−angle, and so on cross-terms, that also change with the increase in the bond stretch thus contaminating the derived value.

Direct Evaluation of Individual Atom−Atom Nonbonding Interactions from ab Initio Calculations of Second Derivatives. Though determination of bond stretching serves the purpose of demonstrating the method, it is not as pivotal or challenging as, for example, the conformationally and energetically critical nonbonded interactions. These interactions, composed of electrostatic, steric, and dispersion contributions are crucial to the calculation and understanding of the structure and energetics of interactions of biomolecular systems, crystal packing, and liquid properties.[22,46,79−82]

Thus, the appropriate second derivative of the energy, in this case of dimers, can be utilized in a similar manner to explore the important atom−atom nonbonded interactions in biomolecular and pharmacologic functional groups.[24,83−85] Characteristic partial atomic charges and higher order multipoles,[86−89] as well as van der Waals and polarization effects of particular interatomic interactions, can all be studied through this technique.[24,83] In addition, the ab initio derivatives in directly probing the nature of the atom−atom interactions yield information on transferability, anisotropy of the nonbonding repulsion, polarizability, and charge flux, i.e., the reaction of partial charges (electron distribution) to distortion of molecular geometry. In the repulsive regime they can also probe the recently discussed effects of short-range electrostatic penetration.[90]

We can demonstrate how the second derivative extracts the nature of the atom−atom nonbonding interactions by considering a simple system, the water−water dimer (here taken in the D_{2h} configuration (Figure 5).

For the sake of discussion the first water molecule is labeled "A" and the second "B". The energy for this system can then be represented by

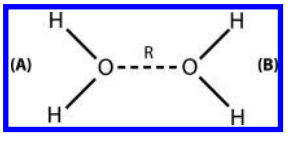


Figure 5. D_{2h} water dimer.

DOI: 10.1021/acs.jctc.5b00666
J. Chem. Theory Comput. 2015, 11, 5555−5572

$$E_{AB} = E_{Intra}(A) + E_{intra}(B) + E_{inter}(H_A - H_B) + E_{inter}(O_A - O_B) + E_{inter}(O_A - H_B) + E_{inter}(O_B - H_A) \quad (5)$$

where the first two terms are the intramolecular energies of waters A and B respectively; the third term represents the interaction between the hydrogens in molecule A with those in molecule B and similarly for the oxygens in the fourth term. The final two terms represent the interactions of the oxygen in one molecule with the hydrogens in the other.

If we use the same overall strategy as that with the bond function earlier and take the first derivative of the energy with respect to, say, the x coordinate of the oxygen in water A, then it follows that

$$\partial E_{AB}/\partial x_{OA} = \partial E_{intra}(A)/\partial x_{OA} + \partial E_{inter}(O_A{-}O_B)/\partial x_{OA} + \partial E_{inter}(O_A{-}H_B)/\partial x_{OA} \quad (6)$$

where the terms involving the intramolecular energy of molecule B and the nonbonded interactions between the hydrogens on the two waters and between O_B and H_A, are all eliminated as they do not depend on the coordinates of O_A.

If we now take the second derivative with respect to the x-coordinate of O_B (i.e., $\partial(\partial E_{AB}/\partial x_{OA})/\partial x_{OB}$), and assume for simplicity the commonly used 12−6 Lennard-Jones potential with a simple Coulombs law to represent electrostatics as in eq 1 (i.e., $E_{LJ} = A_{ij}/r_{ij}^{12} - C_{ij}^{6}/r_{ij}^{6}$; $E_{elec} = q_i q_j/ r_{ij}$), then the second derivative with respect to x_{OA} and x_{OB} becomes

$$\partial^2 E_{AB}/\partial x_{OB}\partial x_{OA} = \partial^2 E_{inter}(O_A{-}O_B)/\partial x_{OB}\partial x_{OA} = 156A_{O-O}/r_{OAOB}^{14} - 56C_{O-O}/r_{OAOB}^{8} + 2q_O q_O/r_{OAOB}^{3} \quad (7)$$

That is, the only term that remains is the oxygen−oxygen interaction. The second derivative with respect to the coordinates of the two oxygens *gives direct information on the pairwise interaction between the two oxygens*.

It then follows that the key to exploring the properties of isolated pairwise nonbonded interactions is to fit the ab initio second derivatives with an analytical representation of the nonbonding interaction and assess the ability of this analytic representation to describe the energy surface as reflected in these derivatives.[83]

The protocol described earlier was applied to characterize the O···O nonbonded interaction.[83] Hartree−Fock calculations were carried out on the water dimer in Figure 5 at various intermolecular distances, and the second derivatives of the energy were extracted. (Of course, the protocol can be carried out at any level of QM calculation.) The values of the Lennard-Jones and Coulomb parameters for the O−O intermolecular interaction were then determined. A linear least-squares method was carried out to fit analytical second derivative (as given in eq 7) to the ab initio values. Since the dispersion interactions are not accounted for by Hartree−Fock calculations, the R^{-6} term was omitted from fitting the ab initio second derivatives.

Charge Flux and Polarizability. The first, perhaps disturbing, observation that emerged from the analysis of the fit to these quantum mechanical energy derivatives was that the resulting partial charges and characteristic repulsive parameters which emerged *were significantly different depending on the direction of the derivative*. This of course cannot occur based on the assumptions underlying all current analytic potential functions. That is, the parameters resulting from the fit to the Hessian elements, H_{xx}^{OO} ($\partial^2 E_{AB}/\partial x_{OB}\partial x_{OB}$), H_{yy}^{OO}, and H_{zz}^{OO}, differ substantially, as seen in Figure 6.[83] This figure shows the charge q,

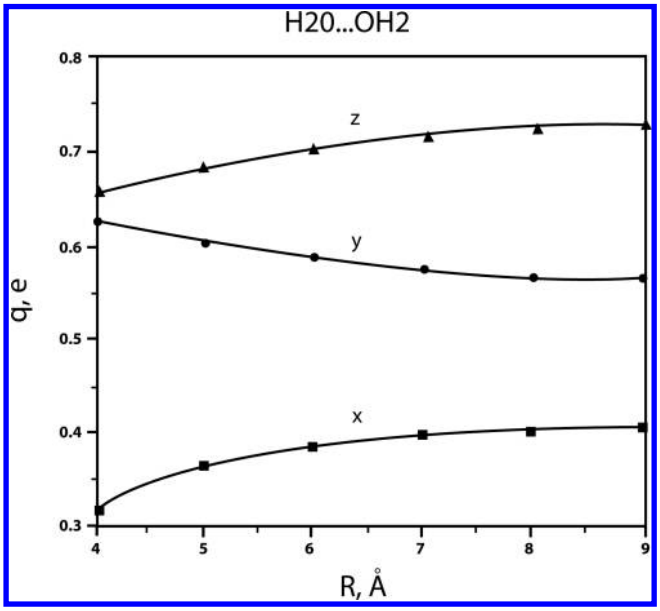


Figure 6. Effective partial atomic charge on the oxygen in H_2O as derived from the water dimer in Figure 5. The abscissa is the distance between the two oxygens. The curves marked x, y, z correspond to the Hessian elements H_{xx}^{OO}, H_{yy}^{OO}, and H_{zz}^{OO}, respectively. At large distances the repulsion is negligible and the interaction becomes purely electrostatic; i.e., $H_{xx}^{OO} \rightarrow q^2/R^3$.

as obtained from the three Cartesian components of the Hessian as a function of distance between the oxygens. At large distances the repulsion becomes negligible and the interaction becomes purely electrostatic—allowing the partial atomic charges to be determined from the values of the second derivatives in this range.

Analysis of the discrepancy between the "charges" reveals that the difference is due to the charge flux. That is, if we determine the electrostatic force on O_B due to the $O_A \cdots O_B$ interaction, it is actually written as

$$\partial E_{Coul}(O_A{-}O_B)/\partial x_{OB} = \partial(q_{OA}q_{OB}/r_{OAOB})/\partial x_{OB} = 1/r_{OO}[\partial \mathbf{q}_{OA}/\partial \mathbf{x}_{OB} + \partial \mathbf{q}_{OB}/\partial \mathbf{x}_{OB}] - q_{OA}q_{OB}/r_{OO}^{2}\ldots \quad (8)$$

(More rigorously we should include the change in charge on the hydrogens with respect to the oxygen position, but for the sake of demonstrating the phenomena, we omit these terms). The first two terms account for the reorganization in charge in O_A and O_B, respectively, due to a change in the position of O_B, *and reflects the fact that the charges are, in fact, not constant but depend on the atomic coordinates*. Note polarization is the change in charge on an atom due to an external field while charge flux is the fluctuation in atomic charge due to distortions in intramolecular geometry; *thus the first term, $\partial q_{OA}/\partial x_{OB}$, corresponds to polarization, while the second, $\partial q_{OB}/\partial x_{OB}$, corresponds to the charge flux*). It has been found that the effects of the flux are significantly larger than those from polarizability.[91]

DOI: 10.1021/acs.jctc.5b00666
J. Chem. Theory Comput. 2015, 11, 5555−5572

The properties of the derivatives of the molecular dipole, their interpretation as atomic charges, and the geometric dependence of charge distributions (charge flux) has been recognized and studied extensively for many years, particularly with regard to IR intensities.[92−94] More recently charge flux for a variety of molecular families has been investigated to study H-bonds,[95] the electronic structural changes underlying multipolar derivatives and flux[96] and with an eye toward its incorporation into force fields for more accurate simulations[36,85,97] as discussed later. In addition there is a huge amount of literature on polarization and its inclusion in force fields. The reader is referred to the many excellent reviews on this subject as well as to the inclusion of atomic multipoles to describe electrostatic interactions.[22,79,98−103]

Further analysis of the QM derivatives in Figure 6 reveals that the major effect is in fact due to the charge flux and that the effect of polarizability is negligible.[83] This may be shown as well by considering the derivative of the dipole moment of an isolated water monomer with respect to the same coordinate, again calculated from QM (using the same basis set). For consistency we will take this monomer as water B in the preceding discussion.

Since the molecular dipole moment is defined as $\mu = \Sigma(q_i x_i + q_i y_i + q_i z_i)$, where x, y, and z are the Cartesian coordinates of atom i, we can write this derivative as

$$\partial \mu_B / \partial x_{OB} = q_{OB} + x_{OB} \partial q_{OB} / \partial x_{OB} \ldots \quad (9)$$

where we have again omitted the derivatives of the charge on the hydrogens with respect to change in x_{OB} for simplicity. Calculation of the three derivatives $\partial \mu_B / \partial x_{OB}$, $\partial \mu_B / \partial y_{OB}$, and $\partial \mu_B / \partial z_{OB}$ of the monomer reveals that the values again diverge *and are virtually identical to those derived from the corresponding components of the second energy derivatives calculated for the dimer* given in Figure 6. This confirms that the fluctuation in charge in the dimer is due to the charge flux, with polarization effects being negligible (since here we are considering the isolated monomer where there is no contribution of an external field from intermolecular interactions).

Determination of Charge Distributions. Though not an emphasis of this Perspective, we note that eqs 7 and 9 suggest a direct method for obtaining partial atomic charges (and, in fact, higher multipoles). As seen later this is strictly true only for planar molecules where the derivatives are taken perpendicular to the plane of planar molecules (eliminating charge flux). On the other hand they do provide information on both the multipole interactions and charge flux, which provides the opportunity to extract both. Thus, this is not as robust a procedure for obtaining charges as, for example, the elegant methods based on partitioning of electron densities either from crystallography, such as in Hirshfeld's method,[104] or ab initio, as in the widely used distributed multipole method introduced by Stone.[88] Nor of course are they as easily implemented as the widely applied method of fitting the molecular electrostatic potential (MEP).[8,105]

The disadvantages of the latter methods lie in the fact that there is no unique way to partition molecular electron density, and thus these methods are model dependent. Furthermore, there is no guarantee that these charges will reproduce the molecular multipoles, forces, and field (calculated from the same wave function). In fact a more straightforward method (than that of the energy derivatives), i.e., obtaining atomic charges and higher atomic multipoles from derivatives of the quantum mechanical molecular multipoles, has been presented.[94,96,97,106] This is a natural definition as can be seen for the classical case of the charge as the derivative of the molecular dipole in eq 9. Analogous definitions hold for all atomic multipoles as the derivatives of the next higher molecular multipoles.[94,106] (Though rigorously this only holds for derivatives perpendicular to the plane of planar molecules, as noted earlier, due to the absence of the flux contributions in this direction). Though these quantum derivative methods are not as robust as the cited methods, they have the advantage that the atomic multipoles derived from these derivatives are molecular invariants and, thus, more amenable to studies of trends and transferability than atomic multipoles that are obtained with alternative methods. Furthermore, it can be shown that this set of charges and atomic multipoles not only recapitulate the molecular multipoles *and external field* but also are uniquely appropriate for calculating interatomic forces.

Given these properties, it would seem appropriate to compare charges and multipoles derived from other methods with those obtained by the multipole derivatives—at least for planar molecules. A preliminary comparison of charges and dipoles obtained from derivatives of molecular multipoles, electrostatic potentials, and population analysis has been carried out and their ability to account for the MEP at distances from 2.5−5 Å compared.[106] The qualitative values were similar though derivative derived multipoles gave somewhat better fits to the electrostatic field.[106]

Geometry Dependent Charges and Inclusion of Charge Flux in Force Fields. Equation 8 recapitulates the well-recognized fact that charges do in fact vary with molecular geometry. Thus, in fact, as noted earlier, eq 9 is more rigorously written as

$$\partial \mu / \partial x_j = \sum \partial (q_i x_i) / \partial x_j = q_i + \sum x_i \, \partial q_i / \partial x_j \quad (10)$$

where the derivatives $\partial q \partial x$, referred to as charge flux, reflect the geometry dependence of the atomic charges.

Though this geometry dependence has been known for quite awhile,[22,36,107−112] along with the fact that this charge flux is a major contribution to intermolecular forces crucial to MD,[109,113] the geometry dependence of charges has not been widely incorporated into force fields. In part to address this need, a simple and computationally inexpensive formalism was developed, to include the effect of charge flux and geometrically dependent charges in MM and MD simulations.[85] The methodology was applied to alkanes, aldehydes, ketones, and amides.

Examination of the results in planar molecules, where the charges and charge flux can be defined exactly,[85,94,106] along with results of IR spectra[114] suggest that the major contribution to the charge flux is due to distortion of the local internals in which the atom is contained. Given this observation, the geometric dependence of the bond increment 116 (charge flux) was expressed in terms of the length of the bond, as well as the internals surrounding the bond[85] (eq 11)

$$dq = dq_0 + j_b(b - b_0) + \sum_{b'} j_{b'}(b' - b'_0) + \sum_{\theta} j_\theta(\theta - \theta_0) + \sum_{\theta} j_\phi(1 \pm \cos(n\phi)) \quad (11)$$

where b is the bond, b' is any bond connected to b, θ is any valence angle that contains the bond, *and* ϕ is any torsion angle that contains the bond, b. b_0 and θ_0 are reference values, and the sign in front of the cosine term depends on the periodicity, n. The constant dq_0 is the intrinsic or equilibrium bond increment for that bond, j_b is the charge flux constant for distortions along that bond, $j_{b'}$, and j_θ and j_ϕ are the corresponding constants for the flux

DOI: 10.1021/acs.jctc.5b00666
J. Chem. Theory Comput. 2015, 11, 5555−5572

Table 7. Results from Least-Squares Fit of Charge and Flux Parameters to Dipoles and Dipole Derivatives of Alkanes, Aldehydes, Ketones, and Amides

group	observables	parameters	S_{rel}[a]	S_{μ}[a]	$\Delta\Theta$[a]	$\langle\mu^2\rangle^{1/2}_{QM}$[a]
alkanes	6681	32	0.165	0.085	4.96	0.084
aldehydes/ketones	6927	70	0.139	0.084	1.05	3.00
amides	5523	36	0.058	0.15	1.03	4.04

[a] S_{rel} is the rms relative deviation, S_{μ} is the rms deviation of the dipole moment, $\Delta\Theta$ is the rms deviation in the angle of the dipole, and $\langle\mu^2\rangle^{1/2}_{QM}$ is the rms value of the total dipole moment over all configurations from QM.

along bond b due to distortions of the adjacent bonds, angles, and torsions.

The atomic charges are obtained from the charge increments by[115]

$$q_i = q_{i0} + \sum_j dq_{ij} \tag{12}$$

where q_{i0} is a reference charge on atom i and the sum is over all bonds, j, containing atom i. The expression for the charges as obtained from eqs 11 and 12 can then be substituted into eq 1, thus including charge flux in the force field. This approach was taken by Krimm et al.[36] in their derivation of the SDFF for proteins.

Deriving the Flux Constants. To derive the flux constants in a set of distorted configurations of alkanes, aldehydes, ketones, and amides were selected and the ab initio molecular dipoles and dipole derivatives were calculated.[85] The constants in eq 11 were then obtained by fitting these dipole moments and dipole moment derivatives calculated by substituting the expressions for q as obtained from eqs 11 and 12 into eq 10.[85] That is by optimizing the sum of squares of the differences in the ab initio and analytical expressions for these quantities as given in eq 13 with respect to the all charge and charge flux parameters (in eq 11).

$$S^2 = \sum \omega(\mu_{calc} - \mu_{qm})^2 + \sum (\nabla\mu_{calc} - \nabla\mu_{qm})^2 \tag{13}$$

A total of 200 configurations of 12 compounds for these families were used resulting in a training set of ~19,000 dipole and dipole derivative observables from which a total of 138 charge and flux constants were derived (or a parameter to observable ratio of 1:138).

The results of the fit are given in Table 7. As can be seen the results for the polar compounds are better than those for the alkanes, with a roughly 3% relative deviation in the dipole moment ($S_{\mu}/\langle\mu^2\rangle^{1/2}$). Analysis of the results further indicated the effect of torsion flux to be relatively small, though significant in the amides.

The parameters derived in the fit were then applied to fit the dipole surface of molecules outside the training set. A slight degradation in fit was observed (e.g., an average deviation in dipole moment of 0.15 D corresponding to a relative deviation of 4% in relative dipole moment vs ~3% in the training set), which was attributed to the intrinsic inaccuracy in transferability as well as to the assumption of linearity and truncation of the flux series.

Charge Flux: Constituting 40% or More of Force, of the Same Magnitude as Coulomb Interaction. As discussed earlier, charge flux makes a significant contribution to the derivative of the Coulomb interaction with respect to the atomic coordinates, which is the electrostatic force. In fact this contribution can be huge, accounting for 40% or more of the force.[85] This can be seen in Figure 7 where the force on one of the carbon atoms, C, of N-formylformamide, exerted by a point charge, P, of +0.1, is calculated from QM at various C−P

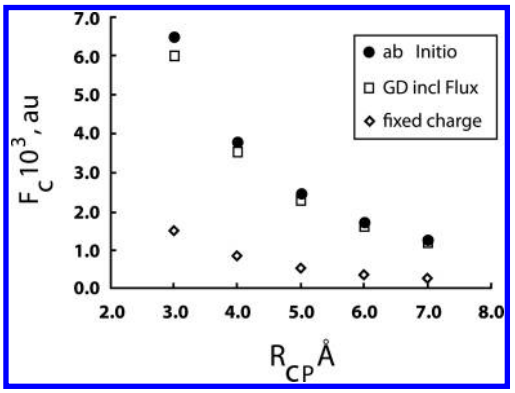


Figure 7. Forces exerted by an external point charge on a carbon atom in N-formylformamide. R_{CP} is the distance between the proton and the carbon atom, and F_c is the force on the carbon atom. "GD incl flux" is the force calculated by substituting the charges calculated by eq 11 into the Coulomb term, and fixed charge are the forces calculated by using potential derived charges in the Coulomb term.

distances. These forces (filled circles in Figure 7) are then compared to those derived from eq 11 including the effect of flux (open squares) and from fixed charges calculated by fitting the electrostatic potential, PD charges (open diamonds). It is clear from the results in Figure 7 that forces calculated from the Coulomb interaction with fixed charges result in major errors while those calculated including flux perform well. This is especially significant since the PD charges were optimized for N-formylformamide itself, while the constants in eq 11 were transferred from the fit to the amides, ketones, and aldehydes discussed earlier.[85]

Results on water demonstrate similar major contributions of charge flux to the forces resulting from intermolecular interactions.[113] Here it was shown that the error associated with the neglect of the charge flux was of the order of −80%. (It was pointed out that, as in the results presented previously, the constant ratio of in-plane vs out-of-plane forces as a function of distance show that these effects are not due to polarization since the latter would decay at large distances.) In addition, although these results were obtained with relatively small basis sets by today's standards (6-31G* or 6-31G**), they are in qualitative agreement with higher level calculations as well as with the analysis of experimental IR intensities of water.[68,116,117]

Krimm et al. also adopted the formalism developed by Dinur and Hagler as embodied in eq 11 for the development of their SDFF alluded to earlier.[67,68] Their studies also showed charge flux effects on structural changes in peptides including pyramidalization of the amide nitrogen and peptide torsion angles.[68]

Anisotropy of Nonbond Parameters: Determination from Energy Second Derivatives. With few exceptions[118,119] essentially all current force fields, including all those commonly used in biomolecular simulations, assume isotropic van der Waals interactions.[22,100] That this assumption was questionable was

DOI: 10.1021/acs.jctc.5b00666
J. Chem. Theory Comput. 2015, 11, 5555−5572

noted by Bondi[120] in his classic assessment of van der Waals radii. The issue was further investigated by Nyburg and Faerman[121] who exploited the large array of crystal structures made available by the Cambridge Crystallographic Database. These authors found anisotropies in the shape of atoms as much as 20% for sulfur, for example, as measured by the difference in the major and minor atomic elliptical axes, and suggested these effects can have significant consequences in packing and energetics.

The information contained in the second derivatives gives us a powerful method to probe this effect.

To this end, a calculation was carried out of ab initio energy derivatives for the usual planar H-bonded water dimer (C_s symmetry), and the oxygen−oxygen interaction was extracted through the second derivatives as described previously. The results were then compared to those obtained for the D_{2h} dimer[83] in Figure 5. The value of the repulsive constant, A_{OO}, found in the hydrogen-bonded orientation, is 220,000 kcal/(mol Å^{12}), *or 4 times the value found for the head-to-head orientation* (61,552 kcal/(mol Å^{12})). Thus, a huge anisotropy is found, which could have significant implications for the representation of these interactions in biomolecular simulations. This could be due to an inherent anisotropy in the atoms' electron distribution caused by chemical bonding or to polarization arising from the hydrogen-bond interaction.

Anisotropy of van der Waals radii has been further verified by both crystallographic analysis and other QM calculations.[122−124] Recent studies[125−127] continue to lend credence to the suggestion by Stone and Price[118] that

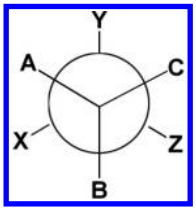
> *"... models of the repulsion which treat molecules as assemblies of spherical atoms are fundamentally inadequate. Atoms in molecules are not spherical, and ... (anisotropy) may amount to 10−15% of the van der Waals radius and have a significant effect on properties because of the way that they modify packing preferences"*

Incorporation of Anisotropy, Penetration, and Charge Transfer into Force Fields. Price and co-workers have developed an analytical model to represent anisotropic nonbond potential and applied them to lattice energy calculations and packing predictions of a variety of crystal systems.[89,118,128,129] In an early study they applied intermolecular perturbation theory to 300 random geometries of cyanuric acid dimers to parametrize exchange-repulsion, penetration, and charge-transfer interactions. These were then used to refine an analytical expression including a cylindrically symmetric anisotropic nonbonded repulsion to these interactions.[130] The resulting force field was then applied to lattice energy calculations of cyanuric chloride and was shown to account for the packing better than isotropic nonbond potentials. The force field has been implemented in the crystal simulation program DMACRYS, which also accounts for polarization and atomic multipole interactions.[89]

Determination of Individual Dihedral Potential Functions Including One- and Two-fold Contributions. One of the most important, yet vexing, terms in force fields is the torsion contribution. This can be seen simply by reflecting on the difficulty encountered in refining the potentials about the critical ϕ and ψ angles in proteins.[131,132]

Experimental data such as energy barriers and vibrational frequencies are sparse and where available do not contain enough information to enable decomposition of a dihedral potential into its various Fourier components. In addition, the torsional coordinate is composed of several dihedral components (for an sp^3-sp^3 bond there are nine dihedral components as depicted in Scheme 1 (A−C−C−X, A−C−C−Y ,and A−C−C−Z, etc.), and

Scheme 1



thus there is redundancy of dihedral coordinates that further makes it difficult to obtain reliable potentials for these dihedrals. The energy expression in force fields is invariably expressed in terms of these individual dihedral potentials, thus further emphasizing the need for determination of these individual components. Fortunately we can use the same second derivative methodology outlined earlier to probe the energetics and determine functional forms and potential constants for these elusive terms. In the following the individual dihedral angles A−C−C−X and so on are denoted by τ_{AX} to distinguish them from the torsion angel ϕ about the C−C bond (often taken as the average of all individual dihedrals).

Following a procedure similar to that used to derive eq 4, it can be shown that the second derivative of the energy (eq 1) with respect to an individual dihedral angle τ(A−C−C−X), τ_{AX}, is given by[133]

$$\partial^2 E/\partial\tau_{AX}{}^2 = \partial^2 E_\phi/\partial\tau_{AX}{}^2 + \partial^2 E_{nonbonded}/\partial\tau_{AX}{}^2 \quad (14)$$

That is, as in the case of the bond stretch, the second derivative annihilates all terms not involving the torsion angles itself, or the nonbond interactions between atoms A and X. (The dihedral derivative can be derived from the Hessian by a suitable transformation of the Cartesian derivative.[133]) Utilizing the procedure described earlier for the bond and nonbond interactions, the constants for the torsion potential can then be obtained by fitting the analytical torsion derivatives calculated from the FF to the corresponding ab initio values.

An example of the information which may be extracted from this second derivative is given in Table 7, where the Fourier components as obtained from the fit to the values of the second derivatives of the H−C−C−H dihedral angle in different configurations of ethane (A, X = H), are given.[133]

It is evident from Table 8 that there is a large one-fold term in the expansion of E_τ(H−C−C−H) in ethane. This one-fold term as well as all terms with multiplicity different from n = 3 do not appear in the total torsional energy profile (E_ϕ) of ethane because

Table 8. Fourier Expansion Coefficients (kcal/mol) of the H−C−C−H Dihedral Potentials in Ethane Obtained by a Least-Squares Fit to ab Initio Second Derivatives (at the 6-31G* Level)[133]

n	K_n
1	1.51
2	0.10
3	0.17
4	0.01
5	0.00
6	0.00

[a] $E_\tau = K_n[1 - \cos(n\tau)]$

DOI: 10.1021/acs.jctc.5b00666
J. Chem. Theory Comput. 2015, 11, 5555−5572

the sum of all such contributions from all of the 1−4 pairs is zero by symmetry. It is therefore impossible to obtain the full individual dihedral potentials in ethane from the rigid torsion curve. This emphasizes the usefulness of the individual derivatives $\partial/\partial\tau$, which, in distinction from $\partial/\partial\phi$, in effect probe the individual dihedral energy surface at distorted configurations and capture in this way "hidden" components of the interaction, which will contribute to the configurations encountered for example in MD, which break symmetry.

SUMMARY

In the coming years it is likely that we will see the evolution of a new class of force fields. Significant advances have been made in the past years including adoption of coupling terms,[36,134,135] widespread inclusion of polarizability,[22,100,103,136] improved representations of electronic distributions,[67,86,119,137] more accurate combination rules,[48,138] and out-of-plane representations.[23,48] However, these advances need to be consolidated. Thus, with the exception of Krimms' FF,[36] none of the current force fields have implemented representations of all of these interactions.

In addition to these terms, if we want to achieve quantitative accuracy, for example calculating ligand binding energies within a few tenths of a kilocalorie,[139] it is likely we will need to account for charge flux and anharmonicity of nonbond interactions and improve the representation of dihedral energy surfaces, including determination of all Fourier terms as discussed previously. Here we reviewed the exploitation of quantum mechanical forces and Hessians to derive force fields including deriving and assessing the importance of coupling terms, anharmonicity, and the validity of different representations of the out-of-plane deformation. The use of these derivatives amplifies the "observable" to parameter ratio by over 2 orders of magnitude and furthermore lends itself to automation. It is expected that the same quantum derivative fitting techniques, especially with advances in computational resources, will enable the determination of the additional interaction terms discussed earlier. Finally it was noted how individual energy derivatives with respect to appropriate interaction coordinates (e.g., interatomic distances, angles, and dihedral angles) can inform on the corresponding interaction. For example the derivative of the energy with respect to the O···O distance in water dimers was shown to reveal the importance of charge flux and also probe the anisotropy of the nonbond O···O repulsion. Likewise the second derivative with respect to an individual dihedral extracts that single rotational potential and allows determination of all Fourier components of the rotation. Thus, these and other derivatives can be invoked to assess the functional form best accounting for these energy surfaces, test transferability, and derive the corresponding force constants.

The preceding discussion is by no means meant to provide an exhaustive review of the work carried out on exploitation of quantum mechanical gradient and Hessians in force field derivation. Hopefully it gives a flavor of the wide applications of the method over the years. It is clear from this summary that though widely used in the materials and polymer community, the vast majority of biomolecular force field developers have yet to avail themselves of this powerful tool. This may well be changing, since, as we saw in the preceding text, recently the creators of one of the most advanced biomolecular force fields, AMOEBA,[100] have utilized QM gradients in their derivation of a new water potential.[46]

AUTHOR INFORMATION

Corresponding Author

*E-mail: athagler@gmail.com.

Notes

The authors declare no competing financial interest.

ABBREVIATIONS:

Å, angstroms; DFT, density functional theory; FF, force field; H-bond, hydrogen bond; HF, Hartree−Fock; MD, molecular dynamics; MEP, molecular electrostatic potential; MM, molecular mechanics; OOP, out-of-plane angle; QDF, quantum derivative fitting; QM, quantum mechanics; TCM, traditional Chinese medicine

REFERENCES

(1) Pullman, B.; Pullman, A. The electronic structure of the purine-pyrimidine pairs of DNA. *Biochim. Biophys. Acta* **1959**, *36*, 343−350.

(2) Hoffmann, R.; Imamura, A. Quantum Mechanical Approach to the Conformational Analysis of Macromolecules in Ground and Excited States. *Biopolymers* **1969**, *7*, 207−213.

(3) Hehre, W. J.; Pople, J. A. Atomic electron populations for some simple molecules. *Chem. Phys. Lett.* **1968**, *2* (6), 379−380.

(4) Momany, F. A.; McGuire, R. F.; Yan, J. F.; Scheraga, H. A. Energy parameters in polypeptides. 3. Semiempirical molecular orbital calculations for hydrogen-bonded model peptides. *J. Phys. Chem.* **1970**, *74* (12), 2424−2438.

(5) Momany, F. A.; McGuire, R. F.; Yan, J. F.; Scheraga, H. A. Energy parameters in polypeptides. IV. Semiempirical molecular orbital calculations of conformational dependence of energy and partial charge in di- and tripeptides. *J. Phys. Chem.* **1971**, *75* (15), 2286−2297.

(6) Cox, S. R.; Williams, D. E. Representation of the molecular electrostatic potential by a net atomic charge model. *J. Comput. Chem.* **1981**, *2* (3), 304−323.

(7) Williams, D. E. Ab initio molecular packing analysis. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1996**, *52* (2), 326−328.

(8) Scrocco, E.; Tomasi, J. Electronic Molecular Structure, Reactivity and Intermolecular Forces: An Euristic Interpretation by Means of Electrostatic Molecular Potentials. *Adv. Quantum Chem.* **1978**, *11*, 115−193.

(9) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11* (4), 431−439.

(10) Hagler, A. T.; Leiserowitz, L.; Tuval, M. Experimental and Theoretical Studies of Barrier to Rotation About N-C-Alpha and C-Alpha-C' Bonds (Phi and Psi) in Amides and Peptides. *J. Am. Chem. Soc.* **1976**, *98* (15), 4600−4612.

(11) Clementi, E.; Cavallone, F.; Scordamaglia, R. Analytical Potentials from "ab Initio" Computations for the Interaction between Biomolecules. 1. Water with Amino Acids. *J. Am. Chem. Soc.* **1977**, *99*, 5531−5545.

(12) Clementi, E. Ab Initio Computational Chemistry. *J. Phys. Chem.* **1985**, *89*, 4426−4436.

(13) Becke, A. D. Perspective: Fifty years of density-functional theory in chemical physics. *J. Chem. Phys.* **2014**, *140* (18), 18A301.

(14) St-Amant, A.; Salahub, D. R. New Algorithm For The Optimization Of Geometries In Local Density Functional Theory. *Chem. Phys. Lett.* **1990**, *169* (5), 387−392.

(15) Bajorath, J.; Kraut, J.; Li, Z. Q.; Kitson, D. H.; Hagler, A. T. Theoretical-Studies on the Dihydrofolate-Reductase Mechanism - Electronic Polarization of Bound Substrates. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88* (15), 6423−6426.

(16) Andzelm, J. W.; Nguyen, D. T.; Eggenberger, R.; Salahub, D. R.; Hagler, A. T. Applications of the Adiabatic Connection Method to Conformational Equilibria and Reactions Involving Formic-Acid. *Comput. Chem.* **1995**, *19* (3), 145−154.

(17) Lii, J.; Allinger, N. L. Directional Hydrogen bonding in the MM3 Force Field. I. *J. Phys. Org. Chem.* **1994**, *7*, 591−609.

(18) 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.

DOI: 10.1021/acs.jctc.5b00666
J. Chem. Theory Comput. 2015, 11, 5555−5572

A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179−5197.
(19) Jorgensen, W. L.; Maxwell, D. S.; Tirado-rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225−11236.
(20) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(21) Zhou, T.; Huang, D.; Caflisch, A. Quantum Mechanical Methods for Drug Design. *Curr. Top. Med. Chem.* **2010**, *10* (1), 33−45.
(22) Lopes, P. E. M.; Guvench, O.; MacKerell, A. D., Jr. Current Status of Protein Force Fields for Molecular Dynamics Simulations. In *Molecular Modeling of Proteins, Methods in Molecular Biology*; Kukol, A., Ed.; Methods in Molecular Biology; Springer: New York, NY, USA, 2015; Vol. *1215*, pp 47−71, DOI: 10.1007/978-1-4939-1465-4_3.
(23) Maple, J. R.; Dinur, U.; Hagler, A. T. Derivation of force fields for molecular mechanics and dynamics from ab initio energy surfaces. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 5350−5354.
(24) Dinur, U.; Hagler, A. T. New Approaches to Empirical Force Fields. In *Reviews in Computational Chemistry*,; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 1991; Vol. *2*, pp 99−164, DOI: 10.1002/9780470125793.ch4.
(25) *Probe Manual*; Biosym Technologies: San Diego, CA, USA, 1989.
(26) Palca, J. Computer Models: Cooperation on New Molecules. *Nature* **1986**, *322*, 586.
(27) Hagler, A. T.; Maple, J. R.; Thacher, T. S.; Fitzgerald, G. B.; Dinur, U. Potential Energy Functions for Organic and Biomolecular Systems. *Computer Simulation of Biomolecular Systems*; van Gunsteren, W. F., Weiner, P. K., Eds.; ESCOM: Leiden, The Netherlands, 1989; Vol. *1*, pp 149−167.
(28) Maple, J. R.; Hwang, M. J.; Stockfisch, T. P.; Dinur, U.; Waldman, M.; Ewig, C. S.; Hagler, A. T. Derivation of Class II Force Fields. 1. Methodology and Quantum Force Field for the Alkyl Functional Group and Alkane Molecules. *J. Comput. Chem.* **1994**, *15* (2), 162−182.
(29) Ercolessi, F.; Adams, J. B. Interatomic Potentials from First-Principles Calculations: the Force-Matching Method. *Eur. Lett.* **1994**, *26*, 583−588.
(30) Hwang, M. J.; Stockfisch, T. P.; Hagler, A. T. Derivation of Class II Force Fields. 2. Derivation and Characterization of a Class-II Force Field, CFF93, for the Alkyl Functional Group and Alkane Molecules. *J. Am. Chem. Soc.* **1994**, *116* (6), 2515−2525.
(31) Ermer, O.; Lifson, S. Consistent Force Field Calculations. III. Vibrations, Conformations, and Heats of Hydrogenation of Non-conjugated Olefins. *J. Am. Chem. Soc.* **1973**, *95* (13), 4121−4132.
(32) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 1. *J. Am. Chem. Soc.* **1989**, *111* (23), 8551−8566.
(33) Maple, J. R.; Hwang, M. J.; Stockfisch, T. P.; Dinur, U.; Waldman, M.; Ewig, C. S.; Hagler, A. T. Derivation of Class II Force-Fields. 1. Methodology and Quantum Force-Field for the Alkyl Functional Group and Alkane Molecules. *J. Comput. Chem.* **1994**, *15* (2), 162−182.
(34) Dillen, J. L. M. An Empirical Force Field. I. Alkanes. *J. Comput. Chem.* **1995**, *16*, 595−609.
(35) Allinger, N. L.; Chen, K.; Lii, J. An Improved Force Field (MM4) for Saturated Hydrocarbons. *J. Comput. Chem.* **1996**, *17* (5−6), 642−668.
(36) Palmo, K.; Mannfors, B.; Mirkin, N. G.; Krimm, S. Potential energy functions: from consistent force fields to spectroscopically determined polarizable force fields. *Biopolymers* **2003**, *68* (3), 383−394.
(37) Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*; Dover: New York, 1980; p 59.
(38) Maple, J. R.; Hwang, M. J.; Jalkanen, K. J.; Stockfisch, T. P.; Hagler, A. T. Derivation of Class II Force Fields: V. Quantum Force Field for Amides, Peptides, and Related Compounds. *J. Comput. Chem.* **1998**, *19* (4), 430−458.
(39) Hwang, M. J.; Ni, X.; Waldman, M.; Ewig, C. S.; Hagler, A. T. Derivation of class II force fields. VI. Carbohydrate compounds and anomeric effects. *Biopolymers* **1998**, *45* (6), 435−468.
(40) Ewig, C. S.; Berry, R.; Dinur, U.; Hill, J. R.; Hwang, M. J.; Li, H. Y.; Liang, C.; Maple, J. O. N.; Peng, Z. W.; Stockfisch, T. P.; et al. Derivation of class II force fields. VIII. Derivation of a general quantum mechanical force field for organic compounds. *J. Comput. Chem.* **2001**, *22* (15), 1782−1800.
(41) Sun, H.; Mumby, S. J.; Maple, J. R.; Hagler, A. T. An Ab-Initio Cff93 All-Atom Force Field for Polycarbonates. *J. Am. Chem. Soc.* **1994**, *116* (7), 2978−2987.
(42) Lifson, S.; Warshel, A. Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane and n-Alkane Molecules. *J. Chem. Phys.* **1968**, *49* (11), 5116.
(43) Hagler, A. T.; Huler, E.; Lifson, S. Energy Functions for Peptides and Proteins. I. Derivation of a Consistent Force Field Including the Hydrogen Bond from Amide Crystals. *J. Am. Chem. Soc.* **1974**, *96* (17), 5319−5327.
(44) Maple, J. R.; Hwang, M. J.; Stockfisch, T. P.; Hagler, A. T. Derivation of Class II Force Fields 0.3. Characterization of a Quantum Force Field for Alkanes. *Isr. J. Chem.* **1994**, *34* (2), 195−231.
(45) Huang, L.; Roux, B. Automated Force Field Parameterization for Non-Polarizable and Polarizable Atomic Models Based on Ab Initio Target Data. *J. Chem. Theory Comput.* **2013**, *9*, 3543−3556.
(46) Laury, M. L.; Wang, L.-P.; Pande, V. S.; Head-Gordon, T. L.; Ponder, J. W. Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *J. Phys. Chem. B* **2015**, *119*, 9423−9437.
(47) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* **1986**, *7* (2), 230−252.
(48) Halgren, T. A. Merck molecular force field.I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17* (5−6), 490−519.
(49) Chen, K.-H.; Allinger, N. L. Molecular mechanics (MM4) study of saturated four-membered ring hydrocarbons. *J. Mol. Struct.: THEOCHEM* **2002**, *581* (1−3), 215−237.
(50) Hagler, A. T.; Osguthorpe, D. J.; Dauber-Osguthorpe, P.; Hempel, J. C. Dynamics and conformational energetics of a peptide hormone: vasopressin. *Science (Washington, DC, U. S.)* **1985**, *227*, 1309−1315.
(51) Wu, S.; Zhang, Y. ANGLOR: a composite machine-learning algorithm for protein backbone torsion angle prediction. *PLoS One* **2008**, *3*, e3400−e3400.
(52) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: a free tool to discover chemistry for biology. *J. Chem. Inf. Model.* **2012**, *52* (7), 1757−1768.
(53) Gu, Q.; Xu, J.; Gu, L. Selecting diversified compounds to build a tangible library for biological and biochemical assays. *Molecules* **2010**, *15* (7), 5031−5044.
(54) Zhanel, G. G.; Ennis, K.; Vercaigne, L.; Walkty, A.; Gin, A. S.; Embil, J.; Smith, H.; Hoban, D. J. A Critical Review of the Fluoroquinolones. Focus on Respiratory Tract Infections. *Drugs* **2002**, *62* (1), 13−59.
(55) Perucca, P.; Savio, M.; Cazzalini, O.; Mocchi, R.; Maccario, C.; Sommatis, S.; Ferraro, D.; Pizzala, R.; Pretali, L.; Fasani, E.; et al. Structure-activity relationship and role of oxygen in the potential antitumour activity of fluoroquinolones in human epithelial cancer cells. *J. Photochem. Photobiol., B* **2014**, *140*, 57−68.
(56) Bradner, W. T. Mitomycin C: a clinical update. *Cancer Treat. Rev.* **2001**, *27* (1), 35−50.
(57) Zargar, H.; Aning, J.; Ischia, J.; So, A.; Black, P. Optimizing intravesical mitomycin C therapy in non-muscle-invasive bladder cancer. *Nat. Rev. Urol.* **2014**, *11* (4), 220−230.
(58) Asif, M. Biological Overview on Thiirane Derivatives. *SOP Trans. Appl. Chem.* **2014**, *1* (1), 1−10.
(59) Testero, S. A.; Bouley, R.; Fisher, J. F.; Chang, M.; Mobashery, S. Exploration of mild copper-mediated coupling of organotrifluoroborates in the synthesis of thiirane-based inhibitors of matrix metalloproteinases. *Bioorg. Med. Chem. Lett.* **2011**, *21* (9), 2675−2678.

DOI: 10.1021/acs.jctc.5b00666
J. Chem. Theory Comput. 2015, 11, 5555−5572

(60) Singh, A.; Lucki, I. Antidepressant-Like Activity Of Compounds With Varying Efficacy At 5-Ht1a Receptors. *Neuropharmacology* **1993**, *32* (4), 331−340.
(61) Fulton, B. S.; Knapp, B. I.; Bidlack, J. M.; Neumeyer, J. L. Synthesis and pharmacological evaluation of hydrophobic esters and ethers of butorphanol at opioid receptors. *Bioorg. Med. Chem. Lett.* **2008**, *18* (16), 4474−4476.
(62) Norrby, P.-O.; Wärnmärk, K.; Åkermark, B.; Moberg, C. Unusual Conformational-Determining Interactions in Oxymethylpyridines: An *ab Initio* Study and an Improved Method for Refining Molecular Parameters. *J. Comput. Chem.* **1995**, *16* (5), 620−627.
(63) Dasgupta, S.; Goddard, W. A. Hessian-biased force fields from combining theory and experiment. *J. Chem. Phys.* **1989**, *90*, 7207−7215.
(64) Norrby, P.; Liljefors, T. Parameterization with Simultaneous Utilization of Experimental and Quantum Mechanical Data. *J. Comput. Chem.* **1998**, *19* (10), 1146−1166.
(65) Hill, J. R.; Sauer, J. Molecular mechanics potential for silica and zeolite catalysts based on ab initio calculations. 1. Dense and microporous silica. *J. Phys. Chem.* **1994**, *98* (4), 1238−1244.
(66) Palmo, K. I. M.; Krimm, S. Electrostatic Model for Infrared Intensities in a Spectroscopically Determined Molecular Mechanics Force Field. *J. Comput. Chem.* **1998**, *19* (7), 754−768.
(67) Mannfors, B.; Palmo, K.; Krimm, S. A new electrostatic model for molecular mechanics force fields. *J. Mol. Struct.* **2000**, *556*, 1−21.
(68) Mannfors, B.; Palmo, K.; Krimm, S. Spectroscopically determined force field for water dimer: physically enhanced treatment of hydrogen bonding in molecular mechanics energy functions. *J. Phys. Chem. A* **2008**, *112*, 12667−12678.
(69) Tafipolsky, M.; Schmid, R. A Consistent Force Field for the Carboxylate Group. *J. Chem. Theory Comput.* **2009**, *5* (10), 2822−2834.
(70) Waldher, B.; Kuta, J.; Chen, S.; Henson, N.; Clark, A. E. ForceFit: A Code to Fit Classical Force Fields to Quantum Mechanical Potential Energy Surfaces. *J. Comput. Chem.* **2010**, *31* (12), 2307−2316.
(71) Song, C.; Rhee, Y. M. Development of Force Field Parameters for Oxyluciferin on its Electronic Ground and Excited Sates. *Int. J. Quantum Chem.* **2011**, *111*, 4091−4105.
(72) Perez-Angel, E. C.; Seminario, J. M. Ab Initio Analysis and Harmonic Force Fields of Gallium Nitride Nanoclusters. *J. Phys. Chem. C* **2011**, *115* (14), 6467−6477.
(73) Biczysko, M.; Bloino, J.; Brancato, G.; Cacelli, I.; Cappelli, C.; Ferretti, A.; Lami, A.; Monti, S.; Pedone, A.; Prampolini, G.; et al. Integrated computational approaches for spectroscopic studies of molecular systems in the gas phase and in solution: pyrimidine as a test case. *Theor. Chem. Acc.* **2012**, *131*, 1201−1220.
(74) Jorn, R.; Kumar, R.; Abraham, D. P.; Voth, G. Atomistic Modeling of the Electrode−Electrolyte Interface in Li-Ion Energy Storage Systems: Electrolyte Structuring. *J. Phys. Chem. C* **2013**, *117*, 3747−3761.
(75) Barone, V.; Cacelli, I.; De Mitri, N.; Licari, D.; Monti, S.; Prampolini, G. JOYCE and ULYSSES: integrated and user-friendly tools for the parameterization of intramolecular force fields from quantum mechanical data. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3736−3751.
(76) Grimme, S. A General Quantum Mechanically Derived Force Field (QMDFF) for Molecules and Condensed Phase Simulations. *J. Chem. Theory Comput.* **2014**, *10* (10), 4497−4514.
(77) Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. Systematic improvement of a classical molecular model of water. *J. Phys. Chem. B* **2013**, *117* (34), 9956−9972.
(78) Limé, E.; Norrby, P.-O. Improving the Q2MM method for transition state force field modeling. *J. Comput. Chem.* **2015**, *36* (4), 244−250.
(79) Cisneros, G. A.; Karttunen, M.; Ren, P.; Sagui, C. Classical electrostatics for biomolecular simulations. *Chem. Rev.* **2014**, *114* (1), 779−814.
(80) Demerdash, O.; Yap, E.-H.; Head-Gordon, T. Advanced potential energy surfaces for condensed phase simulation. *Annu. Rev. Phys. Chem.* **2014**, *65*, 149−174.
(81) Price, S. L. Predicting crystal structures of organic compounds. *Chem. Soc. Rev.* **2014**, *43* (7), 2098−2111.
(82) Badrinarayan, P.; Choudhury, C.; Sastry, P. G. N. Molecular Modeling. In *Systems and Synthetic Biology*; Singh, V., Dhar, P. K., Eds.; Springer: Dordrecht, The Netherlands, 2015; pp 93−128.
(83) Dinur, U.; Hagler, A. T. Direct Evaluation of Nonbonding Interactions from ab initio Calculations. *J. Am. Chem. Soc.* **1989**, *111*, 5149−5151.
(84) Dinur, U.; Hagler, A. T. The role of nonbond and charge flux in hydrogen bond interactions. The effect on structural changes and spectral shifts in water dimer. *J. Chem. Phys.* **1992**, *97* (12), 9161−9172.
(85) Dinur, U.; Hagler, A. T. Geometry-Dependent Atomic Charges - Methodology and Application to Alkanes, Aldehydes, Ketones, and Amides. *J. Comput. Chem.* **1995**, *16* (2), 154−170.
(86) Berkovitch-yellin, Z.; Leiserowitz, L. The Role of Coulomb Forces in the Crystal Packing of Amides. A Study Based on Experimental Electron Densities. *J. Am. Chem. Soc.* **1980**, *102* (26), 7677−7690.
(87) Berkovitch-Yellin, Z.; Leiserowitz, L. Atom-Atom Potential Analysis of the Packing Characteristics of Carboxylic Acids. A Study Based on Experimental Electron Density Distributions. *J. Am. Chem. Soc.* **1982**, *104*, 4052−4064.
(88) Stone, A. J. Distributed Multipole Analysis, or How to Describe a Molecular Charge Distribution. *Chem. Phys. Lett.* **1981**, *83* (2), 233−239.
(89) Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8478−8490.
(90) Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardère, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P. A General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 2609−2618.
(91) Dinur, U. Charge Flux and Electrostatlc Forces In Planar Molecules. *J. Phys. Chem.* **1991**, *95*, 6201−6211.
(92) Decius, J. C. An Effective Atomic Charge Model for Infrared Intensities. *J. Mol. Spectrosc.* **1975**, *57*, 348−362.
(93) Overend, J.; Hylden, J. L. A charge flow model for the higher derivatives of the molecular dipole moment. I. The model. *J. Chem. Phys.* **1985**, *83*, 2661−2667.
(94) Dinur, U.; Hagler, A. T. Determination of Atomic Point Charges and Point Dipoles from the Cartesian Derivatives of the Molecular Dipole-Moment and 2nd Moments, and from Energy 2nd Derivatives of Planar Dimers 0.2. Applications to Model Systems. *J. Chem. Phys.* **1989**, *91* (5), 2949−2959.
(95) Baranović, G. Intramolecular resonance-assisted hydrogen bonds: a theoretical description by means of atomic charges and charge fluxes. *Spectrochim. Acta, Part A* **2014**, *117*, 465−472.
(96) Torii, H. Electronic structural basis for the atomic partial charges of planar molecular systems derived from out-of-plane dipole derivatives. *J. Phys. Chem. A* **2015**, *119* (13), 3277−3284.
(97) Galimberti, D.; Milani, A.; Castiglioni, C. Charge mobility in molecules: charge fluxes from second derivatives of the molecular dipole. *J. Chem. Phys.* **2013**, *138*, 164115−164123.
(98) Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J. Polarization effects in molecular mechanical force fields. *J. Phys.: Condens. Matter* **2009**, *21* (33), 333102−333123.
(99) Lopes, P. E. M.; Roux, B.; MacKerell, A. D., Jr. Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability. Theory and applications. *Theor. Chem. Acc.* **2009**, *124*, 11−28.
(100) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9* (9), 4046−4063.
(101) Gao, J.; Truhlar, D. G.; Wang, Y.; Mazack, M. J. M.; Löffler, P.; Provorse, M. R.; Rehak, P. Explicit Polarization: A Quantum Mechanical Framework for Developing Next Generation Force Fields. *Acc. Chem. Res.* **2014**, *47* (9), 2837−2845.
(102) Popelier, P. L. A. QCTFF: On the construction of a novel protein force field. *Int. J. Quantum Chem.* **2015**, *115*, 1005−1011.
(103) Khoruzhii, O.; Butin, O.; Illarionov, A.; Leontyev, I.; Olevanov, M.; Ozrin, V.; Pereyaslavets, L.; Fain, B. Polarizable Force Fields for Proteins. In *Protein Modelling*; Náray-Szabó, G., Ed.; Springer Interna-

DOI: 10.1021/acs.jctc.5b00666
J. Chem. Theory Comput. 2015, 11, 5555−5572

tional: Cham, Switzerland, 2014; Chapter 5, pp 91−134, DOI: 10.1007/978-3-319-09976-7_5.
(104) Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta* **1977**, *44* (2), 129−138.
(105) Bonaccorsi, R.; Scrocco, E.; Tomasi, J. An approximate expression of the electrostatic molecular potential in terms of completely transferable group contributions. *J. Am. Chem. Soc.* **1977**, *99* (14), 4546−4554.
(106) Dinur, U.; Hagler, A. T. Determination of atomic point charges and point dipoles from the Cartesian derivatives of the molecular dipole moment and second moments, and from energy second derivatives of planar dimers II. Applications to model systems. *J. Chem. Phys.* **1989**, *91* (5), 2959−2970.
(107) Person, W. B.; Zerbi, G. *Vibrational Intensities in Infrared and Raman Spectroscopy*; Person, W. B., Zerbi, G., Eds.; Elsevier: Amsterdam, 1982.
(108) Stern, P. S.; Chorev, M.; Goodman, M.; Hagler, A. T. Computer-Simulation of the Conformational Properties of Retro-Inverso Peptides 0.2. Abinitio Study, Spatial Electron-Distribution, and Population Analysis of N-Formylglycine Methylamide, N-Formyl N'-Acetyldiaminomethane, and N-Methylmalonamide. *Biopolymers* **1983**, *22* (8), 1885−1900.
(109) Gussoni, M.; Castiglioni, C.; Ramos, M. N.; Rui, M.; Zerbi, G. Infrared intensities: from intensity parameters to an overall understanding of the spectrum. *J. Mol. Struct.* **1990**, *224*, 445−470.
(110) Stouch, T. R.; Williams, D. E. Conformational dependence of electrostatic potential derived charges of a lipid headgroup: Glycerylphosphorylcholine. *J. Comput. Chem.* **1992**, *13* (5), 622−632.
(111) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Multiconformational RESP Methodology to Biopolymers: Charge Derivation for DNA, RNA, and Proteins. *J. Comput. Chem.* **1995**, *16* (1), 1357−1377.
(112) Basma, N.; Sundara, S.; Calgan, D.; Vernali, T.; Woods, R. J. Solvated Ensemble Averaging in the Calculation of Partial Atomic Charges. *J. Comput. Chem.* **2001**, *22* (11), 1125−1137.
(113) Dinur, U. Flexible" water molecules in external electrostatic potentials. *J. Phys. Chem.* **1990**, *94* (15), 5669−5671.
(114) Miwa, Y.; Machida, K. Molecular Mechanics Simulations of Thermodynamic Functions and Infrared Spectra of Alkanes. *J. Am. Chem. Soc.* **1988**, *110*, 5183−5189.
(115) Ewig, C. S.; Thacher, T. S.; Hagler, A. T. Derivation of class II force fields. 7. Nonbonded force field parameters for organic compounds. *J. Phys. Chem. B* **1999**, *103* (33), 6998−7014.
(116) Yamaoka, Y.; Machida, K. Anharmonic vibrational wave functions, infrared band intensities, and dipole moments of water. *J. Mol. Spectrosc.* **1983**, *100* (2), 234−244.
(117) Zilles, B. A.; Person, W. B. Interpretation of infrared intensity changes on molecular complex formation. I. Water dimer. *J. Chem. Phys.* **1983**, *79*, 65−77.
(118) Stone, A. J.; Price, S. L. Some New Ideas in the Theory of Intermolecular Forces: Anisotropic Atom-Atom. *J. Phys. Chem.* **1988**, *92*, 3325−3335.
(119) Price, S. L.; Price, L. S. Modelling Intermolecular Forces for Organic Crystal Structure Prediction. *Struc. Bond* **2005**, *115*, 81−123.
(120) Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68* (3), 441−451.
(121) Nyburg, S. C.; Faerman, C. H. A Revision of van der Waals Atomic Radii for Molecular Crystals: N, O, F, S, CI, Se, Br and I Bonded to Carbon. *Acta Crystallogr., Sect. B: Struct. Sci.* **1985**, *41* (1), 274−279.
(122) Badenhoop, J. K.; Weinhold, F. Natural steric analysis: Ab initio van der Waals radii of atoms and ions. *J. Chem. Phys.* **1997**, *107*, 5422−5432.
(123) Batsanov, S. S. Van der Waals Radii of Elements. *Inorg. Mater.* **2001**, *37* (9), 871−885.
(124) Zgarbová, M.; Otyepka, M.; Sponer, J.; Hobza, P.; Jurecka, P. Large-scale compensation of errors in pairwise-additive empirical force fields: comparison of AMBER intermolecular terms with rigorous DFT-SAPT calculations. *Phys. Chem. Chem. Phys.* **2010**, *12* (35), 10476−10493.
(125) Šponer, J.; Šponer, J. E.; Mládek, A.; Jurečka, P.; Banáš, P.; Otyepka, M. Nature and magnitude of aromatic base stacking in DNA and RNA: Quantum chemistry, molecular mechanics, and experiment. *Biopolymers* **2013**, *99* (12), 978−988.
(126) Eramian, H.; Tian, Y.-H.; Fox, Z.; Beneberu, H. Z.; Kertesz, M. On the anisotropy of van der Waals atomic radii of O, S, Se, F, Cl, Br, and I. *J. Phys. Chem. A* **2013**, *117* (51), 14184−14190.
(127) Kim, H.; Doan, V. D.; Cho, W. J.; Madhav, M. V.; Kim, K. S. Anisotropic charge distribution and anisotropic van der Waals radius leading to intriguing anisotropic noncovalent interactions. *Sci. Rep.* **2014**, *4*, 5826−5833.
(128) Price, S. L. Towards More Accurate Model Intermolecular Potentials for Organic Molecules. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; John Wiley and Sons: New York, 2000; Vol. *14*, pp 225−289.
(129) Price, S. L. The computational prediction of pharmaceutical crystal structures and polymorphism. *Adv. Drug Delivery Rev.* **2004**, *56* (3), 301−319.
(130) Mitchell, J. B. O.; Price, S. L.; Leslie, M.; Buttar, D.; Roberts, R. J. Anisotropic Repulsion Potentials for Cyanuric Chloride (C3N3Cl3) and Their Application to Modeling the Crystal Structures of Azaaromatic Chlorides. *J. Phys. Chem. A* **2001**, *105*, 9961−9971.
(131) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Genet.* **2006**, *65* (3), 712−725.
(132) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **2004**, *126* (3), 698−699.
(133) Dinur, U.; Hagler, A. T. A Novel Decomposition of Torsional Potentials into Pairwise Interactions - a Study of Energy 2nd Derivatives. *J. Comput. Chem.* **1990**, *11* (10), 1234−1246.
(134) Ermer, O. Calculation of Molecular Properties Using Force Fields. Applications in Organic Chemistry. *Struct. Bonding* **1976**, *27*, 161−211.
(135) Nevins, N.; Chen, K.; Allinger, N. L. Molecular mechanics (MM4) calculations on alkenes. *J. Comput. Chem.* **1996**, *17* (5−6), 669−694.
(136) Wang, J.; Cieplak, P.; Li, J.; Wang, J.; Cai, Q.; Hsieh, M.; Lei, H.; Luo, R.; Duan, Y. Development of polarizable models for molecular mechanical calculations II: induced dipole models significantly improve accuracy of intermolecular interaction energies. *J. Phys. Chem. B* **2011**, *115* (12), 3100−3111.
(137) Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011**, *7*, 3143−3161.
(138) Waldman, M.; Hagler, A. T. New Combining Rules for Rare-Gas van der Waals Parameters. *J. Comput. Chem.* **1993**, *14* (9), 1077−1084.
(139) Stouch, T. R. The errors of our ways: taking account of error in computer-aided drug design to build confidence intervals for our next 25 years. *J. Comput.-Aided Mol. Des.* **2012**, *26* (1), 125−134.

DOI: 10.1021/acs.jctc.5b00666
J. Chem. Theory Comput. 2015, 11, 5555−5572