Developing accurate molecular mechanics force fields for conjugated molecular systems

Hainam Do* and Alessandro Troisi*

Department of Chemistry and Centre for Scientific Computing, University of Warwick, Coventry CV4 7AL, United Kingdom

E-Mails: *H.Do@warwick.ac.uk*, *A.Troisi@warwick.ac.uk*

Abstract

A rapid method to parameterize the intramolecular component of classical force fields for complex conjugated molecules is proposed. The method is based on a procedure of force matching with a reference electronic structure calculation. It is particularly suitable for those applications where molecular dynamics simulations are used to generate structures that are therefore analysed with electronic structure methods, because it is possible to build force fields that are consistent with the electronic structure calculations that follow the classical simulations. Such applications are commonly encountered in organic electronics, spectroscopy of complex systems and photobiology (e.g. photosynthetic systems). We illustrate the method by parameterizing the force fields of a molecule used in molecular semiconductors (2,2-dicyanovinyl-capped *S, N*-heteropentacene or DCV-SN5), a polymeric semiconductor (thieno[3,2-b]thiophene-diketopyrrolopyrrole TT-DPP) and a chromophore embedded in a protein environment (15,16-Dihydrobiliverdin or DBV) where several hundreds of parameters need to be optimized in parallel.

1. Introduction

Classical Molecular Dynamics (MD) simulations have recently found application in the study of optical and electronic properties of materials and biomolecules. For this class of applications, MD simulations are used to generate the structures explored by the system under the experimental condition and to interpret experimental results in conjunction with other theoretical methods, e.g. quantum chemical methods or other phenomenological theories. An important area where MD simulations have become particularly important is organic electronics, i.e. the study of electronic and optical properties of polymers and small molecules.[1, 2] The structural[3-7] and dynamic[8, 9] disorders of all organic materials determine their electronic properties and MD simulation are used to correlate the chemical detail of the system with the microstructure. Large-scale electronic structure calculations are then used to correlate the local structure with the observable electronic structure properties.[6, 10] Another new class of applications of classical MD is the elucidation of experiments in ultrafast electronic spectroscopy of biomolecules. It was recently shown that long quantum coherences can be observed in biological molecules[11-13] and the aim of atomistic simulations is to explain how the (classical) environment can influence the (quantum) evolution of the electronic states.[14-18] Also for the applications in photobiology, MD simulations are often coupled with electronic structure methods to describe aspects of the quantum dynamics.

For all applications where MD is a preliminary step toward the study of electronic structure properties it is very difficult to find sufficiently accurate force fields (FFs). All single

molecules, biological chromophores, and semiconducting polymers of interest display a complex chemical structure with extended pi-conjugation that prevents the use of standard FFs. For example, the bonds between sp2 carbon atoms usually have different bond orders (and therefore distances and force constants). The MM3 force field, specifically designed to describe conjugated carbon framework,[19] cannot properly describe systems containing many heteroatoms within the conjugated frame.[20] Errors of the order of 0.05 Å on the bond distances can be tolerable for thermodynamic properties but lead to unacceptable electronic structure properties. FFs suitable for large conjugated molecules need to use many atom types and parameters and, unlike the most common FFs, they also tend to be extremely system specific and not transferable. Moreover, when one uses classical simulation methods as input for electronic structure calculations it is desirable that the equilibrium structure of the classical simulation and the electronic structure calculations are as close as possible, i.e. the force field should be *consistent* with the electronic structure calculation that follows it.

In this paper we propose a method to find rapidly the FF parameters for intramolecular interaction in medium-to-large conjugated molecules so that the empirical forces are as close as possible to those computed with any predetermined electronic structure calculation method. In the following section we provide a brief overview of the currently available methodologies for FF parameterisation, with stress on the ones that are closer to our method and based on the idea of "force matching". Our method is described in section 3. Section 4 illustrates the methodology with several examples focusing on the appropriate set up of the parameter optimization.

2. Background

The parameters for the most commonly used molecular mechanics FFs such as CHARMM[21] and AMBER[22, 23] have been derived to reproduce the experimental vibrational and crystallography data for a class of compounds, e.g. biological molecules.[24] Each force field often includes a standardized procedure to determine atomic point charges from electrostatic potential fitting of ab initio calculations. It is possible to introduce parameters for a new molecule "by analogy", i.e. finding the closest match between atom types of an existing force field on the basis of chemical intuition. Much effort has been devoted into developing automatic protocols for parameterizing an arbitrary molecule in this way. For example in the general AMBER FFs (GAFF),[25] parameters for small organic molecules compatible with the standard AMBER FFs can be generated automatically after assigning atom type (on the basis of chemical environment) and using a combination of look-up table, empirical rules and ab initio calculations for the definition of the parameters. An analogous scheme (named CGenFF) is available for the CHARMM FF.[26, 27]

Although these tools have significantly broaden the range of biological systems that can be investigated, they fail when the local structure of the molecule (bond lengths, rotational barriers) is not determined by the local environment but by the global electronic structure, as in conjugated molecules and polymers. In these cases it is well known that partial charges and dihedral parameters have very limited transferability. It is therefore desirable to derive parameters for new molecules from quantum mechanical data, and a number of independent research group have contributed to this problem.[28] JOYCE – a program to derive all-atom and

united-atom FFs for small molecules has been developed by the Barone group.[29, 30] Inputs to this program are QM equilibrium geometry, energy, gradient and the hessian matrix. FFs parameters are obtained by minimizing a cost function, which measures the errors between QM energy, gradient and hessian and those calculated from the FFs. The drawback of this approach is its complicated weighting scheme that has to be determined by the users for every molecule. Recently, GAAMP, General Automated Atomic Model Parameterization has been proposed by Huang and Roux[28] in an attempt to derive CHARMM or AMBER compatible FFs based on the foundation of CGenFF[26] and GAFF.[25] They particularly pay attention to obtain CHARMM or AMBER compatible charges and reliable FFs for soft dihedrals. Another effort is the development of FFs ToolKit by Mayne et al. to automatically parameterise CHARMM compatible FFs for small molecules.[31] Also Grimme developed a procedure to automatically parameterize FFs from QM input data – a Quantum Mechanically Derived FFs (QMDFF).[32] The FF is constructed by using the equilibrium structure, the Hessian matrix, the atomic partial charges, and the covalent bond orders. A similar approach was also carried out by the Ayers group aiming to produce an automated procedure for parameterising AMBER-compatible FFs for transition metal complexes.[33] The force constant of the harmonic terms are derived from the knowledge of the ab initio Hessian matrix and the torsion potentials are obtained from relaxed potential energy surface scan. Efforts in the same spirit have been presented for metalloproteins[34] and Metal Organic Frameworks.[35, 36] An attempt to make the procedure as automatic as possible is made by the QuickFF[37] program, which reads in the equilibrium structure, atomic point charges and the Hessian matrix and give out the FF parameters.

Employing electronic structure calculations either in the gas phase or in a continuum solvent to determine parameters (or missing parameters) for FFs bear the risk that the newly parameterized FFs may not necessarily transferable to the actual system under investigation in the condensed phases. A better approach that mitigates this problem is to map the "classical" potential energy surface and/or its first derivatives of the investigated system onto the ab inito ones by optimizing a set of pre-defined parameters. This approach implicitly incorporates many-body effects and is often known as the force-matching (FM) technique - first proposed by Ercolessi and Adams.[38] In the force-matching method, all unknown parameters can be optimised simultaneously by minimizing an objective function that measures the difference between the FF and the ab initio forces. The technique is very appealing and has provided a way to parameterize FFs for systems that are difficult or even impossible by any other means. For example, interaction potentials have been parameterized for metals,[38, 39] transition metal complexes,[40, 41], anions[42], reactive FFs,[43, 44], ionic liquids,[45] flexible water models,[46] and microporous materials.[47] Traditionally, an ensemble of equilibrium structures to use in the FM procedures is generated from ab initio MD simulations, which also come with the reference ab initio forces.[38, 48, 49] Despite its success, the problem with this scheme is that sampling configuration space with an ab initio method is only limited to small systems due to the fact that proper sampling of the equilibrium ensemble is extremely expensive with electronic structure calculations. In an attempt to parameterize bimolecular FFs using the FM method, Maurer *et al.* employed a quantum mechanics/molecular mechanics (QM/MM) approach to derive FFs that reproduce the steric, electrostatic, and dynamic properties of the QM subsystem.[50] Works along this line were also carried out to produce FFs for azole-bridged dinuclear platinum anticancer drugs[51], zinc in metalloproteins[41], and the 11-cis protonated schiff base chromophore of Rhodopsin.[52] Recently, the method has also been extended to

parameterized FFs along a reaction path.[53] Although QM/MM FM is a step forward comparing to the original scheme, it still suffers from the time limit due to the computational cost in the QM region.

To circumvent this, Wang and coworkers proposed the adaptive FM (AFM) method.[54, 55] Instead of using ab initio MD, the AFM approach uses MM FFs to generate ensembles of equilibrium structures followed by QM/MM calculations on each configuration with the MM region represented by point charges. The procedure starts with a guessed FF and is repeated until convergence of the FF parameters is reached. Similar approaches have also been suggested to parameterize a highly accurate polarizable and two rigid no-npolarizable water models[56, 57] or to refine intramolecular AMBER FF parameters.[58] Recently, AFM has also been employed to derive a simple FF for graphene (PPBE-G), which provide good agreement in comparison with DFT and experiential values for several experimental properties, including the Young's modulus, bending rigidity, and thermal conductivity.[59]

In this work, we propose and test a methodology to develop accurate FFs for conjugated molecules, focusing on the *intramolecular* components where the available FFs are known to perform very poorly. To the best of our knowledge such automatic methods have not been employed for the parameterization of the potential energy in complex conjugated polymers or large chromophores. These systems present a somewhat different challenge with respect to the other mentioned in this section because they are characterized by a large number of atom types and therefore a very large number of parameters to be optimized in parallel. The method, which resemble in spirit the AFM approach, yield parameters consistent with any predefined electronic structure calculation methods in a quasi-automatic fashion and is designed for the specific applications in material science and photobiology outlined in the first section.

3. Methods

In our work, the best FF parameters are defined as those that minimize the differences between the forces computed with the FFs and the forces computed with a reference electronic structure calculation method. These differences are calculated for a set of molecular geometries close to the equilibrium structure of the molecule of interest. In other words, we consider M geometries (k = 1,…, M) of a molecule consisting of N atoms (i = 1, …, N). For each geometry k it is possible to compute the ab initio forces on atom i, $\mathbf{f}_{i,k}^{AB}$. Similarly, it is straightforward to compute the equivalent forces $\mathbf{f}_{i,k}^{FF}$ from an empirical FF defined by the parameter set $\{p_j\}$. Here, we define the objective function $O(\{p_j\})$ which essential the root mean square deviations (rmsd) of the forces as

$$O(\{p_j\}) = \sqrt{\frac{1}{3MN}\sum_{k=1}^{M}\sum_{i=1}^{N}\left\|\mathbf{f}_{i,k}^{AB} - \mathbf{f}_{i,k}^{FF}\right\|} \quad (1)$$

Where $\|...\|$ denote the norm.

The optimization process starts from an initial guessed FF with appropriate atom-type assigned. The force constants and torsional energy barriers for this initial guess can be taken from the literature e.g. from CHARMM, AMBER or OPLS[60] etc. FFs family depending on the choice of the FF style employed, while equilibrium bonds and angles are taken from the QM

equilibrium structures. An initial set of *M* geometries close to the equilibrium structure are then generated from an MD simulation (in this case using NAMD[61]) at the temperature of interest using the guessed potential followed by *M* computations of the ab initio forces of these geometries. The FF parameters set $\{p_j\}$ (hence the FF forces effectively) are then modified until the objective function $O(\{p_j\})$ is minimized. After the FF is parameterized, the newly constructed FF is used as a new guess FF and the procedure is repeated. In general, 2-3 iterations are sufficient as shown in section 4. In practice, the method produces a FF that best matches the ab initio forces for structure visited at a given temperature. It is therefore suitable for all the "rigid" degrees of freedom, i.e. all intermolecular degrees of freedom except the torsions whose barrier is easily overcome at room temperature (e.g. around C-C single bond).

Here, the objective function is optimized using a Monte Carlo (MC) minimization technique. The algorithm proceeds by iteratively selecting randomly a parameter, modifying it, computing the objective function and accepting or rejecting the new parameter if the objective function decreases. Each parameter p_j is modified by adding a random number uniformly distributed between $-p_j^{\max}$ and $p_j^{\max}$. The variable $p_j^{\max}$ is initialized to $0.5 \times p_j$ at the beginning of the simulation and is adjusted during the simulation in order to keep the acceptance rate of the MC move close to 25%. The simulation is divided in "blocks" of attempted moves, each consisting of 100 times the number of parameters moves. The acceptance rate for parameter p_j, $acc(j)$, is evaluated at the beginning of each MC block and the parameter $p_j^{\max}$ is adjusted according to $p_j^{\max} = [(acc(j) - 0.25) + 1] \times p_j^{\max}$, i.e. it is increased (decreased) if the acceptance rate is higher (lower) than 25%. The optimization is deemed to be converged and is terminated when the change in the objective function is lower than a threshold, here set to 10^{-10} kcal/mol/Å. For all studied systems, the optimized objective function is always very similar if the procedure is repeated with a different random seed, indicating that there is a single minimum or a number of equivalent minima. This observation suggests that one can use faster methods to optimize the objective, e.g. based on gradient descent. However, the MC procedure is more easily modifiable to deal with complicated situations where multiple non-equivalent minima are presents, for example introducing a simulated-annealing procedure, and we cannot exclude that such situations may present themselves for different systems. It should be noted that the electronic structure calculations (before minimizing the objective function) represent approximately 60-70 % of the computer time and any time gain in the FF optimization would not radically change the overall time of the procedure.

For any specific problem, one needs to decide on the parameters to be optimized and the number of reference structures needed (this will be discussed in the following section). In this work, since we are only interested in the intramolecular interactions, the intermolecular part (electrostatic and Van der Waals interactions) is not optimised. Instead, the non-bonded parameters are taken from the existing force fields directly and the charges are computed following the same approach as that of the FF family that is employed. For example, for the DCV-SN5 and DBV molecules for which CHARMM FF style are used, the van der Waals parameters are taken from the CHARMM FF and the charges are computed using the RESP[62] method at the HF/6-31G* level of theory. For the other systems considered, for which a OPLS[60] FF is used, we used the corresponding van der Waals parameters and charges computed using

the CHELP[63] method at the B3LYP/6-31G* level of theory. The intermolecular parameters of the force field are kept constant during the optimization of the intramolecular parameters. Therefore, if one is interested in exploring alternative schemes for intermolecular interactions[64, 65] the intramolecular degrees of freedom need to be re-parameterized. The force calculations are performed here at the B3LYP/6-31G* level of theory, to illustrate the methodology with a readily available and frequently used density functional and basis set. All QM calculations are carried out using the Gaussian03[66] package.

4. Results

To exemplify the possible applications of the method we derive accurate FFs for (i) a large rigid conjugated molecule, (ii) a semiconducting polymer flexible degrees of freedom and (iii) a large chromophore embedded in light-harvesting biological complexes. Before going into each of these systems, we shall first examine a simpler example to discuss more in detail the properties of the parameterization and how it should be set up. The full set of optimized FF parameters is given in the supporting information.

I. A test case using the diketopyrrolopyrrole (DPP) molecule

Our initial test molecule DPP is a planar molecule commonly found in diketopyrrolopyrrole based semiconducting polymers.[67] There are 14 atoms in this molecule and the number of FF parameters required to be optimized are 78. In this example, we shall explain our FM procedure by parameterizing a new FF for this molecule using the OPLS FF style. The layout of the atom-type is shown in Figure 1. Having optimised the structure, computed the point charges and set up an initial guessed FF, we then perform MD simulations at 300 K in vacuum. For this and the other molecules we extract a set of structures from snapshots separated by 1 ps along the MD. This time interval is close to the period of the slowest normal modes of the molecule and the structures can be considered a virtually uncorrelated set around the equilibrium geometry. A critical parameter of the method is the number of structures used to optimize the FF parameters. With too few structures the FF would be inaccurate because of overfitting (few data with respect to the number parameters to optimize) and the minimized objective function would increase as the number of structure considered increase (Figure 2). With a sufficient number of structures the minimized objective function becomes insensitive to the number of structures used. Figure 2 shows that the minimized objective function (*minimized force rmsd*) for this molecule begins to converge at about 80 structures and thus we have used 100 structures in the FM procedure in this case. In a molecule with N_a atom, for which N_S structures (force calculations) are considered, the ratio between the number of forces that are matched and the number of parameters N_P is $a = 3 \times N_a \times N_S / N_P$. The test reported in Figure 2 suggests that a value of a close to 50 should be ideal. As the number of parameters increases roughly proportionally to the number of atoms, the number of force calculations remains approximately constant, no matter how large the molecule is. For the other molecules considered in this work, the number of structures used for the FF optimization was set to be $N_S >\sim 50 \times N_P / 3 \times N_a$.

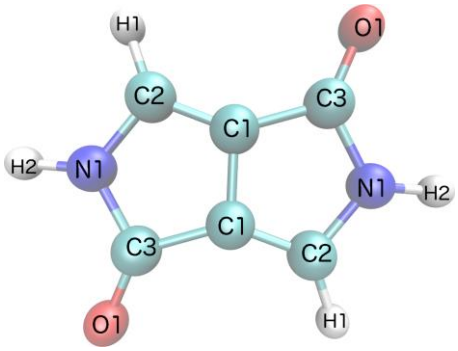


Figure 1. The geometry of the DPP molecule with the atom-type labelled.

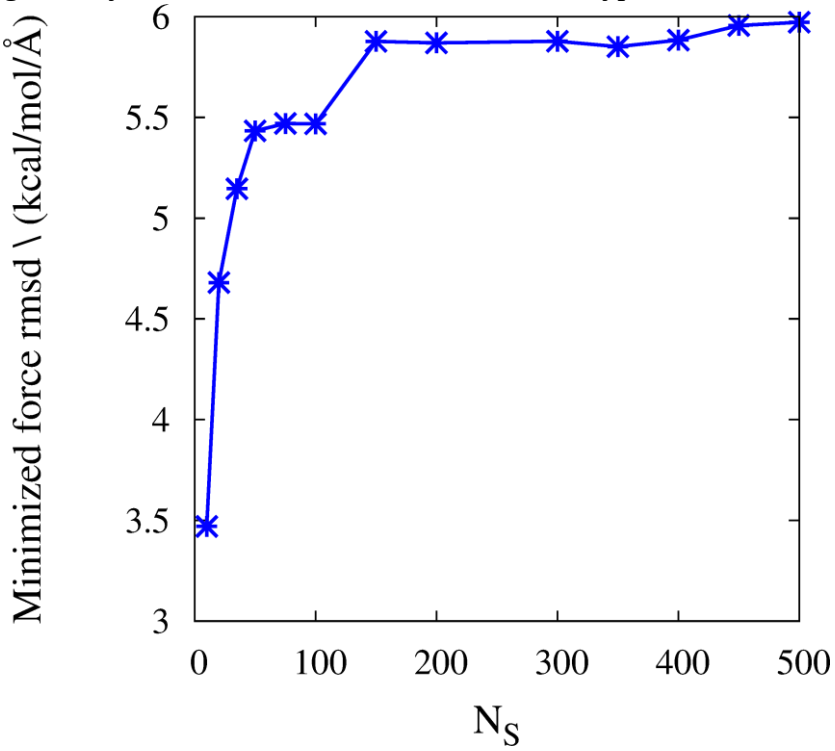


Figure 2. Optimized rmsd between the ab initio and FF forces (i.e. the value of the minimized objective function) versus the number of the structures used.

Due to the planarity of the extended pi-conjugation systems, the equilibrium dihedral angles that contain two sp2 atoms in the middle are either 0° or 180°. This means for those dihedral angles it is appropriate to represent the torsional potential using only the second term in the OPLS expression $E_{dihed} = \frac{V_2}{2}\left[1 - \cos\left(2f - d\right)\right]$, where δ is set to 180°. Note that there are no soft dihedral angles in this molecule, i.e. torsional barriers that are crossed and experimentally accessible temperature. The soft dihedral angles appear in the semiconducting polymer systems and are parameterized separately using the torsion scanning approach as discussed in the next section. The advantage of fixing the phases d is that we can reduce significantly the number of parameters to be optimized, thus saving the computational costs. We make a comparison between two FFs, one obtained from a full parameterization i.e. with the phases are included in the FF optimisation procedure and one obtained from the optimization process where the

phases are fixed at 180°. Table 1 shows that the rmsd between the ab initio and FF forces in the two cases (the optimized objective functions) are more or less the same. In addition, we also analyse the rmsd of the bonds, angles and dihedral angles of the FF optimized geometry with respect to the ab initio optimized geometry. Both FFs give excellent rmsd for bonds and angles, while the rmsd of the dihedral angles in the case where the phases are fixed is improved. These results confirm that for planar and "rigid" molecules it is best to optimize only the energy barriers.

Table 1: A comparison between two FFs for DPP molecule. The first one comes from a fit that involves both the energy barriers (V) and the phase (δ) and the second one comes from a fit that only involves the energy barriers (V).

	Minimized force rmsd (kcal/mol/Å)	**Bond rmsd (Å)**	**Angle rmsd (°)**	**Dihe rmsd (°)**
Fit both V and δ	4.936	0.0056	0.316	0.789
Fix δ at 180° and optimize V	4.941	0.0055	0.316	0.040

After the FF is optimized (based on an initial guess), It may be necessary to re-parameterize the newly derived FF by repeating the whole procedure using the current optimized FF as the new guess. Ideally, the process should be repeated until the rmsd of the forces (before parameter optimization) converges to the same value as the minimized force rmsd (after parameter optimization). The whole FF optimization procedure is considered to be converged if the rmsd of the forces computed for any arbitrary set of structures is the same and equal to the minimized objective function (the minimized force rmsd). Figure 3 shows a plot of the rmsd of the forces as a function of the number of iterations. The figure shows that the FF optimization process converges after three iterations. In fact, two iterations should be sufficient for most cases.

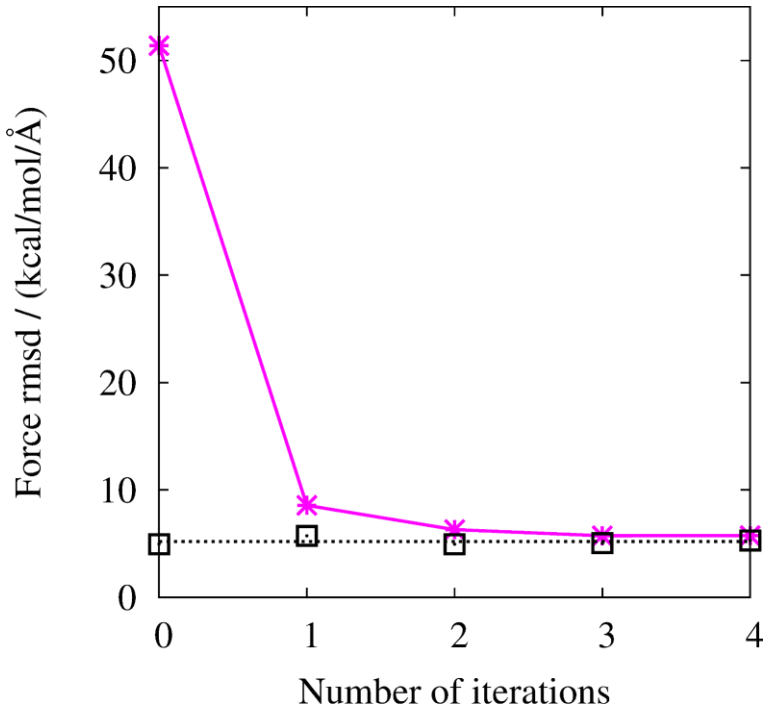


Figure 3: The rmsd of the forces (purple line with stars) and the minimized rmsd of the forces (black dotted line with squares) as a function of the number of iterations. In each iteration, the optimization process starts at the purple star and finishes at the black square. The whole optimization process converges when the star and the square merged.

The quality of our FF is evident in the scatter plot of Figure 4. This figure shows a comparison of the forces calculated using our FF and those from ab initio calculations. The figure was created with 100 configurations in the equilibrium ensemble, which were *not* included in the FF fitting. The red pluses are forces calculated using the initial guessed FF compared with ab initio forces. All the points would lie on the blue diagonal line if the FF predicts forces identical to the ab initio forces. Here, the rmsd of the forces between our FF and ab initio is ~5 kcal/mol/Å.

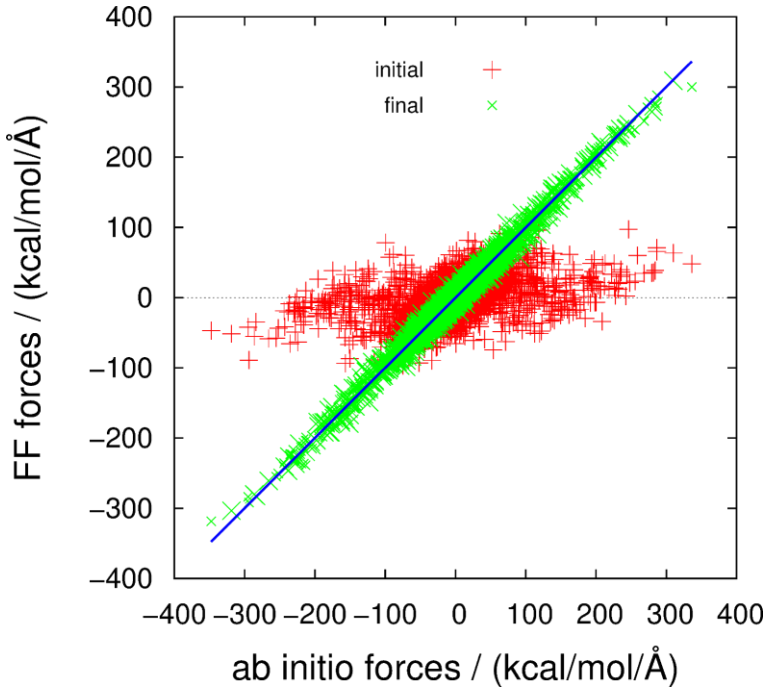


Figure 4. Scatter plot of the forces for the initial guessed FF (red pluses) and the first iteration optimized FF (green crosses). The *x*-axis corresponds to the ab initio forces. The rmsd of the forces is 5.03 kcal/mol/Å**.**

Although a scatter plot of the forces somewhat gives us an insight into the quality of the FF, it may not be very intuitive to appreciate whether a rmsd of the forces of the order of 5 kcal/mol/Å represent a good force field. To establish a concrete reference for this quantity, we perform a comparison wherein the rmsd between the FF forces and the B3LYP/6-31G* forces is compared with the rmsd between the forces computed using other electronic structure methods and the B3LYP/6-31G* target forces. The rmsd of the energies are also compared. Table 2 shows that AM1 gives the largest deviation for the forces and energies from those computed at B3LYP/6-31G*. Calculations performed at HF/6-31G* are slightly closer to the target and calculations at the B3LYP/3-21G* level are the closest to the B3LYP/6-31G* target. However the FF with optimized parameters yield forces that are closer to the B3LYP/6-31G* target than either HF/6-31G* or B3LYP/3-21G*.

Table 2. Forces and energies rmsd between calculations at B3LYP/6-31G* levels and other levels of theory or the optimized FF, for 100 arbitrary geometries not used in the FF optimization process.

Methods used	Forces rmsd (kcal/mol/ Å)	Energy rmsd (kcal/mol)
FF	5.03	1.17
AM1	23.21	3.50
HF/6-31G*	19.28	2.86
B3LYP/3-21G*	8.80	1.46

II. Examples of practical interest

A. Molecule forming semiconducting crystals, DCV-SN5

DCV-SN5 is a recently synthesised[68] molecule, which possesses exceptionally high power conversion efficiency of 6.5% when used as a donor in organic solar cells. To elucidate its unusual property, the dynamics of the charge and exciton transfer processes within the molecular crystal structure needs to be thoroughly understood and performing MD simulations is an important step in this type of study.[69] This molecule represents a classic example where any standard force field would be inaccurate. The complex fused ring structure with 5 heteroatoms makes it very difficult to extract parameters by comparing with similar molecules. The out of plane dynamics is likely very important for the electron and exciton dynamics and needs to be captured correctly.

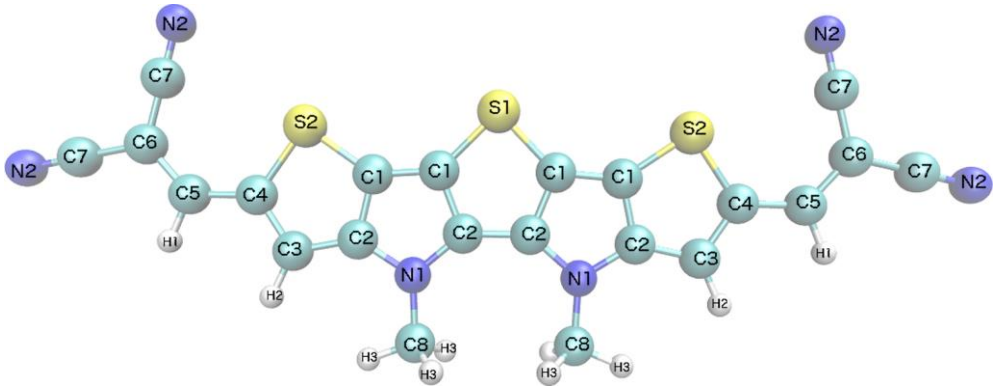


Figure 5. The geometry of the DCV-SN5 molecule with the atom-type labelled.

Since the equilibrium geometry of the molecule is planar (other conformations cannot be visited at room temperature and certainly not in the crystal phase) we do not optimize the phase of the torsional potential as we illustrated for the DPP molecule. Here, we have employed the CHARMM FF style for this molecule. We use 150 structures for the optimization process. The

rmsd of the forces and energies after three iterations are 4.61 kcal/mol/Å and 2.74 kcal/mol, respectively. The force rmsd is slightly smaller than that of the DPP molecule even though DVC-SN5 is much larger and has nearly doubles the number of parameters of DPP. Even though our FF is designed to match the ab initio forces, it gives excellent rmsd also when the equilibrium bonds and angles are compared with the ab initio counterpart (Table 3).

Table 3. Summary of the results of the FF parameterization for DCV-SN5 and TT-DPP.

	N_a	N_p	N_s	Force rmsd (kcal/mol/Å)	Bond rmsd (Å)	Angle rmsd (°)	Energy rmsd (kcal/mol)
DCV-SN5	41	131	150	4.611	0.0035	0.703	2.74
TT-DPP	31	165	150	6.245	0.0054	0.827	2.74

B. TT-DPP Semiconducting Polymer

The structure property relationship of semiconducting polymers is still poorly understood and, in particular, it is still not clear how the polymer disorder influences the electronic structure and the charge mobility.[70] For many amorphous polymers molecular simulations are the main tools to formulate hypotheses on the microscopic structure[3-7] and these are severely limited (among other things) by the lack of reliable force fields for the ever-increasing number of oligomers used. Thus, in this section we report the FF parameters for a DPP based polymer TT-DPP (Figures 6) constructed using our FM approach. The polymer is one of the most studied members of a new generation of semiconducting polymers with extremely promising charge mobility and efficiency in organic solar cells.[67] The model shown in Figure 6 contains all parameters needed for the simulation of an oligomer of TT-DPP of arbitrary length (the alkyl side chains, for which plausible parameters exists, have been removed). To be consistent with our previous works on semiconducting polymers,[3, 6, 71] here we parameterize the FFs for these polymers using the OPLS FF style.

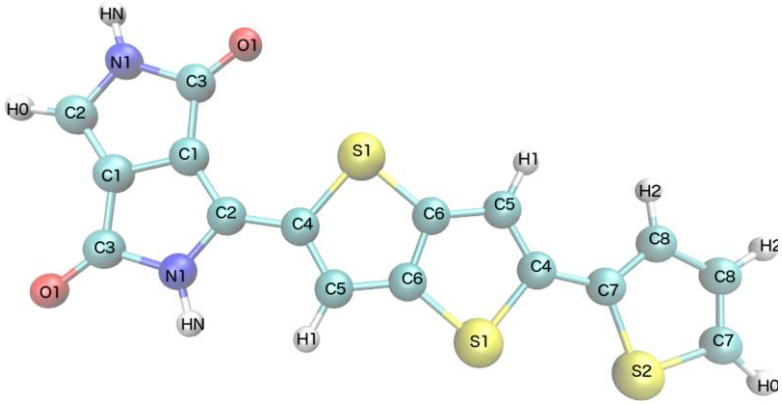


Figure 6. The geometry of a monomer of the TT-DPP semiconducting polymer with the atom-type labelled.

Unlike the rigid molecule considered in the previous section TT-DPP contains flexible torsional degrees of freedom between the fused rings. We proceed by dividing the FF optimization procedure into two stages. The first stage is to obtain all of the parameters apart from those torsion parameters for the flexible dihedral angles and the second stage is to obtain the missing torsion parameters. In the first stage, the dihedral angles whose parameters are not optimized are kept fixed at 0° or 180°. This geometry is typically close (or identical) to the equilibrium geometry and, by symmetry, the forces that tend to displace the molecule along the frozen degree of freedom are zero. In the TT-DPP system, there are two flexible dihedral angles C1-C2-C4-C5 and C5-C4-C7-C8. The former is fixed at 180° and the latter is frozen at 0° (i.e. in the conformation of Figure 6). The FF optimization follows exactly the same procedure as those of DPP and DCV-SN5 and is described in details in the method section (section 3). Having obtained the FF parameters from the first stage, we then perform a torsional scan to derive the torsion potentials for each missing flexible dihedral angle. This is done by first carrying out an ab initio scan (constrained optimization) at the B3LYP/6-31G* level of theory then followed by a "classical" scan using the FF derived from the first stage. The torsion potential for the dihedral angle of interest is the difference between the two scans as illustrated in Figure 7 for the flexible dihedral angle (C1-C2-C4-C5). The torsion potential is finally fitted to a truncated cosine series to obtain the coefficients (the energy barriers) following the OPLS FF style.

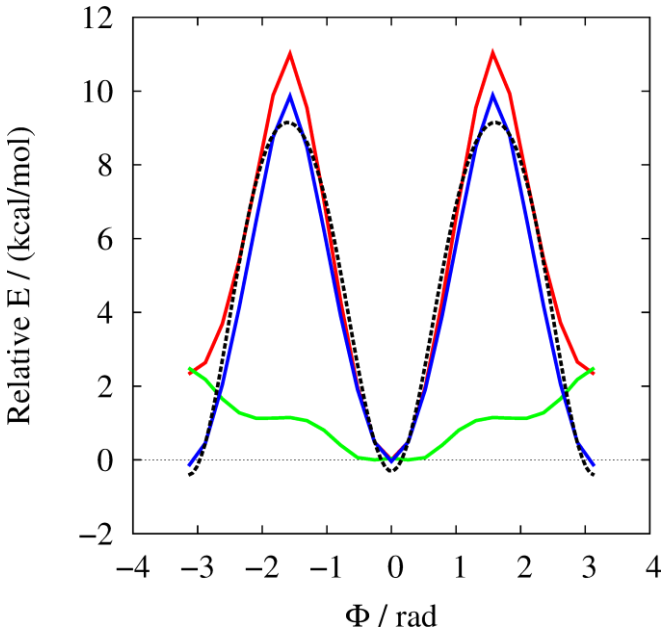


Figure 7. Plots of a torsional scan for the C1-C2-C4-C5 dihedral angle in TT-DPP. The vertical-axes is the relative energies with respect to the minimum energy. The red line is the ab initio scan, the green line is the FF scan and the blue line is the torsion potential, i.e. the difference between the ab initio and FF scans. The dotted black line is the fit of the torsion potential to an OPLS truncated cosine series for the C1-C2-C4-C5 dihedral angle in TT-DPP.

When one deals with a very large molecule that requires a long optimization time, it may be necessary to cut the molecule into smaller segments. Take the TT-DPP as an example. We can cut this molecule say into three smaller segments DPP, thienothiophene (TT), and thiophene (T). Then, we could parameterize the FF for each of these small segments individually and finally join them together by running a FF optimisation for a dimer e.g. DPP-TT/TT-T and only allow the missing parameters to be optimised while the rest keep fixed at those obtained from the FF optimization of the monomers. Another way is to optimize the FF parameters for the dimers DPP-TT and TT-T. However, this will leave us with two sets of parameters for TT segments to choose from, which may not be ideal. Nevertheless, we found that, for the sizes of the molecules presented in this study, it is possible to do the FF optimization for the whole molecule in one go, and therefore the truncated approaches was not investigated in depth. Considering that the number of ab initio calculations is approximately constant, the computational cost for parameterizing the force field is proportional to the cost of a single point force calculation, scaling at worst as the cube of the number of atoms for DFT calculations. For example, the optimization of the DPP force field costs about five CPU hours, suggesting that the complete force field optimization is easily achievable for most commonly encountered problems.

Table 3 summarises the energetic and structural results obtained using our new FF for the semiconducting polymers TT-DPP. When the equilibrium properties are considered the rmsd of the bonds is close to 0.01 Å and that of the angles is close to 1 degree, i.e. very small and only slightly larger than those found for the test molecule DPP.

C. DBV – A Chromophore Embedded in the Protein Environment

DBV is a chromophore that is found in the PE545 complex which is a primary antenna of the cryptophyte algae *Rhodomonas CS24* that live in both marine and freshwater environments.[72] These organisms exhibit maximal photosynthetic activity at very low light intensities even though they have fewer types of antenna proteins than others and also given the large average center-to-center separations of chromophores in their antenna.[16] Thus, this complex has attracted much attention and has been characterised by several theoretical studies,[16, 18] which employ MD simulations as a prerequisite step. In these works, standard transferable FFs such as GAFF[25] are often used. However, as outlined in our introduction, for this kind of system, transferable FFs may not be able to capture the structural properties and may affect the quality of the post-quantum calculations. Therefore, for these applications we seek to improve their intramolecular interaction with the assumption that the protocol developed in CHARMM for the intermolecular interaction is valid (and in any case its improvement is outside the scope of this work). Here, we show an example of how to utilize the FM procedure to tailor an accurate FF for chromophores embedded in the protein environment by minimizing the rmsd of the FF and the ab initio forces.

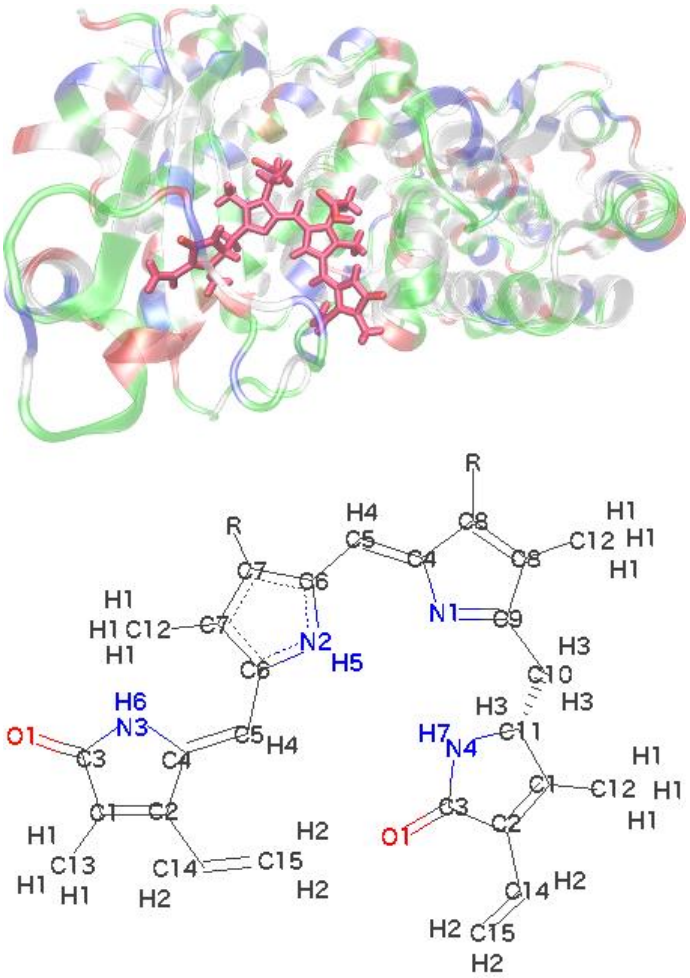


Figure 8. The chemical structure of the DBV molecule with the atom-type labelled. The -R group is the -CH_2CH_2COOH group and is not optimized in our FF. The upper panel shows the chromophore inside the protein.

The procedure for parameterizing the FF for DBV is slightly more complicated than that of the other systems. Since we want to have a FF that best describes the DBV within the protein environment, the equilibrium ensemble of the structures of the chromophore are drawn from the actual MD simulations of the whole system, which include both the protein scaffold and the solvent (water) at ambient condition (T = 300 K and P = 1 atm). In other words, the FF is designed specifically to reproduce the forces and hence the structural properties of the chromophore when embedded inside the protein with its main conformation locked. The upper panel in Figure 8 shows a snapshot of a DBV chromophore inside the PE545 (PDB code 1XG0[72]). The protein complex is surrounded by 13000 water molecules in a cubic periodic box. CHARMM[26] FF is employed for the protein and TIP3P is used for water. The system is equilibrated for 1 ns followed by 100 ps in which the trajectories are stored every 50 fs. Prior to the MD simulations, the geometry of the molecule is optimized in vacuum at the B3LYP/6-31G* level of theory and its point charges are computed using HF/6-31G* (to be consistent with CHARMM charges). The atom types are chosen based on the local environment around the atoms i.e. by judging the ab initio equilibrium bonds and angles that involved the atom. The atom type selection is given in the lower panel in Figure 8 and results in a total of 574 parameters to be optimized (for this extremely large case we performed a single iteration of the

optimization procedure from the initial guessed FF). The FM procedure is performed using forces computed in the gas phases on the structures generated from the MD simulations including the protein environment. For those dihedral angles which have the equilibrium values close to either 0° or 180° (≤ 4°), their phases are not optimized and fixed at 180°. The rmsd between the FF and B3LYP/6-31G* forces is 6.65 kcal/mol/Å, which is aligned with results obtained for the other molecules considered. The error seems therefore rather an intrinsic limitation of the force field analytical form that can be improved with the introduction or more complicated expressions of the potential energy, which can then be fitted with the same method proposed here.

5. Conclusion

We have proposed a rapid method to parameterize the intramolecular components of classical force field for complex conjugated molecules such as those commonly encountered in organic electronic materials and spectroscopy of biological molecules. We have used for this problem a force matching procedure and demonstrated the flexibility of the approach by parameterizing medium-sized rigid molecules, oligomers of semiconducting polymers, and chromophores embedded in a protein environment. The systems considered required the optimization from few tens to few hundreds force field parameters in parallel and exemplify the typical cases for which the proposed method is useful. The method is particularly suitable for those applications where MD simulations are used to generate structures that are therefore analysed with electronic structure methods, because it is possible to build force fields that are consistent with the electronic structure calculations that follow. This is particularly important for the calculation of excited states where innovations are constantly introduced in electronic structure theory and one may need a rapid way to reparametrize the FF to align it to new electronic structure methods.

Since we have considered systems for which it is very difficult to find FF parameters by transferring them from related molecule, our optimized force field is, almost by construction, not transferable to other molecules, unless they share the same conjugated core. A procedure to develop more transferable FFs can be based on similar ideas by defining a large set of molecules sharing the same parameters and optimizing the global objective function for all of them.

In this paper we have considered only standard force fields such as CHARMM or AMBER to illustrate immediately how one can generate better parameters than those commonly used for the same type of problem and obtained by parameter-transferring procedures such as GAFF[25] or CGenFF.[26] However the procedure can be used to parameterize different and more general functional forms of the FF, for example including coupling between degrees of freedom and anharmonicity (i.e. more fitting parameters). The residual difference between ab initio and empirical forces can be reduced by improving the FF functional form but a different type of study with smaller molecules and a range of functional forms would be needed to follow this route.

Our iterative scheme based on sampling from a gradually improving force field can explore a large uncorrelated sample of structures with the correct thermal weighting. There are cases where this procedure fails, e.g. if one is interested in describing the reactivity with a force field (or, in general, any process determined by high energy barriers). In these cases the various

procedures used to fit the *global* ab initio PES are more appropriate and few examples are given by refs. [73-75] for systems with fewer degrees of freedom, sometime also based on force matching. It is not surprising that the most convenient strategy for FF optimization depends on the number of degrees of freedom of the chemical system and the physics that one needs to model within that a classical FF. It is hoped that the tool we proposed in this paper can simplify, accelerate and improve the computational study of medium-to-large sized organic conjugated molecules, which requires a specialist but increasingly important use of MD simulations.

Acknowledgments. We are grateful to ERC for funding this work (Grant "ESTYMA").

References

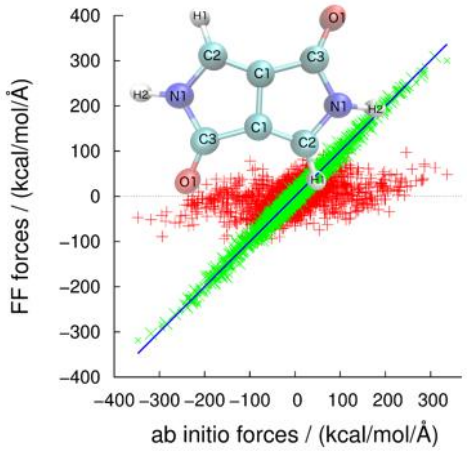
1. D. Andrienko, in *Supramolecular Materials for Opto-Electronics*, ed. N. Koch, The Royal Society of Chemistry, Cambridge, 2015, DOI: 10.1039/9781782626947-00309, pp. 309-362.
2. W. Barford, D. G. Lidzey, D. V. Makhov and A. J. H. Meijer, *J. Chem. Phys.*, 2010, 133, 044504.
3. T. Qin and A. Troisi, *J. Am. Chem. Soc.*, 2013, 135, 11247-11256.
4. M. Mladenović and N. Vukmirović, *Adv. Funct. Mater.*, 2015, 25, 1915-1932.
5. S. Athanasopoulos, J. Kirkpatrick, D. Martínez, J. M. Frost, C. M. Foden, A. B. Walker and J. Nelson, *Nano Lett.*, 2007, 7, 1785-1788.
6. T. Liu and A. Troisi, *Adv. Funct. Mater.*, 2014, 24, 925-933.
7. Y. Olivier, L. Muccioli, V. Lemaur, Y. H. Geerts, C. Zannoni and J. Cornil, *J. Phys. Chem. B*, 2009, 113, 14102-14111.
8. A. Troisi and G. Orlandi, *J. Phys. Chem. A*, 2006, 110, 4065-4070.
9. J. Kirkpatrick, V. Marcon, J. Nelson, K. Kremer and D. Andrienko, *Phys. Rev. Lett.*, 2007, 98, 227402.
10. N. Vukmirović and L.-W. Wang, *J. Phys. Chem. B*, 2009, 113, 409-415.
11. G. S. Engel, T. R. Calhoun, E. L. Read, T.-K. Ahn, T. Mancal, Y.-C. Cheng, R. E. Blankenship and G. R. Fleming, *Nature*, 2007, 446, 782-786.
12. H. Lee, Y.-C. Cheng and G. R. Fleming, *Science*, 2007, 316, 1462-1465.
13. E. Collini, C. Y. Wong, K. E. Wilk, P. M. G. Curmi, P. Brumer and G. D. Scholes, *Nature*, 2010, 463, 644-U669.
14. S. Shim, P. Rebentrost, S. Valleau and A. Aspuru-Guzik, *Biophysical Journal*, 2012, 102, 649-660.
15. J. Huh, S. K. Saikin, J. C. Brookes, S. Valleau, T. Fujita and A. Aspuru-Guzik, *J. Am. Chem. Soc.*, 2014, 136, 2048-2057.
16. C. Curutchet, J. Kongsted, A. Muñoz-Losa, H. Hossein-Nejad, G. D. Scholes and B. Mennucci, *J. Am. Chem. Soc.*, 2011, 133, 3078-3084.
17. L. Viani, C. Curutchet and B. Mennucci, *J. Phys. Chem. Lett.*, 2013, 4, 372-377.
18. M. Aghtar, J. Strümpfer, C. Olbrich, K. Schulten and U. Kleinekathöfer, *J. Phys. Chem. Lett.*, 2014, 5, 3131-3137.
19. N. L. Allinger, Y. H. Yuh and J. H. Lii, *J. Am. Chem. Soc.*, 1989, 111, 8551-8566.

20. V. Marcon and G. Raos, *J. Phys. Chem. B*, 2004, 108, 18053-18064.
21. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *Journal of Computational Chemistry*, 1983, 4, 187-217.
22. J. Wang, P. Cieplak and P. A. Kollman, *Journal of Computational Chemistry*, 2000, 21, 1049-1074.
23. S. J. Weiner, P. A. Kollman, D. T. Nguyen and D. A. Case, *Journal of Computational Chemistry*, 1986, 7, 230-252.
24. S. Lifson and A. Warshel, *J. Chem. Phys.*, 1968, 49, 5116.
25. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *Journal of Computational Chemistry*, 2005, 26, 114-114.
26. K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. Mackerell, *Journal of Computational Chemistry*, 2010, 31, 671-690.
27. K. Vanommeslaeghe, M. Yang and A. D. MacKerell, *Journal of Computational Chemistry*, 2015, 36, 1083-1101.
28. L. Huang and B. Roux, *Journal of Chemical Theory and Computation*, 2013, 9, 3543-3556.
29. V. Barone, I. Cacelli, N. De Mitri, D. Licari, S. Monti and G. Prampolini, *Phys, Chem. Chem. Phys.*, 2013, 15, 3736-3751.
30. I. Cacelli, A. Cimoli, P. R. Livotto and G. Prampolini, *Journal of Computational Chemistry*, 2012, 33, 1055-1067.
31. C. G. Mayne, J. Saam, K. Schulten, E. Tajkhorshid and J. C. Gumbart, *Journal of Computational Chemistry*, 2013, 34, 2757-2770.
32. S. Grimme, *Journal of Chemical Theory and Computation*, 2014, 10, 4497-4514.
33. S. K. Burger, M. Lacasse, T. Verstraelen, J. Drewry, P. Gunning and P. W. Ayers, *Journal of Chemical Theory and Computation*, 2012, 8, 554-562.
34. O. Wise and O. Coskuner, *Journal of Computational Chemistry*, 2014, 35, 1278-1289.
35. J. K. Bristow, D. Tiana and A. Walsh, *Journal of Chemical Theory and Computation*, 2014, 10, 4644-4652.
36. M. A. Addicoat, N. Vankova, I. F. Akter and T. Heine, *J. Chem. Theor. Comput.*, 2014, 10, 880-891.
37. L. Vanduyfhuys, S. Vandenbrande, T. Verstraelen, R. Schmid, M. Waroquier and V. Van Speybroeck, *J. Comput. Chem.*, 2015, 36, 1015-1027.
38. F. Ercolessi and J. B. Adams, *EUROPHYSICS LETTERS*, 1994, 26, 583-588.
39. B. A. a. F. E. a. J. A. M. Xiang-Yang Liu and James, *Modelling and Simulation in Materials Science and Engineering*, 1996, 4, 293.
40. C. M. Handley and R. J. Deeth, *Journal of Chemical Theory and Computation*, 2012, 8, 194-202.
41. R. Wu, Z. Lu, Z. Cao and Y. Zhang, *J. Chem. Theor. Comput.*, 2011, 7, 433-443.
42. L. B. Wright, P. M. Rodger and T. R. Walsh, *RSC Adv.*, 2013, 3, 16399-16409.
43. A. Jaramillo-Botero, S. Naserifar and W. A. Goddard, *Journal of Chemical Theory and Computation*, 2014, 10, 1426-1439.
44. C. Knight, C. M. Maupin, S. Izvekov and G. A. Voth, *J. Chem. Theor. Comput.*, 2010, 6, 3223-3232.
45. T. G. A. Youngs, M. G. Del Pópolo and J. Kohanoff, *J. Phys. Chem. B*, 2006, 110, 5697-5707.
46. J. Sala, E. Guàrdia and M. Masia, *Comp. Phys. Comm.*, 2011, 182, 1954-1957.
47. A. Gabrieli, M. Sant, P. Demontis and G. B. Suffritti, *Micropor. & Mesopor. Mater.*, 2014, 197, 339-347.

48. S. Izvekov, M. Parrinello, C. J. Burnham and G. A. Voth, *J. Chem. Phys.*, 2004, 120, 10896.
49. S. Izvekov and G. A. Voth, *J. Phys. Chem. B*, 2005, 109, 6573-6586.
50. P. Maurer, A. Laio, H. W. Hugosson, M. C. Colombo and U. Rothlisberger, *J. Chem. Theor. Comput.*, 2007, 3, 628-639.
51. K. Spiegel, A. Magistrato, P. Maurer, P. Ruggerone, U. Rothlisberger, P. Carloni, J. Reedijk and M. L. Klein, *J. Comput. Chem.*, 2008, 29, 38-49.
52. M. Doemer, P. Maurer, P. Campomanes, I. Tavernelli and U. Rothlisberger, *J. Chem. Theor. Comput.*, 2014, 10, 412-422.
53. Y. Zhou and J. Pu, *J. Chem. Theor. Comput.*, 2014, 10, 3038-3054.
54. O. Akin-Ojo, Y. Song and F. Wang, *J. Chem. Phys.*, 2008, 129, 064108.
55. F. Wang, O. Akin-Ojo, E. Pinnick and Y. Song, *Mol. Simul.*, 2011, 37, 591-605.
56. L.-P. Wang, J. Chen and T. Van Voorhis, *J. Chem. Theor. Comput.*, 2013, 9, 452-460.
57. L.-P. Wang, T. J. Martinez and V. S. Pande, *J. Phys. Chem. Lett.*, 2014, 5, 1885-1891.
58. R. M. Betz and R. C. Walker, *J. Comput. Chem.*, 2015, 36, 79-87.
59. D. Wei, Y. Song and F. Wang, *J. Chem. Phys.*, 2011, 134, 184704.
60. W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, 118, 11225-11236.
61. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten, *J. Comput. Chem.*, 2005, 26, 1781-1802.
62. W. D. Cornell, P. Cieplak, C. I. Bayly and P. A. Kollmann, *J. Am. Chem. Soc.*, 1993, 115, 9620-9631.
63. C. M. Breneman and K. B. Wiberg, *J. Comput. Chem.*, 1990, 11, 361-373.
64. Q. Wang, J. A. Rackers, C. He, R. Qi, C. Narth, L. Lagardere, N. Gresh, J. W. Ponder, J.-P. Piquemal and P. Ren, *J. Chem. Theor. Comput.*, 2015, 11, 2609-2618.
65. Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder and P. Ren, *J. Chem. Theor. Comput.*, 2013, 9, 4046-4063.
66. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. Montgomery, J. A., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. and Pople.
67. H. Bronstein, Z. Chen, R. S. Ashraf, W. Zhang, J. Du, J. R. Durrant, P. Shakya Tuladhar, K. Song, S. E. Watkins, Y. Geerts, M. M. Wienk, R. A. J. Janssen, T. Anthopoulos, H. Sirringhaus, M. Heeney and I. McCulloch, *J. Am. Chem. Soc.*, 2011, 133, 3272-3275.
68. A. Mishra, D. Popovic, A. Vogt, H. Kast, T. Leitner, K. Walzer, M. Pfeiffer, E. Mena-Osteritz and P. Bäuerle, *Adv. Mater.*, 2014, 26, 7217-7223.
69. J. Aragó and A. Troisi, *Phys. Rev. Lett.*, 2015, 114, 026402.
70. R. Noriega, J. Rivnay, K. Vandewal, F. P. V. Koch, N. Stingelin, P. Smith, M. F. Toney and A. Salleo, *Nat Mater*, 2013, 12, 1038-1044.

71. D. L. Cheung, D. P. McMahon and A. Troisi, *J. Phys. Chem. B*, 2009, 113, 9393-9401.
72. A. B. Doust, C. N. J. Marai, S. J. Harrop, K. E. Wilk, P. M. G. Curmi and G. D. Scholes, *J. Mol. Biol.*, 2004, 344, 135-153.
73. A. Pukrittayakamee, M. Malshe, M. Hagan, L. M. Raff, R. Narulkar, S. Bukkapatnum and R. Komanduri, *J. Chem. Phys.*, 2009, 130.
74. J. Chen, X. Xu, X. Xu and D. H. Zhang, *J. Chem. Phys.*, 2013, 138.
75. M. Gastegger and P. Marquetand, *J. Chem. Theor. Comput.*, 2015, 11, 2187-2198.

TOC Graphics and text



A rapid method to parameterize the intramolecular component of classical force fields is proposed and applied to a molecular semicoductor, oligomers of conjugated polymers and a biological chromophore.