

# Systematic and Automated Development of Quantum Mechanically Derived Force Fields: the Challenging Case of Halogenated Hydrocarbons

Giacomo Prampolini<sup>a,\*</sup>, Marco Campetella<sup>b</sup>, Nicola DeMitre<sup>c</sup>,  
Paolo Roberto Livotto<sup>d</sup>, and Ivo Cacelli<sup>a,b</sup>

<sup>a</sup>*Istituto di Chimica dei Composti OrganoMetallici (ICCOM-CNR),  
Area della Ricerca, via G. Moruzzi 1, I-56124 Pisa, Italy*

<sup>b</sup>*Dipartimento di Chimica e Chimica Industriale, Università di Pisa,  
Via G. Moruzzi 13, I-56124 Pisa, Italy*

<sup>c</sup>*Department of Chemistry, University of Cambridge ,  
Lensfield road, Cambridge CB2 1EW, UK*

<sup>d</sup>*Instituto de Química, Universidade Federal do Rio Grande do Sul,  
Avenida Bento Gonçalves 9500, CEP 91501-970 Porto Alegre, Brazil*

August 29, 2016

---

\*Corresponding author

## Abstract

A robust and automated protocol for the derivation of sound force field parameters, suitable for condensed phase classical simulations, is here tested and validated on several halogenated hydrocarbons, a class of compounds for which standard force fields have often been reported to deliver rather inaccurate performances. The major strength of the proposed protocol is that all the parameters are derived from first principles only, since all the information required is retrieved from quantum mechanical data, purposely computed for the investigated molecule. This *a priori* parameterization is carried out separately for the intra- and inter-molecular contributions to the force fields, respectively exploiting the JOYCE and PICKY programs, previously developed in our group. To avoid high computational costs, all quantum mechanical calculations were performed exploiting the density functional theory. Since the choice of the functional is known to be crucial for the description of the intermolecular interactions, a specific procedure is proposed, which allows for a reliable benchmark of different functionals against higher level data. The intramolecular and intermolecular contribution are eventually joined together, and the resulting quantum mechanically derived force field is thereafter employed in lengthy molecular dynamics simulations to compute several thermodynamic properties that characterize the resulting bulk phase. The accuracy of the proposed parameterization protocol is finally validated by comparing the computed macroscopic observables with the available experimental counterparts. It is found that, on average, the proposed approach is capable of yielding a consistent description of the investigated set, often outperforming the literature standard force fields, or at least delivering results of similar accuracy.

# 1 Introduction

The success of molecular simulations, which are nowadays routinely adopted in many fields of chemistry, ranging from simple liquids<sup>1,2</sup> to advanced materials<sup>3</sup> and biological systems,<sup>4,5</sup> is certainly rooted in their capability to bridge the gap between an atomistic description and the macroscopic observables exhibited by the condensed phase. Despite the most accurate framework to handle atoms and molecules is certainly quantum mechanics (QM), accurate estimates of the macroscopic properties often require the exploration of a wide portion of a complex phase space, and their convergence is reached only when a large number of particles is included in the calculations. Therefore, computational convenience and feasibility suggest to resort to techniques based on classical mechanics, as Monte Carlo (MC) and molecular dynamics (MD).<sup>1,2</sup> In these approaches, the chemical identity of the investigated system under study is encoded in the force field (FF): a collection of analytical functions of the positions of the nuclei aimed to describe the energy of the simulated system.

In the past decades, mainly because of the unavailability of proper computational resources, the majority of popular FFs<sup>1,2,4,6-12</sup> were first parameterized using very simple functional forms and heavily relying on experimental data. On the one hand, the main advantage of such procedures was to set up a collection of transferable parameters, which in principle made simulations possible for any target system, without any additional computational expense other than the simulation itself. On the other hand the reliance on empirical parameterization does not ensure the same level of accuracy when the system of interest is characterized by chemical (as peculiar substituents or different solvation) or physical (*e.g.* not standard thermodynamic conditions or external fields) features different from those for which they were developed. This drawback often requires a refinement<sup>4,11,13-15</sup> of the original set of empirical parameters, or, at least, extensive comparative benchmarks,<sup>16-20</sup> to *a posteriori* select the FF giving the best performances.

In this framework, the massive increase of computational resources has given new impulse to an alternative route, which relies on the possibility to parameterize the whole FF solely based on information retrieved through QM calculations, specifically performed on the target systems. As recently noted by Jorgensen’s group,<sup>15</sup> such a choice implies

abandoning the idea of transferability in favor of the higher accuracy<sup>21</sup> that should derive from the FF specificity. From a more fundamental point of view, these quantum mechanically derived force fields (QMD-FF)<sup>22,23</sup> enforce, in a certain sense, the aforementioned bridging role of computer simulations between the micro- and macro-scales. The growing attention to novel *and automated* parameterization strategies, *either* based solely on QM data, <sup>15,22-45</sup> *or including also experimental information ( see for instance [46-48] and references therein)* is testified by the increasing number of papers concerning this subject. *Besides increasing the accuracy of the standard pairwise additive FFs, more complex FF expressions have been recently introduced,<sup>34,36,37,43-45,49,50</sup> which are capable of going beyond the two body approximation, and explicitly accounting for many-body effects.<sup>51</sup>* Clearly, although promising, most of the procedures that lead to QMD-FFs should be thoroughly validated, by comparing their performances both with transferable FFs and the available experimental data.

In this work a QMD-FF parameterization protocol is tested and validated through the aforementioned comparison for a number of target molecules. *Despite the dimensions of the molecules considered for validating the approach are rather small, it might be worth mentioning that the here discussed QMD-FF procedure was developed in view of applications on much more complex and large systems. For this reason, since the vast majority of MD studies on condensed phase systems is carried out through standard FFs, second order (albeit important) effects, as many body contributions, and FF expressions more complex than the formulae implemented in popular MD engines were not considered in this work.*

~~The~~ *Concretely, the* QMD-FF is obtained essentially by joining the intra- and inter-molecular FF parameters derived solely from QM data by means of the JOYCE<sup>22,33</sup> and PICKY<sup>32</sup> schemes, implemented in recent years in our laboratory. This procedure was first<sup>40</sup> tested on the benzene molecule, finding thermodynamic, structure and dynamic properties of the liquid phase in good agreement with both experimental measures and the recently reviewed<sup>16</sup> results obtained with popular empirical FFs. Here we continue the validation process, applying our QMD-FF parameterization route to a new set of target molecules, chosen among a more challenging class of compounds. The aim of the present study is three-fold:

- i) To test the validity of the JOYCE/PICKY protocol in producing QMD-FFs whose performances should be comparable or, hopefully, better than empirical transferable FFs.
- ii) To further<sup>40</sup> investigate on the relationship between the quality of the QM description of the two body interaction potential energy surface (IPES) and the final thermodynamic properties of the connected condensed phase.
- iii) To apply the protocol on a class of rather similar compounds, and investigate how the fitted parameters reflect chemical diversity/similarity, thus verifying the degree of transferability of the proposed QMD-FFs within this class.

While JOYCE has been tested for many different compounds and in several applications,<sup>22,30,32,33,52–58</sup> the PICKY protocol has been validated only for pyridine<sup>32</sup> and, more recently, benzene.<sup>40</sup> Yet, one of the strengths of the proposed protocol and, more in general, of any QMD-FF, is the intrinsic capability to predict the thermodynamic macroscopic properties of a given substance in its condensed phases, regardless of the availability of experimental data. One step forward along this path may be traced by verifying if and at what level of accuracy QMD-FFs are able to handle those compounds where empirical FFs fail or at least give inaccurate results.

A comprehensive benchmark of the performances of three different popular FFs (namely GAFF,<sup>10</sup> OPLS<sup>11</sup> and CGenFF<sup>21</sup>) was performed on more than 150 small organic molecules, recently<sup>17,59</sup> reported in literature and published in the <http://virtualchemistry.org> website, thus building an extremely useful reference database for the development of new FFs. Therein, the accuracy and reliability of each of the investigated FFs are assessed from the comparison of a selected set of computed thermodynamic properties with their experimental counterparts. From the analysis of these published results, it appears that a challenging class of compounds is constituted by the halogenated hydrocarbons. For these compounds, the benchmarked FFs yield very sparse results, with performances that remarkably depend on the specific molecule, the choice of the parameters set and the considered property. For instance, the standard deviation computed on all investigated liquid compounds with respect to the experimental density is 88, 43 and 39 kg/m<sup>3</sup>, for GAFF,

OPLS and CGenFF respectively, whereas it increases 188, 81 and 56 kg/m<sup>3</sup> when the calculation is limited to halogenated compounds. Despite CGenFF, where halogenated

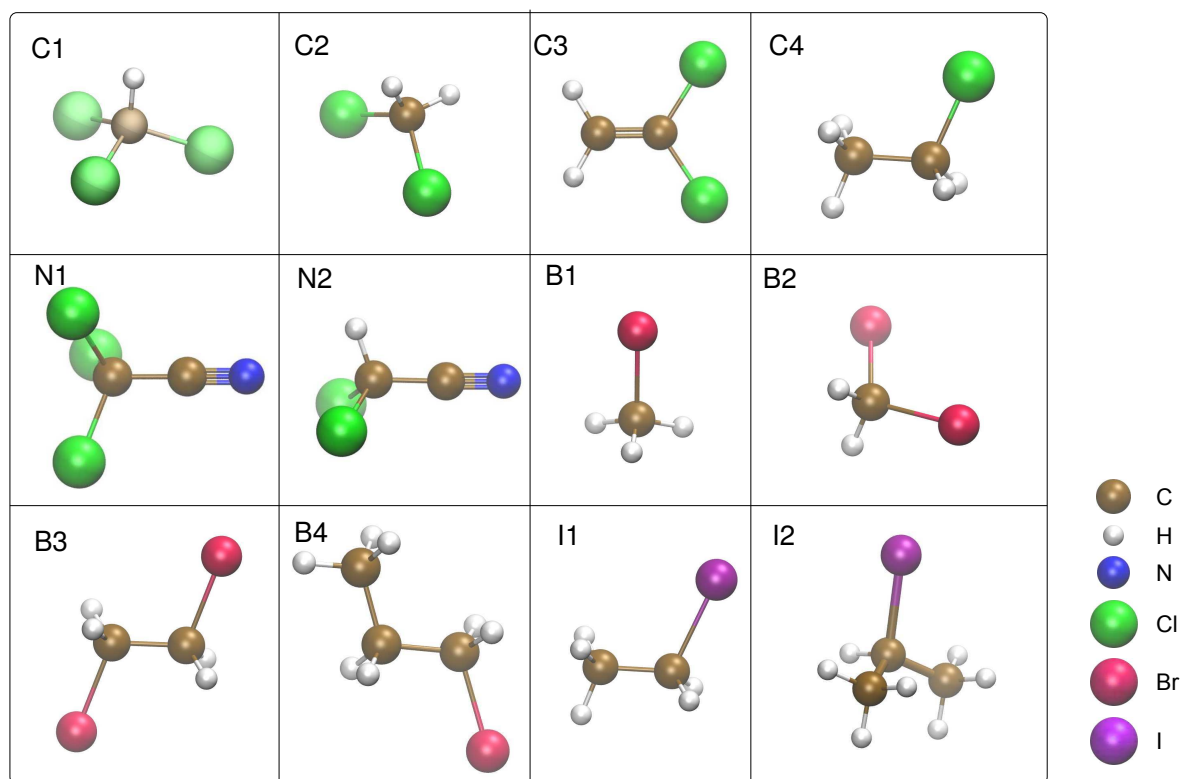


Figure 1: Benchmark set of halogenated hydrocarbons employed in this work. Top row: chlorinated compounds (**C1**: chloroform, CHCl<sub>3</sub>; **C2**: dichloromethane, CH<sub>2</sub>Cl<sub>2</sub>; **C3**: 1,1-dichloro-ethene, CH<sub>2</sub>=Cl<sub>2</sub>; **C4**: chloroethane, CH<sub>3</sub>CH<sub>2</sub>Cl); middle row: chlorinated acetonitriles (**N1**: trichloro-acetonitrile, CCl<sub>3</sub>CN and **N2**: dichloro-acetonitrile, CHCl<sub>2</sub>CN) and brominated compounds (**B1**: bromo-methane, CH<sub>3</sub>Br and **B2**: dibromo-methane, CH<sub>2</sub>Br<sub>2</sub>); bottom row: brominated (**B3**: 1,2-dibromo-ethane, BrCH<sub>2</sub>CH<sub>2</sub>Br and **B4**: 1-bromo-propane, CH<sub>3</sub>(CH<sub>2</sub>)<sub>2</sub>Br) and iodinated alkanes (**I1**: iodo-ethane, CH<sub>3</sub>CH<sub>2</sub>I and **I2**: 2-iodo-propane, CH<sub>3</sub>CHICH<sub>3</sub>).

hydrocarbons were explicitly included<sup>21</sup> in the set of molecules employed in the parameterization, seems to give on average the best results on the benchmarked properties,<sup>17,59</sup> for several of these compounds the reported accuracy is worse than that found for similar organic substances, as alkanes or alcohols.

In this work, exploiting the reference data<sup>17</sup> available on the <http://virtualchemistry.org> website, we intend to verify if and to what extent QMD-FFs are able to accurately reproduce the thermodynamic properties of a representative ensemble of halogenated hydro-

carbons. To this end, ten compounds, whose molecular formulas are displayed in Figure 1, have been selected among the halogenated hydrocarbons reported in References [17,59].

In effect, besides their importance as rather common substituents in both biological and synthetic compounds as dyes, drugs, *etc.*, part of the interest gained in recent years by halogenated compounds among computational chemists<sup>60</sup> is due to the challenging interpretation of the capability of atoms in group 17 to establish a specific type of interaction, namely the halogen bond.<sup>61</sup> Within these interactions, despite being substituents of high electronegativity, halogens attract Lewis base atoms or establish directional contacts among themselves, behaving both as donor and as acceptors.<sup>62</sup> The presence of a nucleophilic and an electrophilic region on the halogen’s surface has been explained in terms of  $\sigma$ -hole:<sup>63</sup> a region of positive electrostatic potential located on top of the halogen along the axis of the covalent Carbon-Halogen bond, surrounded by a negative (nucleophilic) region. Nevertheless, the point charges used for halogens in literature FFs are generally negative (in the -0.1 – -0.2 range), at least for single or double-substituted compounds, and are incapable of describing also the acid character of the  $\sigma$ -hole. To circumvent this problem, several different methods, that account for polarization<sup>64,65</sup> or an explicit  $\sigma$ -hole (ESH) through an additional massless interacting center<sup>66–68</sup> have been proposed and recently reviewed.<sup>60</sup> Nonetheless, as the authors of Ref. [67] have observed (see for instance Figure 5 therein), further tuning of the LJ terms seems to be required, as in general the repulsive branch of dimer interaction energy curves still suffers of a remarkable inaccuracy.<sup>67,69</sup> To overcome this difficulty, in Ref. [68] the LJ potential parameters are set to zero.

In this context, a QMD-FF parameterization route, which minimizes ~~the~~*objective functions based on the* difference between the FF and QM IPES (which correctly consider the halogen bond properties) could yield good average interaction potentials, through the simultaneous modification of both LJ parameters and partial charges. In particular, even without the exploitation of an ESH, the balance between the spherical short-ranged LJ term and the long-ranged electrostatic potential given by both the halogen and the nearest carbon atom, should be able to add part of the required anisotropy to the description of the system. To further investigate the behavior of QMD-FFs, based on standard

model potential functions (*i.e.* without considering polarization nor ESH), in presence of halogen bond forming molecules, two more targets were included in the benchmark set, *i.e.* trichloro- and dichloro-acetonitrile. In fact, in addition to the interactions between halogens, already accounted for in the rest of the benchmark set, these compounds provide an example of intermolecular contributions that stem from C-H $\cdots$ N and C-Cl $\cdots$ N contacts.<sup>70</sup> Standard transferable FFs parameters are not able to mimic such interactions, thus a more accurate FF seems required. Aside from this aspect, once the reliability of the present QMD-FF is validated, most of the thermodynamics properties that will be computed for **N1** and **N2** can be considered as reliable predictions, filling the gap caused by limited number of experimental data available for these last targets.

## 2 Computational Details

### 2.1 Parameterization Protocol

As previously reported,<sup>22,32,33,40</sup> the QMD-FF parameterization here proposed is based on optimizing the FF parameters which allow for a classical description of a given system of target molecules versus the representation obtained at a QM level. This is separately done for one isolated molecule ( $A$ ) and for an interacting dimer ( $AB$ ), exploiting the classical partition of the total FF energy ( $E_{AB}^{FF_{tot}}$ ) of a pair of molecules:

$$E_{AB}^{FF_{tot}} = E_A^{FF_{intra}}(\bar{b}_A, \bar{\theta}_A, \bar{\psi}_A) + E_B^{FF_{intra}}(\bar{b}_B, \bar{\theta}_B, \bar{\psi}_B) + E_{AB}^{FF_{inter}}(\bar{r}_{AB}) \quad (1)$$

$E_A^{FF_{intra}}$  and  $E_B^{FF_{intra}}$  drive respectively molecule  $A$  and molecule  $B$  flexibility, which in turn depends only on the collection of internal coordinates (bonds ( $\bar{b}$ ), angles ( $\bar{\theta}$ ) or dihedrals ( $\bar{\psi}$ )) of the related monomer. On the contrary,  $E_{AB}^{FF_{inter}}$  depends on the relative position and orientation of the two monomers, being a function of the set of distances  $\bar{r}_{AB}$  between all interaction sites of molecule  $A$  and those of molecule  $B$ . According to the procedures devised in our group, the intramolecular part of the QMD-FF,  $E^{FF_{intra}}$  is obtained through the JOYCE protocol,<sup>22,33</sup> while the intermolecular contribution,  $E^{FF_{inter}}$ , is optimized through the PICKY approach.<sup>32,40</sup> Both software packages are freely available at <http://www.pi.iccom.cnr.it/joyce> and at <http://www.pi.iccom.cnr.it/picky>. As shown in Figure 2, once parameterized, the intramolecular contribution (blue path on the left



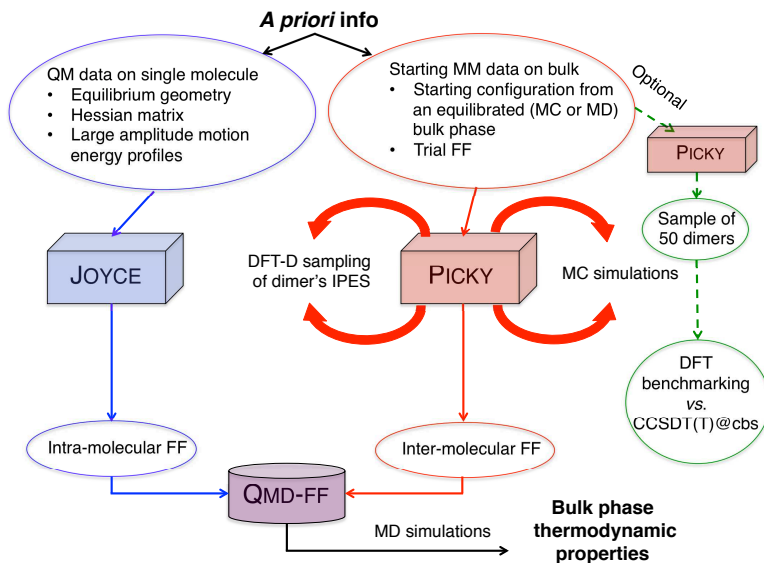


Figure 2: The parameterization route to QMD-FF proposed in this work for halogenated hydrocarbons. The intramolecular and intermolecular parameterization steps are highlighted with blue and red frames, while the (optional) CCSD(T)@cbs validation is displayed in green on the right.

side) and the intermolecular term (red path on the right side) are then joined together to build a QMD-FF. A brief description of the parameterization route to QMD-FF illustrated in Figure 2 is given in the following. Further details can be found in the Supporting Information and in previous works.<sup>22, 32, 33, 40</sup>

- i) For each compound,  $E^{FF_{intra}}$  is parameterized by means of the JOYCE protocol,<sup>22, 33</sup> using the optimized geometry, its Hessian matrix and, when required, a number of torsional energy scans as reference QM data. Further details of the intramolecular parameterization *and objective function* can be found in the Supporting Information.
- ii) Considering the scarce flexibility of most of the investigated compounds, the parameterization of  $E^{FF_{inter}}$  is performed following PICKY route II,<sup>40</sup> *i.e.* based on rigid MC simulations. Within this framework, the QMD-FF intermolecular parameters are obtained through an iterative procedure, which consists in minimizing at each step the squared difference between the QM interaction energies ( $\Delta E^{inter}$ ) computed over a set of dimer geometries extracted from MC simulations and the ones

computed by means of the QMD-FF intermolecular term ( $E^{FF_{inter}}$ ), parameterized in the previous cycle. In other words, a non linear fitting of the **functional objective function**

$$I^{inter} = \frac{\sum_{k=1}^{N_{geom}} [(\Delta E_k^{inter} - E_k^{FF_{inter}})^2] e^{-\alpha \Delta E_k^{inter}}}{\sum_{k=1}^{N_{geom}} e^{-\alpha \Delta E_k^{inter}}} \quad (2)$$

is performed, where  $N_{geom}$  is the number of geometries considered for the target dimer,  $\alpha$  a Boltzmann-like weight. The convergence of the iterative procedure is monitored by means of the quantity  $\Delta P$ , which is defined as

$$\Delta P = \left( \frac{1}{N_{geoms}} \sum_i^{N_{geoms}} [(E_i^{FF_{inter}})_c - (E_i^{FF_{inter}})_{c-1}]^2 \right)^{\frac{1}{2}} \quad (3)$$

where  $(E_i^{FF_{inter}})_c$  is the intermolecular energy computed with the QMD-FF parameters obtained in cycle  $c$ . Thus  $\Delta P$  measures the difference between the intermolecular energies obtained with the FF parameters of in two consecutive cycles. Further details for PICKY parameterization can be found in the original papers<sup>32,40</sup> and in the Supporting Information.

- iii) Once both the intra-molecular and the intermolecular contributions have been parameterized, they are joined together to form a QMD-FF, suitable for MD simulations.
- iv) Eventually, extended MD simulations are performed with the resulting QMD-FF, on systems composed of a large number of molecules in the NPT ensemble. Several macroscopic properties are then computed on the resulting bulk phases, and the reliability and the accuracy of the employed QMD-FF is validated by comparing the simulated properties with those computationally obtained adopting literature FFs and with their experimental counterparts.

## 2.2 QM Calculations

All QM calculation are performed with the Gaussian09 package.<sup>71</sup> Energy optimizations on isolated monomers are carried out for each of the compounds displayed in Figure 1 at

DFT level, employing the standard B3LYP functional with the Dunning’s correlation consistent cc-pv-Dz basis set. The global minimum is obtained by simultaneously optimizing all internal coordinates, storing the resulting Hessian matrix, necessary for JOYCE parameterization. When required, relaxed energy scans are performed by an optimization of all degrees of freedom, but the scanned coordinate.

Like the intramolecular term, also the intermolecular energy of a pair of monomers is computed at DFT level, but an empirical correction<sup>72</sup> is applied to the benchmarked functionals, to take dispersion interactions into account, exploring several combinations of functional, dispersion correction and basis set. Additional details and the specific combinations are given in Table **A** in the Supporting Information.

For a restricted number of the investigated species, namely **C1**, **C2**, **N1** and **B3**, the performances of each of the considered DFT-D functionals are benchmarked over a large set (50) of dimer arrangements against high level QM data. As reference, the so called ”gold standard”<sup>73</sup> of quantum chemistry is here adopted, *i.e.* the Coupled Cluster method, considering single, double and (perturbatively) triple excitations, estimated at the complete basis set limit (CCSD(T)@cbs). Additional details regarding QM calculations can be found in the Supporting Information.

## 2.3 Simulations

All MC simulations are carried out on systems of 216 molecules with an ”in house” code, named MCGBLJ.<sup>74</sup> All runs are performed in the isothermal isobaric ensemble (NPT), and the shape and the volume of the computational box is changed, by attempting to vary a randomly selected edge of the box. Only trial displacements of molecular centers of mass and rotations of main inertia axes are attempted during each MC run, in order to maintain the starting internal geometry of each monomer. All the aforementioned trial moves are selected randomly and not sequentially, to preserve the detailed balance condition. During all runs, the short range intermolecular interactions were truncated at  $R_c = 12$  Å, employing the standard energy correction. Charge-charge long range interactions are handled with the Ewald method, using a convergence parameter  $\alpha$  of  $5.36/2R_c$ .

MD simulations are performed resorting to the GROMACS5.1 engine.<sup>75</sup> Systems of

1000 molecules are first equilibrated for 1 ns in the NPT ensemble at a chosen temperature and at a pressure of 1 atm, keeping them constant through the Berendsen thermostat and barostat, respectively. Thereafter, 5 ns NPT production runs are carried out for each species, employing the velocity rescale *with a stochastic term*<sup>76</sup> and Parrinello-Rahman<sup>77</sup> schemes, setting  $\tau_T$  and  $\tau_P$  to 0.1 and 5.0 ps. In all simulations, *given the larger number of molecules (hence larger simulation box) considered in MD with respect to MC runs* a cut-off of 13 Å was employed for both charge-charge and LJ terms, whereas long-range electrostatics was accounted through the particle mesh Ewald (PME) procedure. In most runs, bond lengths are constrained at their equilibrium value using the LINCS algorithm, allowing for a 1 fs time step.

Finally, two additional 1 ns runs are carried out for each compound, saving the trajectories every 4 fs, one in the NPT and one in the NVE ensemble, in order to compute the specific heat and the viscosity, respectively, as detailed in the Supporting Information. In both simulations, the constraints on the bond lengths are removed, and the time step consequently decreased to 0.1 fs.

## 2.4 Thermodynamic Properties

The quality of the QMD-FFs is assessed by computing bulk density ( $\rho$ ), vaporization enthalpy ( $\Delta H^{vap}$ ), volumetric expansion coefficient ( $\alpha$ ), isothermal compressibility ( $\kappa_T$ ), specific heat at constant pressure ( $c_P$ ) and viscosity ( $\eta$ ) for each of the investigated systems. Details on the calculations of the above quantities from the MD trajectories can be found in the Supporting Information.

# 3 Results

## 3.1 Preliminary CAM-B3LYP-D results

The first attempt to parameterize halogenated compounds was performed on the two smallest molecules of the benchmark set displayed in Figure 1, namely **C1** ( $\text{CHCl}_3$ ) and **C2** ( $\text{CH}_2\text{Cl}_2$ ). In the light of the good results recently achieved with the benzene liquid,<sup>40</sup> this preliminary parameterization was carried out adopting the same combination

of DFT functional (CAM-B3LYP), basis set (6-311+G(2d,2p)) and dispersion correction (D3 Grimme’s correction<sup>78</sup>) adopted therein to describe the QM IPES. As far as the **C2** is concerned, dichloromethane condensed phase was reproduced rather well by the PICKY QMD-FF. For instance, NPT simulations at 298 K and 1 atm yielded an average density of 1297 kg/m<sup>3</sup> and a vaporization enthalpy of 28.5 kJ/mol, which well compares with the experimental<sup>79</sup> values (1327 kg/m<sup>3</sup> and 28.8 kJ/mol), and increases the accuracy<sup>17,59</sup> of the empirical FFs (1258 or 1200 kg/m<sup>3</sup> and 26.5 or 23.4 kJ/mol with GAFF or OPLS).

Unfortunately, a very different situation raised during the parameterization of chloroform (**C1**). In fact, after the fifth parameterization cycle, the MC density was stabilized around 1260 kg/m<sup>3</sup>, against an experimental<sup>80</sup> value of 1479 kg/m<sup>3</sup>, showing an error (> 15%) even worse ( $\sim 8\%$ ) than the one found with empirical FFs. This could be interpreted as a failure of the chosen functional in describing the chloroform IPES. Once again,<sup>40</sup> the problem is then shifted on assessing the quality of the IPES given by a certain functional. A straightward route to solve this dilemma could stand in carrying out a number of PICKY parameterizations based on different functionals and choose the one giving the best agreement with the target properties. This procedure has however at least two drawbacks. First, it is very time-consuming, given the plethora of DFT functionals available and the fact that the whole parameterization protocol has to be repeated for each of the benchmarked ones. Next, and probably more important from our point of view, a FF validation against experimental measures would somehow renege on the intention of maintaining the *a priori* nature of QMD-FFs, that could instead be exploited when such experimental information is scarce, hardly available or totally missing.

An alternative route can be devised though, resorting to the DFT benchmarking proposed in Ref. [40]. Therein, the performance of different DFT functionals with respect to reference CCSD(T)@cbs data was connected to the accuracy of the macroscopic bulk phase properties predicted by the MD simulations carried out with QMD-FFs parameterized over the tested functionals. The results reported in Ref. [40] showed that a sound validation of a given functional should not be assessed by comparing with higher level data for only one or few dimer arrangements, but instead carefully evaluated over a sig-

nificant number of geometries, possibly sampled taking into account their thermodynamic accessibility. Accordingly, a CCSD(T)@cbs database was built<sup>40</sup> therein for the investigated molecule, by computing the interaction energies of the 50 dimers sampled in the first PICKY cycle. Following this idea, a PICKY sampling is here performed on the final configuration of the initial MC run (see Section 2.1) of four test molecules, chosen among Figure 1 benchmark set, namely **C1**, **C2**, **N1** and **B3**.

### 3.2 DFT benchmarking

DFT benchmarking against CCSD(T)@cbs data was first performed on the two compounds already investigated, **C1** and **C2**. For each of the 50 dimer conformations selected in PICKY’s first cycle, the intermolecular energy was computed at CCSD(T)@cbs level and, thereafter, through a number of different combinations of DFT functionals, basis set and dispersion correction schemes. Among the tested combinations, the most significative have been selected and labeled CAM-B3LYP-D, B2PLYP-D, B97-D, PBE0-D and B3LYP-D (see Table **A**). As briefly discussed in the Supporting Information, all DFT-D calculations were performed with and without applying the CounterPoise<sup>81</sup> (CP) correction.

The results for both molecules are shown in the first two panels of Figure 3 and summarized in Table 1, where the quality of the DFT-D description is assessed based on the relative error with respect to reference intermolecular energies averaged over the 50 sampled dimers,

$$\langle \delta \rangle = \frac{1}{50} \sum_{k=1}^{50} (\Delta E_k^{DFT} - \Delta E_k^{CC})$$

and on the achieved standard deviation  $\sigma_{DFT}$  ( $\sigma_{DFT} = (\langle \delta^2 \rangle - \langle \delta \rangle^2)^{\frac{1}{2}}$ ). Since the CP corrected results are systematically better than the uncorrected ones in almost all cases, only the former have been reported in Table 1. For dichloromethane, the best descriptions are given by CAM-B3LYP, PBE0 and B3LYP dispersion corrected functionals. The first and the third slightly overestimate (*i.e.* less attractive interactions) the reference energies (by 0.08 and 0.02 kJ/mol, respectively), while PBE0-D underestimates CCSD(T)@cbs by almost the same extent (0.07 kJ/mol). On the contrary, for chloroform, CAM-B3LYP-

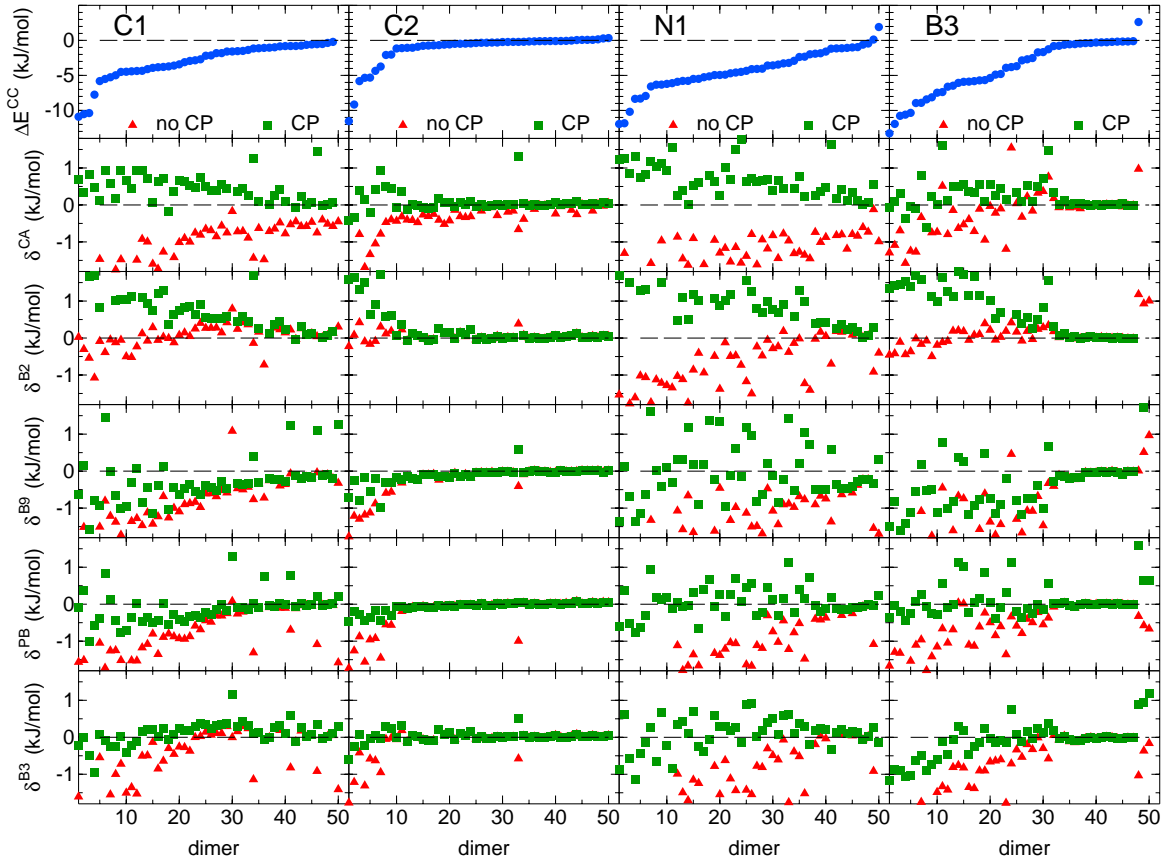


Figure 3: DFT benchmarking against reference CCSD(T)@cbs intermolecular energies ( $\Delta E^{CC}$ , blue circles in top panels) computed on the first 50 dimers extracted by PICKY. From left to right: **C1**, **C2**, **B3** and **N1**. For each molecule, the error achieved with the chosen functional ( $\delta^{XX} = \Delta E^{XX} - \Delta E^{CC}$ ; **XX** = **CAM-B3LYP-D**, **B2PLYP-D**, **B97-D**, **PBE0-D** and **B3LYP-D**) is shown in the first four panels with red triangles and green squares for CP uncorrected and corrected results, respectively.

D's  $\langle \delta \rangle$  and  $\sigma_{DFT}$  are significantly high (+0.59 and 0.84 kJ/mol), while the two best IPES descriptions are given by PBE0-D ( $\langle \delta \rangle = -0.11$  kJ/mol) and B3LYP-D ( $\langle \delta \rangle = +0.11$  kJ/mol) functionals. However, B3LYP-D shows less sparse data, yielding a standard deviation of 0.32 kJ/mol against 0.42 kJ/mol achieved with PBE0-D.

Since B3LYP-D appears to be the most accurate among the tested functionals, at least for the first two investigated halogens, two full parameterizations were carried out with the PICKY protocol described in Section 2, sampling the QM IPES by means of such functional. The resulting B3LYP-D based QMD-FFs, for **C1** and **C2** respectively, were thereafter employed in MD simulations. The final bulk densities achieved from these

Functional	<b>C1</b>		<b>C2</b>	
	$\langle \delta \rangle$ (kJ/mol)	$\sigma_{DFT}$ (kJ/mol)	$\langle \delta \rangle$ (kJ/mol)	$\sigma_{DFT}$ (kJ/mol)
CAM-B3LYP-D	+0.59	0.84	+0.08	0.28
B2PLYP-D	+1.03	1.34	+0.29	0.60
B97-D	-0.19	0.76	-0.12	0.26
PBE0-D	-0.11	0.42	-0.07	0.15
B3LYP-D	+0.11	0.32	+0.02	0.15

Functional	<b>N1</b>		<b>B3</b>	
	$\langle \delta \rangle$ (kJ/mol)	$\sigma_{DFT}$ (kJ/mol)	$\langle \delta \rangle$ (kJ/mol)	$\sigma_{DFT}$ (kJ/mol)
CAM-B3LYP-D	+0.82	1.05	+0.47	1.02
B2PLYP-D	+1.23	1.48	+1.02	1.39
B97-D	-0.04	0.73	-0.25	0.89
PBE0-D	+0.06	0.40	+0.11	0.44
B3LYP-D	+0.10	0.45	-0.11	0.49

Table 1: DFT benchmark against reference CCSD(T)@cbs data, computed for **C1** (top left), **C2** (top right), **N1** (bottom left) and **N1** (bottom right). Both  $\langle \delta \rangle$  and  $\sigma_{DFT}$  (see text for definition) refer to CP corrected energies. All functional labels refer to Table **A** in the Supporting Information.

simulation were compared to the ones obtained with the previous CAM-B3LYP-D based QMD-FFs. While details of these parameterizations are given in Tables **L** and **M** in the Supporting Information, in Table 2 shows the difference ( $\delta\rho$ ) between the bulk density achieved with the QMD-FFs and the experimental one.

molecule	$\langle \delta \rangle^{\text{CAM-B3LYP-D}}$ (kJ/mol)	$\langle \delta \rangle^{\text{B3LYP-D}}$ (kJ/mol)	$\delta\rho^{\text{CAM-B3LYP-D}}$ (%)	$\delta\rho^{\text{B3LYP-D}}$ (%)
<b>C1</b>	+0.59	+0.11	-17.4	-4.3
<b>C2</b>	+0.08	+0.02	-2.6	-0.7

Table 2: Columns 2 and 3: average error  $\langle \delta \rangle$  of the investigated functional with respect to reference CCSD(T)@cbs energies, computed over the 50 dimer sets extracted by PICKY. Columns 4 and 5: percentage error  $\delta\rho$  of the final MD density with respect to its experimental counterpart.  $\rho$  was obtained from 5 ns NPT simulations at 298 K and 1 atm, carried out with QMD-FFs parameterized on either CAM-B3LYP-D (fourth column) or B3LYP-D (last column) data.

As far as the **C2** compound is concerned, the aforementioned slight overestimation of the reference IPES (*i.e.* less attractive interaction energies) reflects, as could be expected, in a small underestimation of the condensed phase density. Despite the good performances



of both functionals, B3LYP-D shows a better agreement with the CCSD(T)@cbs data and it’s worth noticing that this in turn reflects in a better reproduction of the experimental liquid density. Turning to chloroform (**C1**), the surprisingly scarce performances of the CAM-B3LYP-D functional when compared to the reference energies ( $\langle \delta \rangle \simeq 0.6$  kJ/mol) lead to the already mentioned unacceptable error in the predicted density (more than 15% lower than the experiment). On the contrary, the simulations performed with the B3LYP-D parameterized QMD-FF yielded a significant improvement, being the final density underestimated by only  $\sim 4\%$ , again consistently with the overestimation of the reference interaction energy.

These data seem to indicate that the proposed protocol to evaluate DFT accuracy can give important *a priori* hints on the reliability of a given functional to lead to accurate predictions of bulk properties. To confirm this idea, the DFT benchmarking protocol was applied to two additional molecules, **B3** and **N1**: the former allows for extending the B3LYP-D based QMD-FFs to Bromine containing compounds, while the latter may serve to evaluate B3LYP-D accuracy in presence of cyano-groups. As for the former benchmark compounds, the results for these latter molecules are displayed in Figure 3 (last two columns) and their most important features summarized in Table 1.

As found for chloroform, the two best IPES descriptions for the brominated-alkane **B3** are given by PBE0-D and B3LYP-D functionals, but their behavior is reversed, because they respectively slightly overestimate (+0.11 kJ/mol,) and underestimate (-0.11 kJ/mol) the reference energies. Conversely, the CAM-B3LYP-D description again severely overestimates the energies with respect to the CCSD(T) values. Similar results were obtained for the trichloro acetonitrile (**N1**), where again both PBE0-D and B3LYP-D give the best results. It is worth recalling that both the average error and the standard deviation should be considered in the evaluation of each benchmarked functional. This becomes apparent, for instance, when handling the B97D-D functional: despite for **N1** it shows the best  $\langle \delta \rangle$ , it is evident from both Figure 3 and the computed  $\sigma_{DFT}$  that the small  $\langle \delta \rangle$  value comes from the compensation of severely underestimated and overestimated energies rather than from a consistent agreement. In conclusion, considering the more consistent results achieved on the first tested molecules, the B3LYP-D functional was

adopted in all PICKY parameterizations presented in this work.

### 3.3 QMD-FF results for halogenated hydrocarbons

Another aim of the present work is to further validate the protocol for QMD-FF parameterization here adopted and sketched in Figure 2. It is worth recalling that most of the QMD-FF approaches tend to favor accuracy over transferability. In other words, rather than attempting to find "universal" parameters or at least parameters shared by a very broad class of similar molecules (*e.g.* small organic liquids, bio-molecules, polymers, *etc.*), QMD-FFs can be exploited to account in some detail for chemical diversity, specifically tailoring FF parameters on each target molecule. For this reason, rather than evaluating the same set of parameters on a very broad class of benchmark molecules, one can focus on a relatively small number of molecules, here the one displayed in Figure 1, validating the procedure that leads to a specific parameter set for each individual molecule.

#### 3.3.1 Intramolecular Parameterization

As shown in the scheme displayed in Figure 2, the proposed QMD-FF parameterization starts for each molecule by building an intra-molecular FF through the JOYCE procedure. The JOYCE standard protocol (see References [22,33] and Supporting Information for additional details), was applied to all the molecules in the benchmark set displayed in Figure 1. An example of the achieved accuracy is shown for 2-iodo-propane (**I2**) in Figure 4. The comparison between QM and QMD-FF computed vibrational frequencies, displayed in the central panel, highlights the accuracy of the intramolecular FF description achieved for each normal mode. Nonetheless, it might be interesting to notice that once again<sup>33,82</sup> the best agreement with QM frequencies is achieved for high frequency (localized) and low frequency (affected by the torsional profiles) modes, while a somewhat smaller accuracy is obtained for intermediate frequencies, where coupling effects can play some role. However, by looking at Figure 4 top panel, the projection coefficients of each QMD-FF normal mode over its QM counterpart appear to indicate a rather reliable matching between the QM and MM descriptions. A very similar accuracy was achieved along the whole benchmark set, as confirmed by Table 3, which reports for each

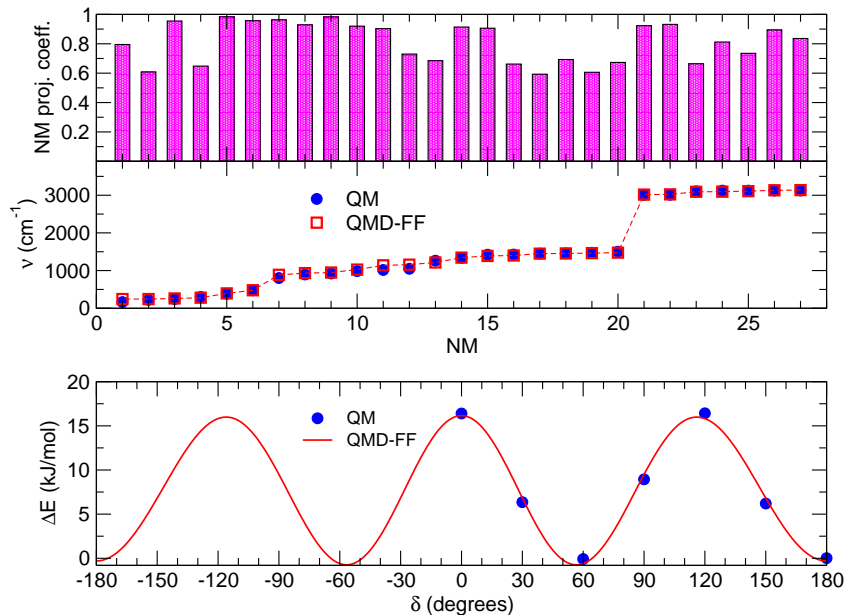


Figure 4: JOYCE parameterization results achieved for **I2** (2-iodo-propane). Top panel: overlap of QM and QMD-FF normal modes; middle panel: vibrational frequencies computed at QM level and by means of the JOYCE parameterized QMD-FF; bottom panel: QM and QMD-FF torsional energy profile along the principal dihedral angle.

compound  $\sigma_J$  (JOYCE standard deviation<sup>22,33</sup>) and  $\sigma_\nu$  (the standard deviation computed from the squared differences between the vibrational frequencies computed at QM level and the ones computed with the QMD-FF).

The final intramolecular parameters obtained for all molecules are reported in detail in Table 4 and in Tables **B** - **E**. By looking at Table 4, where the intramolecular harmonic parameters are reported for the first four investigated molecules, the QMD-FF capability to specifically account for chemical diversity, even within a class of rather similar molecules, clearly appears. For instance, the C-Cl bond, which takes place in the **C1-C4** compounds within different chemical surroundings, is described by JOYCE with different force constants  $k_s$  and equilibrium positions  $R_0$ , respectively spanning a range between 962 - 1624 kJ/mol  $\text{\AA}^{-2}$  and 1.75 and 1.83  $\text{\AA}$ . On the same foot, the Carbon - Halogen bond is consistently described along the series, being (see Supporting Information for details)  $\sim 1000$  kJ/mol  $\text{\AA}^{-2}$ , 800 kJ/mol  $\text{\AA}^{-2}$  and 500 kJ/mol  $\text{\AA}^{-2}$  the average

molecule	$\sigma_J$ (kJ/mol)	$\sigma_\nu$ (cm <sup>-1</sup> )
<b>C1</b>	0.016	55
<b>C2</b>	0.016	46
<b>C3</b>	0.011	59
<b>C4</b>	0.018	60
<b>N1</b>	0.005	66
<b>N2</b>	0.010	29
<b>B1</b>	0.020	40
<b>B2</b>	0.016	53
<b>B3</b>	0.036	55
<b>B4</b>	0.028	46
<b>I1</b>	0.025	42
<b>I2</b>	0.057	52

Table 3:  $\sigma_J$  and  $\sigma_\nu$  (see text for definitions) standard deviations obtained for each investigated molecules by JOYCE intramolecular parameterizations.

force constants found for Cl, Br and I, respectively, and  $\sim 1.80$  Å,  $2.00$  Å and  $2.20$  Å the equilibrium bond lengths.

### 3.3.2 Intermolecular Parameterization

Once all the intramolecular QMD-FFs were prepared, the next step consisted in obtaining the parameters for the intermolecular description. As previously detailed, this was done through the PICKY package, optimizing the QMD-FF dimer IPES against the reference one, sampled through B3LYP-D QM calculations, following the results of Section 3.2.

Table 5 contains the final values of selected quantities monitored during the parameterization cycles. By looking at the third and fourth column, it appears that all parameterizations successfully ended within 5 to 8 cycles, reaching the selected  $\Delta P$  threshold of 0.5 kJ/mol. The bulk densities ( $\rho_F^{MC}$ ) achieved in the last parameterization cycle through the MC simulations and reported in Table 5 cover a broad range of values, from 952 kg/m<sup>3</sup> (chloroform at 273K) to 2549 kg/m<sup>3</sup> (dibromomethane at 298K). This can be interpreted as a first indication that the QMD-FFs are able to maintain the capability of the parent QM description in accounting for chemical diversity. In particular, different densities were predicted by the QMD-FF also for compounds with a very similar chemical formula (*e.g.* **C1** and **C2** or **B1** and **B2**). Indeed it may be worth recalling that, as shown in Figure 2,

<b>C1</b>			<b>C2</b>		
Bond	$k_s$ (kJ/mol Å <sup>-2</sup> )	$R_0$ (Å)	Bond	$k_s$ (kJ/mol Å <sup>-2</sup> )	$R_0$ (Å)
C-H	3288.76	1.09	C-H	3261.98	1.10
C-Cl	961.87	1.79	C-Cl	1181.66	1.80
Angle	$k_b$ (kJ/mol rad <sup>-2</sup> )	$\theta_0$ (degr)	Angle	$k_b$ (kJ/mol rad <sup>-2</sup> )	$\theta_0$ (degr)
H-C-Cl	365.29	108	H-C-H	310.14	112
Cl-C-Cl	820.48	111	H-C-Cl	367.81	113
			Cl-C-Cl	1006.94	113

<b>C3</b>			<b>C4</b>		
Bond	$k_s$ (kJ/mol Å <sup>-2</sup> )	$R_0$ (Å)	Bond	$k_s$ (kJ/mol Å <sup>-2</sup> )	$R_0$ (Å)
C1-H	3418.30	1.10	C1-H1	3139.17	1.10
C1-C2	5063.38	1.33	C1-C2	2340.74	1.52
C2-Cl	1623.78	1.75	C2-H2	3181.66	1.11
			C2-Cl	1140.61	1.83
Angle	$k_b$ (kJ/mol rad <sup>-2</sup> )	$\theta_0$ (degr)	Angle	$k_b$ (kJ/mol rad <sup>-2</sup> )	$\theta_0$ (degr)
H-C1-H	885.93	114	H1-C1-H1	318.76	108
H-C1-C2	301.95	120	H1-C1-C2	374.33	111
C1-C2-Cl	248.23	123	C1-C2-H2	395.09	112
Cl-C2-Cl	243.64	120	C1-C2-Cl	707.99	111
Dihed.	$k_s$ (kJ/mol rad <sup>-2</sup> )	$\phi_0$ (degr)			
H-C1-C2-Cl	85.13	0			

Table 4: Harmonic intramolecular parameters obtained through the JOYCE procedure for chloroform (**C1**), dichloro-methane (**C2**), 1,1-dichloro-ethene (**C3**) and chloro-ethane (**C4**). The labels for force constants and equilibrium position correspond to the quantities entering equations (S4)-(S7) in the Supporting Information.

the only information needed to build a QMD-FF through the proposed parameterization route for a given target molecule is its chemical formula. Since the individual chemical specificity is accounted for at QM level, the adherence to the reference QM description of the final QMD-FF IPES is a key quantity to initially evaluate the FF quality. Here, this first evaluation is achieved through the PICKY standard deviation  $\sigma_P$ , computed in the last cycle over the **functional objective function**  $I^{inter}$  of equation (2) and reported in the second column of Table 5.

Different performances were registered for the investigated molecules, with the best and the worst agreement reached for compound **C4** (0.83 kJ/mol) and **B2** (2.26 kJ/mol). The not negligible difference between these limit values suggests that the employed stan-

Molecule	$\sigma_P$ (kJ/mol)	$N_P$	$\Delta P$ (kJ/mol)	T(K)	$\rho_F^{MC}$ (kg/m <sup>3</sup> )
<b>C1</b>	1.25	6	0.14	298	1408
<b>C2</b>	1.04	7	0.11	298	1310
<b>C3</b>	1.21	7	0.36	298	1248
<b>C4</b>	0.83	6	0.14	273	952
<b>N1</b>	1.37	6	0.28	298	1426
<b>N2</b>	1.27	8	0.40	298	1356
<b>B1</b>	1.18	5	0.49	298	1782
<b>B2</b>	2.40	7	0.48	298	2549
<b>B3</b>	2.26	7	0.41	298	2282
<b>B4</b>	1.20	5	0.39	298	1415
<b>I1</b>	1.60	8	0.10	298	2074
<b>I2</b>	1.40	7	0.17	298	1777

Table 5: Computed densities (last column) achieved during PICKY parameterization in MC runs at 1 atm and temperature T (fifth column). In the second, third and fourth column are reported the final fitting standard deviation ( $\sigma_P$ ), the number of cycles ( $N_P$ ) and the final convergence index ( $\Delta P$ ).

standard FF model functions (*i.e.* LJ and charge-charge potentials only) might not always be sufficiently flexible to accurately mimic the IPES. Although there is no clear systematic distribution of the error over the investigated targets, it may be worth to outline some general trends that can be observed along the investigated set. First, the Chlorine containing compounds, **C1-C4** and **N1-N2**, showed the best agreement ( $\sim 1.1$  kJ/mol and  $\sim 1.3$  kJ/mol in average, respectively) with the QM reference data. About the same accuracy ( $\sim 1.2$  kJ/mol) was found for those Bromo-hydrocarbons, bearing only one Bromine atom. Conversely, the largest deviations ( $\sim 2.3$  kJ/mol) appeared for the two di-bromo alkanes, suggesting that more complex functions should be adopted in this case to improve the adherence to the QM IPES. Finally, **I1** and **I2** show an intermediate error ( $\sim 1.50$  kJ/mol) between the aforementioned two groups. It could be hypothesized that these errors are connected to the limited capability of the adopted model functions to account for halogen bond. On this basis, *ad hoc* modifications of the standard FF functions (*e.g.* the addition of auxiliary interacting sites like ESHs), could possibly solve the problem. *Alternatively, the inclusion of many-body effects through, for instance, polarization functions or even more complex algorithms to compute the FF could also be tempted, at least in the case of the rather small molecules here considered.* Nonetheless, the inclusion of

ESHs *and/or three-body contributions* in a QMD-FF protocol is beyond the aims of this work, and it will be the object of future investigations.

A deeper insight can be gained by monitoring the behavior of the same key quantities along the parameterization cycles, as displayed in Figure 5. In the bottom row the ratio between the bulk density achieved in the MC simulation of the monitored cycle and the final one ( $\rho_F$ ) is displayed for all compounds. In almost all cases, convergence

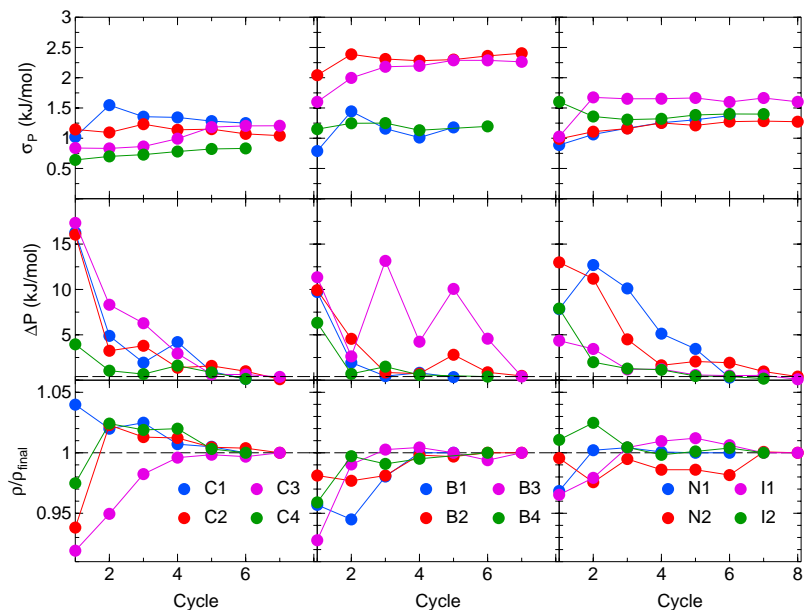


Figure 5: Selected properties monitored along the cycles of PICKY parameterization, achieved for compounds **C1-C4** (left panels), **B1-B4** (center panels), **N1-N2** and **I1-I2** (right panels). From top to bottom: PICKY standard deviation ( $\sigma_P$ ), convergence index ( $\Delta_P$ ) and the ratio between the MC density at each cycle and the final MC density. In the central panels, the  $\Delta_P$  threshold (0.5 kJ/mol) is displayed with a black dashed line.

on density is reached after few cycles, being the changes after the fifth cycle almost negligible. Conversely, as shown in the middle rows, the shape of the QMD-FF IPES converges more slowly, as it often requires two or more additional cycles to fall under the selected threshold. However, the largest  $\Delta_P$  variations are made in most cases within the first cycles, confirming the efficiency of the PICKY sampling algorithm to extract representative samples even with a relatively small number of dimers.

The final intermolecular parameters achieved through PICKY parameterization are

reported for compounds **C1-C4** in Table 6, and in Tables **I-K**, for the rest of the benchmarked molecules. Here, the improved capability of the QMD-FF, with respect to

Molecule	prm	C <sub>1</sub>	C <sub>2</sub>	H <sub>1</sub>	H <sub>2</sub>	Cl
<b>C1</b>	$\sigma$	3.44	-	2.00	-	3.61
	$\epsilon$	2.255	-	0.048	-	0.647
	$q$	-0.236	-	0.259	-	-0.008
<b>C2</b>	$\sigma$	3.09	-	2.85	-	3.53
	$\epsilon$	2.114	-	0.005	-	1.020
	$q$	-0.405	-	0.254	-	-0.052
<b>C3</b>	$\sigma$	3.13	3.66	2.91	-	3.42
	$\epsilon$	1.618	0.105	0.005	-	1.306
	$q$	-0.028	-0.221	0.098	-	0.026
<b>C4</b>	$\sigma$	3.55	3.25	2.83	2.77	3.52
	$\epsilon$	0.598	1.659	0.005	0.0050	1.106
	$q$	0.005	-0.011	0.021	0.077	-0.209

Table 6: PICKY parameters for all investigated chloro-hydrocarbons:  $\sigma$  (Å),  $\epsilon$  (kJ/mol) and  $q$  (a.u.).

transferable FFs, to preserve chemical identity even at atomic level emerges even more clearly than for the intramolecular parameters. To not overwhelm the reader with undue information, the comparison with literature parameter sets will be discussed only with respect to OPLS<sup>11</sup> parameters, since it covers the largest number of the molecules here investigated, while GAFF and CGenFF data were available<sup>17,59</sup> only for few compounds. For instance, the QMD-FF LJ’s  $\sigma$  values for the carbon atoms reported in Table 6, show very different values, going from 3.09 Å, found for dichloromethane to 3.55 Å, obtained for the methyl atom in chloroethane. This range is even more extended (3.68 Å, for molecule **I1**) if one takes into account all the investigated compounds. On the contrary, empirical FFs (see for instance OPLS, in Tables **F-H**), to preserve transferability, tend to adopt the same  $\sigma$  value for all (alkane) carbons, with a consequent loss of precision in the IPES. On the same foot, the strength of the LJ interaction for carbon atoms is constrained to 0.28 kJ/mol, whereas in QMD-FF,  $\epsilon$  may vary from a minimum of 0.1 kJ/mol to a maximum value of 2.3 kJ/mol, only considering the small **C1-C4** set. Similar conclusions can be drawn by extending such comparative analysis to all compounds and all intermolecular parameters, reported in detail in the Supporting Information.



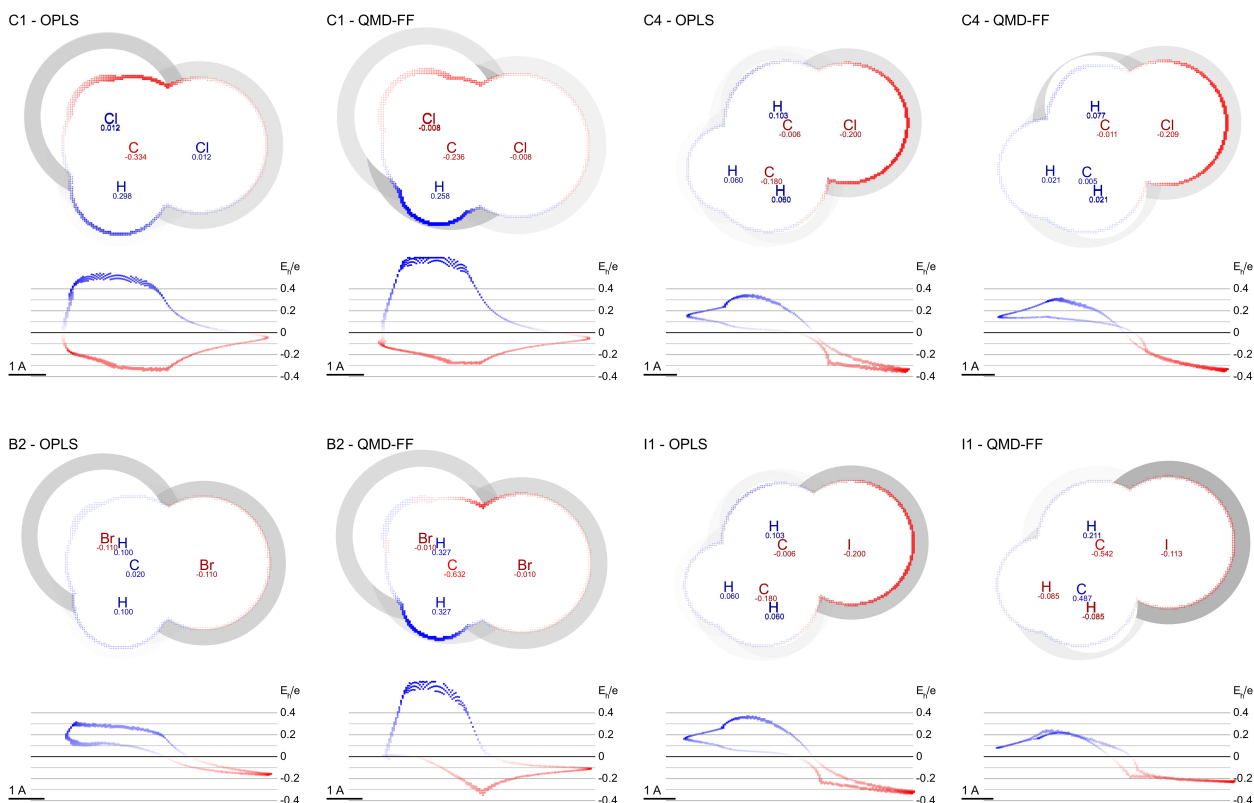


Figure 6: Visual 2D comparison (minimap) of the OPLS<sup>11</sup> and QMD-FF descriptions of molecules **C1**, **C4**, **B2**, **I1**. Each molecule is displayed on the molecular plane containing the X-C-C or X-C-H triplet (X = halogen). For each atom, the assigned charge is reported; LJ parameters are displayed through white (repulsive) circles with radius =  $\sigma$ , and grey (attractive) rings, whose opacity is proportional to  $\epsilon$ . The electrostatic potential (EPS) is plotted along the intersection between the molecular plane and the VdW spheres of radii  $\sigma$ , with colors varying from dark blue (0.7 a.u.) to dark red (-0.4 a.u.). The same values of ESP are also plotted, projected onto the  $x$  axis, below each minimap.

Some of the parameterized QMD-FFs are visualized in Figure 6 (2D diagrams), along with their OPLS counterparts. As concerns the treatment of the halogen atoms, **B2** and **I1** provide interesting insights about the aforementioned  $\sigma$ -hole problem (which is known to be more evident in the heaviest halogens). By looking at the Bromine atom in **B2**, it may be worth noticing that the charge assigned by QMD-FF is close to zero (-0.010 a.u.), while it is more negative in OPLS (-0.110 a.u.). Indeed, the electrostatic potential (EPS, red and blue curves), which is expected to be negative on the nucleophilic side of the atom's VdW surface, is now determined by the strongly negative charge assigned to

the carbon atom: as seen from above the C-Br bond, the EPS around the halogen goes from -0.3 a.u. near the center of the bond to -0.1 on the polar site of the  $\sigma$ -hole, correctly reproducing the electrostatic imbalance between the VdW sphere’s side and pole. The OPLS description shows an inverted trend, with a more negative EPS on top of the Br atom. In the case of **I1**, on the other hand, both the OPLS and QMD-FF descriptions yield a more negative potential in the  $\sigma$ -hole with respect to the surrounding region, however the strongly negative value of  $\simeq$  -0.3 a.u. for OPLS is reduced to  $\simeq$  0.2 a.u. by QMD-FF, and its value is almost constant throughout all the Iodine’s surface, as compared to a steeper slope in OPLS. From these examples, it is evident that the information about the  $\sigma$ -hole is present in the IPES sampled at the QM level, and that the QMD-FF parameterization route is capable of getting close to the expected interaction profile, by adjusting simultaneously the parameters of the halogen and its neighboring atoms. It is also clear, however, that the inclusion of an *ad hoc* center, like a ESH, would be essential for a proper treatment of Iodine and Bromine’s anisotropy.

### 3.3.3 Validation against experiment

Since the ultimate goal of any FF is to deliver accurate prediction of the macroscopic properties, the performances in MD simulations of the twelve QMD-FFs, obtained by joining the parameterized intra- and inter-molecular parts, were eventually tested against available experimental data, as well as compared to those achieved with popular transferable FFs, reported in References [17,59]. For all compounds, extended MD simulations were performed in the NPT ensemble, on systems of one thousand molecules, at 1 atm and at several temperatures. Unless otherwise stated, to make consistent comparisons with literature results, all MD parameters were set as suggested by Caleman and coworkers.<sup>17</sup> Further details concerning the MD runs can be found in the Supporting Information.

The final values of the computed thermodynamic properties for compounds **C1-I2** are reported in Tables 7-11. By looking at the last row of Table 7, where the standard deviation of the computed bulk densities with respect to the experiment is displayed, it appears as the overall performance of the QMD-FF (98 kg/m<sup>3</sup>) is comparable to the one delivered by OPLS (94 kg/m<sup>3</sup>) and remarkably better than the one given by GAFF (254

kg/m<sup>3</sup>). Within the chosen set of benchmark molecules, CGenFF yields instead a better standard deviation (47 kg/m<sup>3</sup>), but it is worth recalling that many of these compounds were included in the CGenFF’s<sup>21</sup> training set, and their parameters tuned to reproduce experimental density and vaporization enthalpy. More specifically, the QMD-FF yields the bulk densities closest to the experiment for **C1**, **C2**, **N1**, **N2** and **I2**, while the percentage error oscillates from a minimum value of less than 1% (**C2**) to a maximum one of  $\sim 9$  % (**I1**). As already pointed out, also the performances of the transferable FFs appear to depend on the chosen target: for instance GAFF outperforms OPLS for **C2**, while the opposite is true for **I2**. The good agreement with the experiment found for QMD-FF and its performance versus transferable FFs is more than encouraging, especially considering both the *a priori* origin of the former and the fact the the experimental bulk density is one of the target properties on which transferable FFs are tuned.

Molecule	$\rho$ (kg/m <sup>3</sup> )				
	GAFF <sup>17</sup>	OPLS <sup>17</sup>	CGenFF <sup>59</sup>	QMD-FF	Exp.
<b>C1</b>	1375	1373	-	1420	1479 <sup>80</sup> - 1480 <sup>83</sup>
<b>C2</b>	1258	1200	1241	1328	1316 <sup>80</sup> - 1327 <sup>79</sup>
<b>C3</b>	1054	1176	1217	1261	1213 <sup>79</sup>
<b>C4</b> (273 K)	898	915	912	969	924 <sup>80</sup>
<b>N1</b>	-	1410	-	1439	1440 <sup>79</sup>
<b>N2</b>	-	1265	-	1378	1369 <sup>79</sup>
<b>B1</b>	1326	1740	-	1813	1662 <sup>83</sup> - 1675 <sup>79</sup>
<b>B2</b>	1946	2474	2402	2570	2469 <sup>79</sup> - 2484 <sup>80</sup>
<b>B3</b>	1807	2335	2170	2311	2169 <sup>83</sup> - 2179 <sup>79</sup>
<b>B4</b>	1229	1408	1398	1446	1345 <sup>80</sup> - 1354 <sup>79</sup>
<b>I1</b>	-	2042	-	2107	1936 <sup>79</sup>
<b>I2</b>	1599	1835	-	1818	1695 <sup>83</sup> - 1737 <sup>80</sup>
St. Dev. (kg/m <sup>3</sup> )	254	94	47	98	-

Table 7: Bulk density ( $\rho$ , kg/m<sup>3</sup>): comparison between experiment and data computed from MD simulations performed with popular FFs and the QMD-FF. Unless otherwise stated, all values refer to T = 298 K.

The other property usually employed in empirical FF tuning is the vaporization enthalpy. In Table 8 the  $\Delta H^{vap}$  computed and experimental values are compared for each compound. By looking at the standard deviations, reported as usual in the last row, it

Molecule	$\Delta H^{vap}$ (kJ/mol)				
	GAFF <sup>17</sup>	OPLS <sup>17</sup>	CGenFF <sup>59</sup>	QMD-FF	Exp.
<b>C1</b>	28.4	29.2	-	30.4	31.3 <sup>84</sup>
<b>C2</b>	26.5	23.4	25.3	31.0	28.8 <sup>79</sup> -29.9 <sup>85</sup>
<b>C3</b>	24.9	24.3	29.9	27.7	26.5 <sup>79</sup>
<b>C4(273 K)</b>	22.9	26.9	24.0	28.9	27.8 <sup>86</sup>
<b>N1</b>	-	35.8	-	42.3	41.7 <sup>70</sup>
<b>N2</b>	-	35.5	-	34.5	33.0 <sup>70</sup>
<b>B1</b>	18.5	25.2	-	25.5	22.8 <sup>79</sup>
<b>B2</b>	28.6	34.2	43.1	40.3	37.4 <sup>84</sup>
<b>B3</b>	34.7	48.8	40.1	51.3	41.7 <sup>84</sup>
<b>B4</b>	31.4	34.8	35.3	39.4	32.2 <sup>79</sup>
<b>I1</b>	-	29.0	-	39.8	32.0 <sup>79</sup>
<b>I2</b>	33.7	33.8	-	40.6	34.1 <sup>79</sup>
St. Dev. (kJ/mol)	4.6	3.7	3.7	4.8	-

Table 8: Vaporization enthalpy ( $\Delta H^{vap}$ , kJ/mol): comparison between experiment and data computed from MD simulations performed with popular FFs and the QMD-FF. Unless otherwise stated, all values refer to T = 298 K.

appears as the QMD-FF error is slightly larger than the one computed with GAFF and  $\sim 1$  kJ/mol worse than OPLS and CGenFF. Consistently with  $\rho$ , where most values were overestimated with respect of the experiment, there is a general overestimation of the total energy of the system given by the QMD-FF. In the previous parameterizations performed on pyridine<sup>32</sup> and benzene<sup>40</sup> molecules, this issue was traced back to the absence of three body effects, given the pure two-body nature of the QMD-FF. As a matter of fact, the *effect of the inclusion of three-body interactions, in the FFs has been recently reported to in the FF, which has been recently*<sup>44</sup> *quantified to account for 15-20% of the total interaction energy of water bulk phase (hence causing not negligible inaccuracies), was found to contribute only for a 2-5%, when organic molecules, more similar to the ones here considered, are taken into account.*<sup>36</sup> *In turn, consistently with the present results, the absence of three-body effects was reported*<sup>36</sup> *to cause a  $\sim 5\%$  overestimation of the bulk density and to lower the  $\Delta H^{vap}$  by 6 to 12%. Here, these inaccuracies* could be also connected to a only partial reproduction of the halogen bond. Indeed, the worst QMD-FF performances are registered essentially for Bromine (**B1**, **B3** and **B4**)

and Iodine (**I2**) containing compound, where the halogen bond is expected to play a more significant role. This notwithstanding, the QMD-FFs are capable of yielding the most accurate predictions on seven over the twelve compounds, reaching a high precision for most of the chlorinated targets.

A more fair playground for QMD and transferable FFs can be set by comparing the performances delivered by the two approaches in the prediction of thermodynamic properties outside the set of target observables employed in the empirical FF tuning. To this end, specific heat at constant pressure ( $c_P$ ), volumetric expansion coefficient ( $\alpha_P$ ) and isothermal compressibility ( $\kappa_T$ ) were computed for all compounds and shown in Tables 9, 10 and 11, respectively, together with the ones available in the literature<sup>17,59</sup> for transferable FFs. As far as  $c_P$  is concerned, despite a systematic underestimation of the experimental values is evident for every compound, in almost all cases when the data are available for comparison, QMD-FFs pair or outperform the standard FFs, resulting in a standard deviation with respect to the experiment of  $21 \text{ J K}^{-1} \text{ mol}^{-1}$ , against  $42 \text{ J K}^{-1} \text{ mol}^{-1}$  and  $26 \text{ J K}^{-1} \text{ mol}^{-1}$ , obtained for GAFF and OPLS, respectively.

Molecule	$c_P$ ( $\text{J K}^{-1} \text{ mol}^{-1}$ )				
	GAFF <sup>17</sup>	OPLS <sup>17</sup>	CGenFF <sup>59</sup>	QMD-FF	Exp.
<b>C1</b>	88	113	-	93	114 <sup>79</sup>
<b>C2</b>	69	75	-	77	101 <sup>79</sup>
<b>C3</b>	-	-	-	82	111 <sup>79</sup>
<b>C4</b>	140	91	-	92	104 <sup>79</sup>
<b>N1</b>	-	104	-	107	-
<b>N2</b>	-	123	-	122	-
<b>B1</b>	-	-	-	66	-
<b>B2</b>	-	-	-	83	-
<b>B3</b>	73	93	-	118	136 <sup>79</sup>
<b>B4</b>	-	-	-	127	-
<b>I1</b>	-	-	-	98	115 <sup>79</sup>
<b>I2</b>	-	-	-	129	-
St. Dev. (kJ/mol)	42	26	-	21	-

Table 9: Specific heat at constant pressure ( $c_P$ ,  $\text{J K}^{-1} \text{ mol}^{-1}$ ): comparison between experiment and data computed from MD simulations performed with popular FFs and the QMD-FF.

Besides employing  $\alpha_P$  in the FF validation, the volumetric expansion coefficient was

also exploited to evaluate the internal consistency of the MD runs. As detailed in equations (S13) and (S14) in the Supporting Information,  $\alpha_P$  was computed in two different manners, either from the fluctuations of the MD run at 298 K (273 K for **C3**) or from the variation of the density evaluated over three runs, purposely carried out at different temperatures. From the values reported in the fifth column of Table 10, it appears that the two estimates are in fairly good agreement, which ranges from 0.03 (found for **B1**, **B3** and **I1**) to  $0.60 \cdot 10^{-3} \text{ K}^{-1}$  (**N1**). Turning to the comparison with the experimental values, again the QMD-FF yields the best results with respect to the other investigated sets

Molecule	$\alpha_P$ ( $10^3 \text{ K}^{-1}$ )				
	GAFF <sup>17</sup>	OPLS <sup>17</sup>	CGenFF <sup>59</sup>	QMD-FF	Exp.
<b>C1</b>	1.91	2.28	-	1.29 (1.06)	1.30 <sup>84</sup>
<b>C2</b>	1.73	2.31	1.71	1.31 (1.52)	1.35 <sup>84</sup> -1.44 <sup>83</sup>
<b>C3</b>	1.79	2.07	1.47	1.47 (1.65)	1.46 <sup>83</sup>
<b>C4</b>	2.90	2.23	1.77	1.43 (1.23)	1.30 <sup>83</sup>
<b>N1</b>	-	1.3	-	1.20 (1.81)	-
<b>N2</b>	-	1.2	-	1.10 (1.30)	-
<b>B1</b>	2.86	1.79	-	1.47 (1.50)	1.67 <sup>83</sup>
<b>B2</b>	1.48	1.44	1.01	0.98 (1.07)	1.02 <sup>84</sup>
<b>B3</b>	1.42	1.06	1.10	0.84 (0.81)	0.92 <sup>84</sup>
<b>B4</b>	1.56	1.28	1.29	1.07 (1.13)	1.23 <sup>79</sup>
<b>I1</b>	-	-	-	0.97 (0.80)	1.17 <sup>79</sup>
<b>I2</b>	0.91	1.13	-	1.03 (1.11)	1.15 <sup>83</sup>
St. Dev. ( $10^3 \text{ K}^{-1}$ )	0.76	0.61	0.25	0.12	-

Table 10: Volumetric expansion coefficient ( $\alpha$ ,  $10^3 \text{ K}^{-1}$ ): comparison between experiment and data computed from MD simulations performed with popular FFs and the QMD-FF. In the fourth column, the number between parenthesis is the  $\alpha$  value computed from the volume and enthalpy fluctuations.

of parameter, remarkably increasing the quality of the prediction. Similar performances were encountered when dealing with  $\kappa_T$ , as reported in Table 11. In fact, the QMD-FF yield data in better agreement with the experiment with respect to the transferable FFs, being  $0.12 \text{ GPa}^{-1}$  the QMD-FF standard deviation, and  $0.19 \text{ GPa}^{-1}$  the maximum error, found for compound **C4**. On the same compound, GAFF, OPLS and CGenFF give an error of 2.86, 1.11 and  $0.42 \text{ GPa}^{-1}$ , respectively, whereas their standard deviation is 1.29,

Molecule	$\kappa_T$ (GPa <sup>-1</sup> )				
	GAFF <sup>17</sup>	OPLS <sup>17</sup>	CGenFF <sup>59</sup>	QMD-FF	Exp.
<b>C1</b>	1.79	2.16	-	1.09	1.03 <sup>84</sup>
<b>C2</b>	1.47	2.22	1.42	0.95	1.03 <sup>79</sup>
<b>C3</b>	1.64	1.68	1.22	1.03	-
<b>C4</b>	3.88	2.13	1.44	1.21	1.02 <sup>79</sup>
<b>N1</b>	-	-	-	1.55	-
<b>N2</b>	-	-	-	0.80	-
<b>B1</b>	4.15	1.36	-	0.91	-
<b>B2</b>	1.45	0.91	0.56	0.58	0.70 <sup>84</sup>
<b>B3</b>	1.17	0.54	0.70	0.70	0.64 <sup>84</sup>
<b>B4</b>	1.47	0.99	0.96	1.29	1.13 <sup>17</sup>
<b>I1</b>	-	-	-	0.58	-
<b>I2</b>	1.26	1.01	-	0.74	-
St. Dev. (GPa <sup>-1</sup> )	1.29	0.82	0.28	0.12	-

Table 11: Isothermal compressibility ( $\kappa_T$ , GPa<sup>-1</sup>): comparison between experiment and data computed from MD simulations performed with popular FFs and the QMD-FF.

0.82 and 0.28 GPa<sup>-1</sup>

Molecule	$\eta$ (10 <sup>-9</sup> m <sup>-1</sup> s <sup>-1</sup> )	
	QMD-FF	Exp.
<b>C1</b>	0.34	0.54 <sup>87</sup>
<b>C2</b>	0.42	0.41 <sup>79</sup>
<b>C3</b>	0.32	0.33 <sup>88</sup>
<b>C4(273K)</b>	0.32	0.28 <sup>89</sup>
<b>N1</b>	0.44	-
<b>N2</b>	0.86	-
<b>B1</b>	0.45	0.32 <sup>90</sup>
<b>B2</b>	0.79	0.98 <sup>79</sup>
<b>B3</b>	1.32	1.59 <sup>79</sup>
<b>B4</b>	0.67	0.53 <sup>91</sup>
<b>I1</b>	0.86	0.58 <sup>92</sup>
<b>I2</b>	0.77	0.70 <sup>92</sup>
St. Dev. 10 <sup>-9</sup> m <sup>-1</sup> s <sup>-1</sup> ( )	0.20	-

Table 12: Shear viscosity ( $\eta$ , 10<sup>-9</sup> m<sup>-1</sup> s<sup>-1</sup>): comparison between experiment and data computed from MD simulations performed with popular FFs and the QMD-FF. Unless otherwise stated, all values refer to T = 298 K.

Finally, the shear viscosity has also been computed, to assess the performances of the QMD-FF on a collective, dynamic property. Results are reported in Table 12, and compared with their experimental counterparts. No data regarding the behavior of transferable FFs in delivering viscosity estimates for these compounds were instead found. The overall standard deviation of  $0.20 \cdot 10^{-9} \text{ m}^{-1} \text{ s}^{-1}$  originates from very different performances, which remarkably depend on the investigated compound, and the maximum error ranges from  $0.01 \cdot 10^{-9} \text{ m}^{-1} \text{ s}^{-1}$ , found for **C2**, to  $0.28 \cdot 10^{-9} \text{ m}^{-1} \text{ s}^{-1}$  for **I1**.

An overview of the comparison between the QMD-FF predictions and the experimental properties for all considered observables reveals some trends in the distribution of the error, that can be connected with the accuracy of the reproduction of the QM IPES given by the QMD-FF. Indeed, as emerges from Table 5, the compounds showing the worst agreement with the QM reference IPES are those potentially more subjected to establish strong halogen bonds, as **B3** ( $\sigma_P = 2.26 \text{ kJ/mol}$ ), **B2** ( $2.26 \text{ kJ/mol}$ ), **I1** ( $1.60 \text{ kJ/mol}$ ) and, to a minor extent, **I2** ( $1.40 \text{ kJ/mol}$ ). Interestingly, these compounds are also among those giving the larger error in the prediction of either density (171, 142, 123 and 101  $\text{kg/m}^3$  for **I1**, **B3**, **I2** and **B2**, respectively), vaporization enthalpy (9.6, 7.2, 6.5, and 2.8  $\text{kJ/mol}$  for **B3**, **I1**, **I2** and **B2**) and viscosity (0.07, 0.28, 0.27,  $0.19 \cdot 10^{-9} \text{ m}^{-1} \text{ s}^{-1}$  for **I1**, **B3**, **B2** and **I2**). On the one side this finding further confirms the close connection between the adherence to the QM IPES and the quality of the QMD-FF in reproducing macroscopic properties. On the other side, the implementation of non-standard potential functions, to better mimic the presence of halogen bonds, could in principle increase the quality of QMD-FF prediction, and certainly deserves further investigation.

## 4 Conclusions

The first of the aims of the present work was to validate the JOYCE/PICKY protocol, by comparing the QMD-FF predictions in terms of macroscopic properties with the available experimental and computed data. In this regard, the overall performance of the QMD-FF over the explored compounds was more than encouraging. In fact, the QMD-FF computed values were not only in rather good agreement with the experimental ones, but in many cases also more accurate than the considered literature computational estimates,



even for those properties, like density and vaporization enthalpy, that transferable FFs were empirically tuned to reproduce. In second place, a connection was found between the ability of a given QMD-FF to yield reliable predictions of macroscopic properties and how precisely it represents the QM IPES. In fact, the somewhat lesser accuracy in the QM IPES description, achieved for those compounds potentially involved in strong halogen bonds, was shown to reflect in the largest errors in terms of density, vaporization enthalpy and viscosity. This suggests that even more reliable predictions of macroscopic properties could be obtained by increasing the quality of the IPES description through the implementation of more complex and flexible potentials (as for instance ESH or polarizable charges) in the analytical functions of the QMD-FF. Finally, the third goal of this study was to address the capability of the proposed parameterization procedure to incorporate the chemical detail, intrinsic in the parent QM description, into the more simple FF model. The subtle differences between apparently similar compounds (as the **C1/C2**, **N1/N2** and **B1/B2** pairs) were found to be accounted for by the QMD-FF, both at intra- an inter-molecular level. Indeed, the variety of bonded and non bonded parameters further testifies the higher specificity of the QMD-FF, remarkably increased with respect to its transferable counterparts. On the same foot, the reliable description of the halogen bond features, intrinsically contained in the QM description, was shown to be partially transferred to the QMD-FF, albeit limited by the simplicity of the standard potential function adopted.

From a more general point of view, this work can be considered a further step of a long-lasting project, started in our group several years ago, aimed to the development of a sound and accurate QMD-FF parameterization procedure, for computer simulations of advanced materials and complex systems. Based on the findings of the present study, three complementary tasks should be pursued in forthcoming work. First, given the good results achieved for the simple molecules investigated in this work, the protocol should be extended to larger molecules, implementing novel solutions to diminish the consequent computational burden (in particular in the QM level), without a significant loss of accuracy. The second task is connected to the former, and concerns the choice of the QM level adopted to sample the dimer IPES. Despite computational feasibility

suggests to resort to DFT based techniques, the proposed (and necessary) benchmarking against CCSD(T) data becomes rapidly infeasible with the growing molecular dimensions. The search and development of new dispersion corrections and/or QM methods able to balance accuracy and computational cost is in this context of the foremost importance. Finally, as mentioned throughout the result discussion, the need of more flexible potential FF functions clearly emerges from this study. In the case of halogenated hydrocarbons, the implementation of an ESH model, whose parameters could be inserted in the QMD-FF and optimized together with the standard LJ and charges, could certainly improve the final results. More in general, the adoption of more complex functions, especially those describing the repulsive branches of the interaction curves, can also be a key strategy to dramatically improve the future QMD-FF performances. *In this framework, further research could involve the implementation also of many body terms, beyond the pure two-body approximation here adopted. On the one hand, this could allow for a more sound representation of the interactions that take place in the bulk, but on the other hand the computational effort to account for such terms during simulations could seriously undermine the possibility to extend the approach to large molecules and complex systems.*

### **Supporting Information**

Additional data and several details about the reported calculations not included in this paper.

## References

- [1] Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon: Oxford, 1987.
- [2] Frenkel, D.; Smith, B. *Understanding Molecular Simulations*; Academic Press: San Diego, 1996.
- [3] Ferrario, M.; Ciccotti, G.; Binder, K. *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 2; Lecture Notes in Physics, Vol. 704*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2006.
- [4] Mackerell, A. D. Empirical force fields for biological macromolecules: Overview and issues *J. Comp. Chem.* **2004**, *25*, 1584–1604.
- [5] Naidoo, K.; Brady, J.; Field, M.; Gao, J.; Hann, M. *Modelling Molecular Structure and Reactivity in Biological Systems; Special Publications; The Royal Society of Chemistry*, 2006.
- [6] Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
- [7] 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**, *7863*, 11225–11236.
- [8] Sun, H. COMPASS: An ab Initio Force-Field Optimized for Condensed-Phase Applications Overview with Details on Alkane and Benzene Compounds *J. Phys. Chem. B* **1998**, *102*, 7338–7364.
- [9] MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.;

- Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–616.
- [10] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. a.; Case, D. a. Development and Testing of a General Amber Force Field. *J. Comp. Chem.* **2004**, *25*, 1157–74.
- [11] Jorgensen, W. L.; Tirado-Rives, J. Potential Energy Functions for Atomic-Level Simulations of Water and Organic and Biomolecular Systems. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6665–70.
- [12] Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholtz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. The GROMOS software for biomolecular simulation: GROMOS05. *J. Comp. Chem.* **2005**, *26*, 1719–51.
- [13] Tiberio, G.; Muccioli, L.; Berardi, R.; Zannoni, C. Toward in Silico Liquid Crystals. Realistic Transition Temperatures and Physical Properties for n-Cyano-biphenyls via Molecular Dynamics Simulations *ChemPhysChem*. **2009**, *10*, 125.
- [14] Burger, S. K.; Ayers, P. W.; Schofield, J. Efficient Parameterization of Torsional Terms for Force Fields. *J. Comp. Chem.* **2014**, *35*, 1438–45.
- [15] Cole, D. J.; Vilseck, J. Z.; Tirado-Rives, J.; Payne, M. C.; Jorgensen, W. L. Biomolecular Force Field Parameterization via Atoms-in-Molecule Electron Density Partitioning *J. Chem. Theory Comput.* **2016**, *12*, 2312–2323.
- [16] Fu, C.-F.; Tian, S. X. A Comparative Study for Molecular Dynamics Simulations of Liquid Benzene *J. Chem. Theory Comput.* **2011**, *7*, 2240–2252.
- [17] Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant *J. Chem. Theory Comput.* **2012**, *8*, 61–74.

- [18] Cui, S.; de Almeida, V. F.; Hay, B. P.; Ye, X.; Khomami, B. Molecular Dynamics Simulation of Tri-n-Butyl-Phosphate Liquid: a Force Field Comparative Study. *J. Phys. Chem. B* **2012**, *116*, 305–13.
- [19] Li, J.; Lakshminarayanan, R.; Bai, Y.; Liu, S.; Zhou, L.; Pervushin, K.; Verma, C.; Beuerman, R. W. Molecular Dynamics Simulations of a New Branched Antimicrobial Peptide: a Comparison of Force Fields. *J. Chem. Phys.* **2012**, *137*, 215101.
- [20] Jahn, D. A.; Akinkunmi, F. O.; Giovambattista, N. Effects of Temperature on the Properties of Glycerol: A Computer Simulation Study of Five Different Force Fields. *J. Phys. Chem. B* **2014**, *118*, 11284–11294.
- [21] Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields *J. Comp. Chem.* **2009**, *31*, 671–690.
- [22] Cacelli, I.; Prampolini, G. Parametrization and Validation of Intramolecular Force Fields Derived from DFT Calculations *J. Chem. Theory Comput.* **2007**, *3*, 1803–1817.
- [23] Grimme, S. A General Quantum Mechanically Derived Force Field (QMDFFF) for Molecules and Condensed Phase Simulations *J. Chem. Theory Comput.* **2014**, 44974514.
- [24] Cacelli, I.; Cinacchi, G.; Prampolini, G.; Tani, A. Computer Simulation of Solid and Liquid Benzene with an Atomistic Interaction Potential Derived from Ab Initio Calculations. *J. Am. Chem. Soc.* **2004**, *126*, 14278–86.
- [25] Amovilli, C.; Cacelli, I.; Cinacchi, G.; Gaetani, L.; Prampolini, G.; Tani, A. Structure and Dynamics of Mesogens Using Intermolecular Potentials Derived from Ab Initio Calculations *Theor. Chem. Accounts* **2007**, *117*, 885–901.
- [26] Verstraelen, T.; Van Neck, D.; Ayers, P.; Van Speybroek, V.; Waroquier, M. The Gradient Curves Method: An Improved Strategy for the Derivation of Molecular

- Mechanics Valence Force Fields from ab Initio Data *J. Chem. Theory Comput.* **2007**, *3*, 1420–1434.
- [27] Stone, A. J.; Misquitta, A. J. Atom-Atom Potentials from Ab Initio Calculations *Int. Rev. Phys. Chem.* **2007**, *26*, 193–222.
- [28] Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-Up Strategy. *J. Chem. Theory Comput.* **2007**, *3*, 1960–1986.
- [29] Akin-Ojo, O.; Song, Y.; Wang, F. Developing *ab initio* quality force fields from condensed phase quantum-mechanics/molecular -mechanics calculations through the adaptive force matching method *J. Chem. Phys.* **2008**, *129*, 64108.
- [30] Cacelli, I.; Lami, C. F.; Prampolini, G. Force-field Modeling through Quantum Mechanical Calculations: Molecular Dynamics Simulations of a Nematogenic Molecule in its Condensed Phases. *J. Comp. Chem.* **2009**, *30*, 366–378.
- [31] Waldher, B.; Kuta, J.; Chen, S.; Henson, N.; Clark, A. ForceFit: A Code to Fit Classical Force Fields to Quantum Mechanical Potential Energy Surfaces *J. Comp. Chem.* **2010**, *31*, 2307–2316.
- [32] Cacelli, I.; Cimoli, A.; Livotto, P. R.; Prampolini, G. An Automated Approach for the Parameterization of Accurate Intermolecular Force-Fields: Pyridine as a Case Study. *J. Comp. Chem.* **2012**, *33*, 1055.
- [33] Barone, V.; Cacelli, I.; De Mitrì, 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*, 3736–51.
- [34] McDaniel, J. G.; Schmidt, J. R. Physically-Motivated Force Fields from Symmetry-Adapted Perturbation Theory. *J. Phys. Chem. A* **2013**, *117*, 2053–66.

- [35] Semrouni, D.; Cramer, C. J.; Gagliardi, L. AMOEBA Force Field Parameterization of the Azabenzenes *Theor. Chem. Accounts* **2014**, *134*, 1590.
- [36] McDaniel, J. G.; Schmidt, J. R. First-Principles Many-Body Force Fields from the Gas Phase to Liquid: a "Universal" Approach. *J. Phys. Chem. B* **2014**, *118*, 8042–8053.
- [37] Schmidt, J. R.; Yu, K.; McDaniel, J. G. Transferable Next-Generation Force Fields from Simple Liquids to Complex Materials *Acc. Chem. Res.* **2015**, *48*, 548–556.
- [38] Chattoraj, J.; Risthaus, T.; Rubner, O.; Heuer, A.; Grimme, S. A multi-scale approach to characterize pure CH<sub>4</sub>, CF<sub>4</sub>, and CH<sub>4</sub>/CF<sub>4</sub> mixtures. *J. Chem. Phys.* **2015**, *142*, 164508.
- [39] Vanduyfhuys, L.; Vandenbrande, S.; Verstraelen, T.; Schmid, R.; Waroquier, M.; Van Speybroeck, V. QuickFF: A program for a quick and easy derivation of force fields for metal-organic frameworks from ab initio input. *J. Chem. Theory Comput.* **2015**, *36*, 1015–27.
- [40] Prampolini, G.; Livotto, P. R.; Cacelli, I. Accuracy of Quantum Mechanically Derived Force-Fields Parameterized from Dispersion-Corrected DFT Data: The Benzene Dimer as a Prototype for Aromatic Interactions. *J. Chem. Theory Comput.* **2015**, *11*, 5182–96.
- [41] Tafipolsky, M.; Ansorg, K. Toward a Physically Motivated Force Field: Hydrogen Bond Directionality from a Symmetry-Adapted Perturbation Theory Perspective *J. Chem. Theory Comput.* **2016**, *12*, 1267–1279.
- [42] Zhu, S. S.; Xu, T.; Yin, S. W.; Wang, Y. Reparameterization of 12-6 Lennard-Jones potentials based on quantum mechanism results for novel tetrahedral N<sub>4</sub> (Td) explosives *Theor. Chem. Accounts* **2016**, *135*, 1–9.
- [43] Paesani, F. Getting the Right Answers for the Right Reasons: Toward Predictive Molecular Simulations of Water with Many-Body Potential Energy Functions *Acc. Chem. Res.* **2016**, acs.accounts.6b00285.

- [44] Cisneros, G. A.; Wikfeldt, K. T.; Ojamäe, L.; Lu, J.; Xu, Y.; Torabifard, H.; Bartók, A. P.; Csányi, G.; Molinero, V.; Paesani, F. Modeling Molecular Interactions in Water: From Pairwise to Many-Body Potential Energy Functions *Chem. Rev.* **2016**, *116*, 7501–7528.
- [45] Wang, Y.; Bowman, J. M.; Kamarchik, E. Five ab initio potential energy and dipole moment surfaces for hydrated NaCl and NaF. I. Two-body interactions *J. Chem. Phys.* **2016**, *144*, 114311.
- [46] Norrby, P.-O.; Liljefors, T. Automated molecular mechanics parameterization with simultaneous utilization of experimental and quantum mechanical data *J. Comput. Chem.* **1998**, *19*, 1146–1166.
- [47] Wang, L.-P.; Martinez, T. J.; Pande, V. S. Building Force Fields: An Automatic, Systematic, and Reproducible Approach *J. Phys. Chem. Lett.* **2014**, *5*, 1885–1891.
- [48] Madarász, Á.; Berta, D.; Paton, R. S. Development of a True Transition State Force Field from Quantum Mechanical Calculations *J. Chem. Theory Comput.* **2016**, *12*, 1833–1844.
- [49] Wang, Y.; Huang, X.; Shepler, B. C.; Braams, B. J.; Bowman, J. M. Flexible, ab initio potential, and dipole moment surfaces for water. I. Tests and applications for clusters up to the 22-mer *J. Chem. Phys.* **2011**, *134*, 094509.
- [50] Medders, G. R.; Babin, V.; Paesani, F. Development of a First-Principles Water Potential with Flexible Monomers. III. Liquid Phase Properties *J. Chem. Theory Comput.* **2014**, *10*, 2906–2910.
- [51] Elrod, M. J.; Saykally, R. J. Many-Body Effects in Intermolecular Forces *Chem. Rev.* **1994**, *94*, 1975–1997.
- [52] Cacelli, I.; Cimoli, A.; De Gaetani, L.; Prampolini, G.; Tani, A. Chemical Detail Force Fields for Mesogenic Molecules *J. Chem. Theory Comput.* **2009**, *5*, 1865–1876.



- [53] Barone, V.; Cacelli, I.; Ferretti, A.; Monti, S.; Prampolini, G. Parameterization and validation of an accurate force-field for the simulation of alkylamine functionalized silicon (111) surfaces. *Phys. Chem. Chem. Phys.* **2010**, *12*, 4201–9.
- [54] Barone, V.; Bloino, J.; Monti, S.; Pedone, A.; Prampolini, G. Theoretical multilevel approach for studying the photophysical properties of organic dyes in solution. *Phys. Chem. Chem. Phys.* **2010**, *12*, 10550–61.
- [55] Pedone, A.; Prampolini, G.; Monti, S.; Barone, V. Realistic Modeling of Fluorescent Dye-Doped Silica Nanoparticles: A Step Toward the Understanding of their Enhanced Photophysical Properties. *Chem. Mater.* **2011**, *23*, 5016–5023.
- [56] De Mitri, N.; Monti, S.; Prampolini, G.; Barone, V. Absorption and Emission Spectra of a Flexible Dye in Solution: A Computational Time-Dependent Approach *J. Chem. Theory Comput.* **2013**, *9*, 4507–4516.
- [57] Prampolini, G.; Yu, P.; Pizzanelli, S.; Cacelli, I.; Yang, F.; Zhao, J.; Wang, J. Structure and Dynamics of Ferrocyanide and Ferricyanide Anions in Water and Heavy Water: An Insight by MD Simulations and 2D IR Spectroscopy *J. Phys. Chem. B* **2014**, *118*, 14899–14912.
- [58] Barone, V.; Cacelli, I.; Ferretti, A.; Prampolini, G.; Villani, G. Proton and Electron Transfer Mechanisms in the Formation of Neutral and Charged Quinhydrone-Like Complexes: A Multilayered Computational Study *J. Chem. Theory Comput.* **2014**, *10*, 4883–4895.
- [59] Fischer, N. M.; van Maaren, P. J.; Ditz, J. C.; Yildirim, A.; van der Spoel, D. Properties of Organic Liquids when Simulated with Long-Range Lennard-Jones Interactions *J. Chem. Theory Comput.* **2015**, *11*, 2938–2944.
- [60] Kola, M. H.; Hobza, P. Computer Modeling of Halogen Bonds and Other  $\sigma$ -Hole Interactions. *Chem. Rev.* **2016**, *116*, 5155–5187.
- [61] Politzer, P.; Lane, P.; Concha, M. C.; Ma, Y.; Murray, J. S. An overview of halogen bonding *J. Mol. Model.* **2007**, *13*, 305–311.

- [62] Desiraju, G. R.; Parthasarathy, R. The nature of halogen-halogen interactions: are short halogen contacts due to specific attractive forces or due to close packing of nonspherical atoms? *J. Am. Chem. Soc.* **1989**, *111*, 8725–8726.
- [63] Clark, T.; Hennemann, M.; Murray, J. S.; Politzer, P. Halogen bonding: the  $\sigma$ -hole *J. Mol. Model.* **2007**, *13*, 291–296.
- [64] Mu, X.; Wang, Q.; Wang, L.-P.; Fried, S. D.; Piquemal, J.-P.; Dalby, K. N.; Ren, P. Modeling organochlorine compounds and the  $\sigma$ -hole effect using a polarizable multipole force field. *J. Phys. Chem. B* **2014**, *118*, 6456–65.
- [65] Adluri, A. N. S.; Murphy, J. N.; Tozer, T.; Rowley, C. N. Polarizable Force Field with a  $\sigma$ -Hole for Liquid and Aqueous Bromomethane. *J. Phys. Chem. B* **2015**, *119*, 13422–32.
- [66] Ibrahim, M. A. A. Molecular mechanical study of halogen bonding in drug discovery *J. Comp. Chem.* **2011**, *32*, 2564–2574.
- [67] Kolá, M.; Hobza, P. On Extension of the Current Biomolecular Empirical Force Field for the Description of Halogen Bonds. *J. Chem. Theory Comput.* **2012**, *8*, 1325–33.
- [68] Jorgensen, W. L.; Schyman, P. Treatment of Halogen Bonding in the OPLS-AA Force Field: Application to Potent Anti-HIV Agents *J. Chem. Theory Comput.* **2012**, *8*, 3895–3901; PMID: 23329896.
- [69] 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*, 10476–93.
- [70] Olejniczak, A.; Katrusiak, A. Structure-property relations in chloroacetonitriles *CrystEngComm* **2011**, *13*, 5212.
- [71] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.; Nakatsuji, H.;

- Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J.; Iyengar, S. S.; Tomasi, J.; ; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Parandekar, P. V.; Mayhall, N. J.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fo, D. J.; Gaussian09, Revision C.01; Gaussian, Inc.; Wallingford CT; 2009.
- [72] Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-Corrected Mean-Field Electronic Structure Methods *Chem. Rev.* **2016**, *116*, 5105–5154.
- [73] Rezáč, J.; Hobza, P. Describing Noncovalent Interactions beyond the Common Approximations: How Accurate Is the Gold Standard, CCSD(T) at the Complete Basis Set Limit? *J. Chem. Theory Comput.* **2013**, *9*, 2151–2155.
- [74] Prampolini, G. Parametrization and Validation of Coarse Grained Force-Fields Derived from ab Initio Calculations *J. Chem. Theory Comput.* **2006**, *2*, 556–567.
- [75] Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit *Bioinformatics* **2013**, *29*, 845–854.
- [76] Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling *J. Chem. Phys.* **2007**, *126*, 014101.
- [77] Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method *J. Appl. Phys.* **1981**, *52*, 7182–7190.

- [78] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104–1–154104–19.
- [79] Lide, D. R. CRC Handbook of Chemistry and Physics, 84th Edition, 2003-2004; CRC press: Boca Raton, FL, 2003; Vol. 53.
- [80] Frenkel, M.; Marsh, K. N. Densities of Halohydrocarbons; Landolt-Börnstein - Group IV Physical Chemistry, Vol. 8J; Springer-Verlag: Berlin/Heidelberg, 2003.
- [81] Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors *Mol. Phys.* **1970**, *19*, 553.
- [82] Cerezo, J.; Santoro, F.; Prampolini, G. Comparing classical approaches with empirical or quantum-mechanically derived force fields for the simulation electronic lineshapes: application to coumarin dyes *Theor. Chem. Accounts* **2016**, *135*, 143.
- [83] Yaws, C. Yaws' Handbook of Thermodynamic Properties for Hydrocarbons and Chemicals; Knovel, 2009.
- [84] Marcus, Y. The properties of solvents; Wiley, 1998.
- [85] An, X.; Hu, H. Enthalpies of Vaporization of Some Multichloro-alkanes *Acta Phys. Chim. Sin.* **1989**, *5*, 565–571.
- [86] Stephenson, R. M.; Malanowski, S. In *Handbook of the Thermodynamics of Organic Compounds*; Springer Netherlands: Dordrecht, 1987; pp 1–471.
- [87] Dean, J. Lange's Handbook of chemistry; McGraw- Hill, Inc., 1999.
- [88] Ketelaar, J. A. A.; de Vries, L.; van Velden, P. F.; Kooy, J. S. Viscosities of cis- and trans- 1-2-dichloroethene, in connection with Eyring's theory of viscous flow *Recueil des Travaux Chimiques des Pays-Bas* **2010**, *66*, 733–745.
- [89] Flick, E. Industrial Solvents Handbook; Noyes Data Corporation: Park Ridge, N.J., 1991.

- [90] Coker, A. K. In *Ludwig's Applied Process Design for Chemical and Petrochemical Plants*, 2010; pp 757–792.
- [91] Rutherford, W. M. Viscosity and density of some lower alkyl chlorides and bromides *J. Chem. Eng. Data* **1988**, *33*, 234–237.
- [92] *Viscosity of Liquids*; Springer Netherlands: Dordrecht, 2007.