The Journal of Chemical Physics

RESEARCH ARTICLE | JULY 24 2020

Automated parameterization of quantum-mechanically derived force-fields including explicit sigma holes: A pathway to energetic and structural features of halogen bonds in gas and condensed phase

Special Collection: Classical Molecular Dynamics (MD) Simulations: Codes, Algorithms, Force fields, and Applications

Marco Campetella ; Nicola De Mitri ; Giacomo Prampolini

J. Chem. Phys. 153, 044106 (2020)
https://doi.org/10.1063/5.0014280

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Automated parameterization of quantum-mechanically derived force-fields including explicit sigma holes: A pathway to energetic and structural features of halogen bonds in gas and condensed phase

Cite as: J. Chem. Phys. 153, 044106 (2020); doi: 10.1063/5.0014280
Submitted: 19 May 2020 • Accepted: 5 July 2020 •
Published Online: 24 July 2020

Marco Campetella,[1] Nicola De Mitri,[2] and Giacomo Prampolini[3,a)]

AFFILIATIONS

[1]Institut des Nanosciences de Paris, Sorbonne Université, CNRS, UMR7588, F-75252 Paris, France
[2]Enthought Ltd., Broers Building, 21 JJ Thomson Avenue, Cambridge CB3 0FA, United Kingdom
[3]Istituto di Chimica dei Composti Organo Metallici (ICCOM), CNR Area della Ricerca, via G. Moruzzi 1, I-56124 Pisa, Italy

Note: This paper is part of the JCP Special Topic on Classical Molecular Dynamics (MD) Simulations: Codes, Algorithms, Force fields, and Applications.
[a)]**Author to whom correspondence should be addressed:** giacomo.prampolini@pi.iccom.cnr.it

ABSTRACT

In classical molecular dynamics, general purpose atomistic force-fields (FFs) often deliver inaccurate results when dealing with halogen bonds (XBs), notwithstanding their crucial role in many fields of science, ranging from material design to drug development. Given the large dimensions of the systems of interest, it would be therefore desirable to increase the FF accuracy maintaining the simplicity of the standard Lennard-Jones (LJ) plus point charge description to avoid an excessive computational cost. A simple yet effective strategy consists in introducing a number of virtual sites able to mimic the so-called "explicit σ-hole." In this work, we present an automated FF parameterization strategy based on a global optimization of both LJ and charge parameters with respect to accurate quantum mechanical data, purposely computed for the system under investigation. As a test case, we report on two homologue series, characterized either by weak or strong XBs, namely, the di-halogenated methanes and the mono-, di-, and tri-substituted acetonitriles, taking into consideration Cl, Br, and I substituents. The resulting quantum mechanically derived FFs are validated for each compound in the gas and in the condensed phase by comparing them to general purpose and specific FFs without virtual sites and to highly accurate reference quantum mechanical data. The results strongly support the adoption of the specific FFs with virtual sites, which overcome the other investigated models in representing both gas phase energetics and the structural patterns of the liquid phase structure related to the presence of XBs.

I. INTRODUCTION

The importance of intermolecular noncovalent interactions (NCIs)[1–4] cannot be overemphasized as they are responsible for the existence of any kind of condensed matter or super-molecular aggregate. Beside the well-known hydrophobic effects,[5] hydrogen bonding,[6–8] or van der Waals interactions,[9] a growing attention has been devoted in the past decade to halogen bonding, as testified by the several dedicated reviews and perspectives.[10–21] Such a success is motivated both by the possibility of multiple applications and by a more fundamental interest. Indeed, on the one hand, XB has been found to play a crucial role in cutting-edge topics such as catalysis,[22,23] material design,[24,25] and drug development.[12,26,27] On the other hand, the counter-intuitive electrophilic behavior of

halogen atoms in the presence of electron-rich groups such as carbonyl, cyano-substituents, π-systems, or even other halogens stimulated a debate that lead to the definition of the σ-hole concept.[28]

In fact, resorting to the seminal paper of Feynman[29] concerning the pure electrostatic nature of intermolecular forces, Clark, Politzer, and co-workers[13,15,28,30–32] first rationalized XB features in terms of σ-hole, i.e., a positive region corresponding to a maximum of the molecular electrostatic potential, located at the tip of the halogen (X). In this framework, XBs are therefore settled by a pure noncovalent electrostatic interaction between a nucleophile and the σ-hole. More recently,[15,18] the σ-hole concept has been extended to other groups, and for instance, chalcogen, pnictogen, and petrel bonding have been defined.[33] Limiting the discussion to XBs, they can be elegantly explained[18,19,28,31,32] through the σ-hole concept as the interaction of a positive region, approximately located on the tip of the involved halogen atom (X), and a negative region, located on the electron donor (D). XB main features,[13,34] such as its directionality (R–X···D angle at ~180°) and consequent anisotropy (head-on interactions), can also be easily rationalized in terms of σ-hole description,[13,28,31,32,35] as well as the variations in XB efficiency along the halogen series. In fact, as the σ-hole is originated by the charge density polarization toward the R–X covalent bond, it is clear that the XB strength diminishes as the X atom polarizability decreases so that the ability of chlorine to form XB is less than that of bromine and iodine, and the ability of fluorine to form XBs is still a subject of debate.[36]

From a computational point of view, the capability of quantum-mechanical (QM) methods to reproduce σ-hole features has been reviewed extensively.[17,20,37–39] Coupled cluster calculations with single, double, and perturbatively included triple excitations, extrapolated at the complete basis set [CCSD(T)/CBS], often referred to as the "gold standard" of quantum chemistry,[40] have an exceedingly high computational cost, which rules out their routine adoption. Yet, they can be safely used as reference[41] to benchmark the performance of cheaper yet less reliable QM methods, as those based on Density Functional Theory (DFT). However, despite the computational convenience, it has been often been reported that the performances of different DFT functionals often depend on the molecular species and on the specific orientation.[17,37–39,42] Given their interest in drug design, Force Field (FF) based methods,[43,44] such as molecular dynamics (MD) simulations, are much more appealing to investigate XBs in large systems.[17,20,21,45–49] Nonetheless, in most popular FFs,[50–54] halogen atoms are usually described by a Lennard-Jones (LJ) potential and a negative single point charge, thus precluding *a priori* a correct representation of the σ-hole. In fact, both LJ and charge–charge model potential functions only depend on the scalar distance between the two involved atoms and, as such, are intrinsically unable to well represent anisotropic features such as the "polar flattening."[15,31,35,55] To overcome this problem, Ibrahim[56] first proposed to add a positive extra-point of charge (EP) at a fixed distance from the halogen atom, thus being able to describe the charge distribution around X more accurately. Since then, several implementations for EP were proposed (see, for instance, Refs. 17, 20, and 21 and references therein) and the term Explicit Sigma Hole (ESH) was introduced.[17,45] Because only a supplementary charge is introduced in the ESH model, most re-parameterizations concern only with the Coulomb electrostatic term of the FF, re-adapting the point charge distribution but leaving unchanged the LJ parameters. In a systematic investigation on ESH properties, Kolář and Hobza pointed out[45] the necessity of a complementary LJ re-parameterization of all the atoms involved. The authors suggested that this would allow in principle for a better reproduction also of the repulsive and dispersion forces, as affected by the σ-hole. Concretely, the comparison of interaction energy curves between several halogen bonded molecular pairs, computed at the CCSD(T)/CBS level and through different FFs with or without ESH implementation,[45] showed that the ESH effect was only able to better tune the well depth and a restricted region of intermolecular separation, while the contact distance and the near attractive region were still subject to large errors. On the same foot, El Kerdawy *et al.*[31] showed that the directional anisotropic features of both dispersion and repulsion interactions involving halogen dimers with argon cannot be reproduced with standard LJ model potentials. Likewise, more recent studies[21,57–59] pointed out that a refinement of the LJ parameters, at least of the halogen atoms, is still necessary to reproduce the fine structural features of liquid bulk structure, even when more complicated models, such as polarizable FFs, are adopted to account for XBs. Following these suggestions, to properly account for XB patterns, a re-parameterization of the whole set of parameters should be performed on most transferable FF libraries, such as the very recent OPLS3 FF,[59] where experimental properties and QM data of an extended training set were employed to tune LJ, charge, and intramolecular parameters.

Thanks to the increase of the available computational resources, an alternative route to FF parameterization has received a lot of attention in the past decade,[60–69] which consists in deriving the parameters solely from QM data, purposely computed for the target under investigation. One of the advantages of such a specific description, whose drawback is the limited transferability of the resulting parameters to different target compounds, stands in the increased accuracy, in turn rooted into the QM description encoded in the parameterized classical FF. In this framework, the possibility to tailor the chemical details delivered by QM calculation into a simple FF model potential has been investigated also in our group in a long lasting effort,[70–72] which recently led us to report on a complete re-parameterization of the LJ (and point charges) parameters for a number of halogenated hydrocarbons,[42] exploiting the PICKY parameterization procedure,[42,67,73,74] an automated protocol developed to produce quantum-mechanically derived FFs (QMD-FFs). The proposed QMD-FFs delivered on average more accurate performances with respect to popular transferable FFs, and some of the σ-hole features, which appear in the reference QM description, were successfully transferred and encoded into the QMD-FF model.[42] Nonetheless, some of the species benchmarked therein presented standard deviations with respect to the QM reference worse than average. Since such species were those where XBs are expected to be stronger, hence enforcing the need to take σ-holes into account,[32] it was then hypothesized[42] that the adoption of more complex FF potentials could significantly reduce the discrepancies with the QM description. To this end, the first step appears to be the introduction of a supplementary virtual site, i.e., the ESH. It is worth mentioning that the ESH approach has the further advantage, for instance, with respect to polarizable FFs, of a rather low, almost negligible, impact on the computational cost, being thus applicable to large-scale simulations. Moreover, a more accurate polarizable model can be

built in a second step on top of a "zero-order" simple QMD-FF model, with refined LJ, point charge, and ESH parameters (QMD-FF+ESH).

The aim of the present work is therefore threefold: (*i*) code and implement within the PICKY[42,67,73,74] procedure an automated and robust protocol to parameterize, with respect to purposely computed QM reference data, both ESH point charge (and, more in general, any virtual site) and LJ atomic parameters; (*ii*) investigate to what extent the resulting QMD-FF+ESH is able to reproduce energetic and structural patterns expected for gas and condensed phases characterized by XBs and evaluate the procedure accuracy with respect to popular transferable FFs; and (*iii*) assess for which species the introduction of ESH is requisite or if a simple re-parameterization of the LJ parameters is sufficient. Two benchmark sets, shown in Fig. 1, will be considered to represent weak and strong XBs, respectively. The first set is constituted by simple di-halogenated methanes, considering chlorine, bromine, and iodine substitutions (**dcm**, **dbm**, and **dim**, top row of Fig. 1). Despite some of these compounds having been the subject of previous investigations employing ESHs,[17,20,45,46,57,58] to our knowledge, this is the first time that they are considered in a systematic way and subjected to a full LJ+ESH re-parameterization with respect to a reliable QM reference. To consider strong XB effects, the second benchmark set considered consists in the whole halo-acetonitriles series (displayed in the last three rows of Fig. 1), the halogen atom being either Cl (**mca**, **dca**, and **tca**), Br (**mba**, **dba**, and **tba**), or I (**mia**, **dia**, and **tia**), where the σ-hole effect is expected to be enhanced by the presence of the electron rich CN group.

II. METHODS

A. QMD-FF parameterization without ESH

A target-specific QMD-FF, without any virtual site, was developed for each of the species included in the two benchmark sets shown in Fig. 1. Thus, QMD-FFs were parameterized by means of the JOYCE[71,75,76] and PICKY[42,67,73,74] parameterization procedures, proposed by some of us some years ago[71,73] and recently improved, automated, and refined.[67,76] According to this protocol, the QMD-FF parameters are derived by minimizing well defined objective functions,[71,73] devised to minimize the differences in some key properties between the classical, FF based description and a reference one, purposely obtained at the QM level. As far as four of the investigated species (i.e., **dcm**, **dbm**, **dca**, and **tca**) are concerned, QMD-FF parameters were taken from our previous work on halogenated hydrocarbons.[42] For the remaining eight compounds, a specific QMD-FF was purposely built by means of the JOYCE and PICKY programs,[77,78] following the same procedure based on QM data adopted previously[42] and described in more detail in the supplementary material.

B. QMD-FF parameterization with ESH

As discussed by Kolář and Hobza[45] and displayed in Fig. 2, the implementation of mass-less ESH virtual sites requires to specify two additional parameters, namely, the distance **d** of the ESH site from the halogen center and the value of the charge q^{ESH} assigned to the extra point. In Ref. 45, the authors show how the best results, with respect to a QM reference, are achieved by the "all-fit" model,

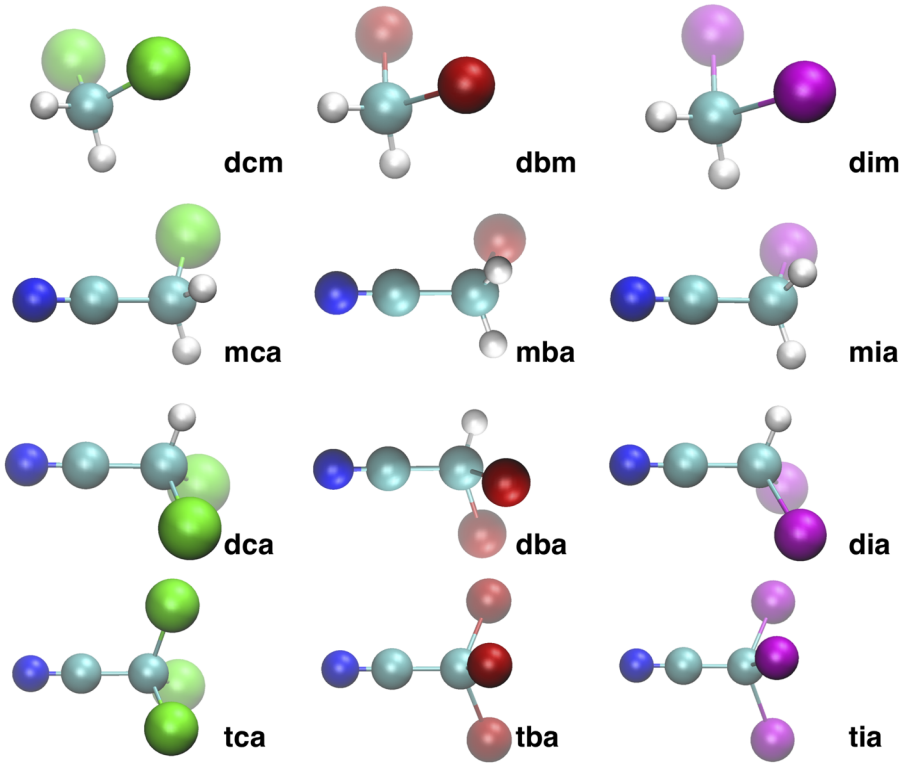

FIG. 1. Benchmark sets of molecules employed in this work. From top to bottom: di-halo-methanes (**dcm**, **dbm**, and **dim** for CH_2Cl_2, CH_2Br_2, and CH_2I_2, respectively) and halo-acetonitriles. Mono-(**mca**, **mba**, and **mia**), di- (**dca**, **dba**, and **dia**), and tri- (**tca**, **tba**, and **tia**) halo substituted acetonitriles are shown in the second, third, and fourth rows, respectively, while chlorine, bromine, and iodine substituted species are displayed in the first, second, and third columns.

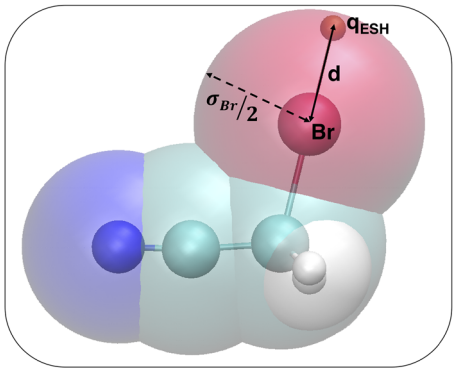

FIG. 2. Modeling of the ESH for the bromo-acetonitrile (**mba**) molecule. The Br-ESH distance **d** is marked with a solid arrow, whereas for bromine, the van der Waals radius ($\frac{\sigma_{Br}}{2}$) is shown with a dashed line.

i.e., when the ESH charge was fitted together with all the rest of the molecular point charges. One of the aims of the present work is to further refine such a model, allowing for all intermolecular parameters (concretely also LJ's ϵ and σ parameters of each atom), rather than only the point charges, to re-adapt to the introduction of the new ESH interaction site.

To this end, resorting to our previous work on halogenated hydrocarbons,[42] we implemented the automated protocol sketched in Fig. 3. The initial step of the whole procedure is to assess a reliable value for the distance **d** (see Fig. 2), at which the ESH should be placed. In fact, in Ref. 45, Kolář and Hobza showed that **d** can be relevant for an accurate description of the σ-hole: notwithstanding the authors report of its best value being around 1.6 Å, such a value was also shown to be dependent both from the target species and from the QM level adopted as reference. Here, we exploit, separately for each compound, the large database of dimer interaction energies, computed at the QM level on several hundred (N_{geoms}) geometries (see Figs. S3–S13 of the supplementary material for a detailed analysis) during the standard QMD-FF parameterizations, represented as blue arrows in Fig. 3.

When the ESH site is introduced, the standard expression of the intermolecular energy given in Eq. (S7) becomes

$$E_{AB}^{FF_{inter+ESH}}(\tilde{r}_{AB};\mathbf{d}) = E_{AB}^{FF_{inter}}(\tilde{r}_{AB}) + \sum_{i=1}^{N_A}\frac{q_i q_B^{ESH}}{r_{i-ESH}} + \sum_{i=1}^{N_B}\frac{q_A^{ESH} q_j}{r_{ESH-j}} + \frac{q_A^{ESH} q_B^{ESH}}{r_{ESH-ESH}}, \quad (1)$$

where $E^{FF_{inter+ESH}}$ now depends explicitly on the ESH charge (q^{ESH}) and implicitly on its position **d**. Consequently, the Picky objective function (S10) also changes to

$$I^{inter+ESH}[\mathbf{d}] = \frac{\sum_{k=1}^{N_{geom}}\left[\left(\Delta E_k^{QM_{inter}} - E_k^{FF_{inter+ESH}}[\mathbf{d}]\right)^2\right]e^{-\alpha\Delta E_k^{QM_{inter}}}}{\sum_{k=1}^{N_{geom}} e^{-\alpha\Delta E_k^{QM_{inter}}}}, \quad (2)$$

and the standard deviation σ_P depends on the distance **d**,

$$\sigma_P[\mathbf{d}] = \sqrt{I^{inter+ESH}[\mathbf{d}]}. \quad (3)$$

To obtain a reliable **d** value, a single PICKY fitting step, over the aforementioned QM database of N_{geom} dimer interaction energies, was carried out through Eq. (2) at a number of selected distances **d**. Eventually, as shown with violet arrows in Fig. 3, $\sigma_P[\mathbf{d}]$ is employed, together with other relevant parameters as $q^{ESH}[\mathbf{d}]$ and the LJ parameters of the X halogen ($\frac{1}{2}\sigma_X[\mathbf{d}]$ and $\epsilon_X[\mathbf{d}]$), to assign the most reliable distance **d** separately for each considered species.

Once the best distance **d** has been determined, the QMD-FF+ESH parameterization procedure is carried out as follows: First, as indicated at the top of Fig. 3, all intramolecular parameters are transferred from the ones previously obtained through the JOYCE protocol for the QMD-FF. Indeed, no additional term needs to be re-parameterized in the $E^{FF_{intra}}$ intramolecular term because the ESH virtual site is at a fixed position with respect to the halogen atom. Conversely, the parameterization of $E^{FF_{inter}}$ is performed through a development version of the PICKY code, which is able to include ESH

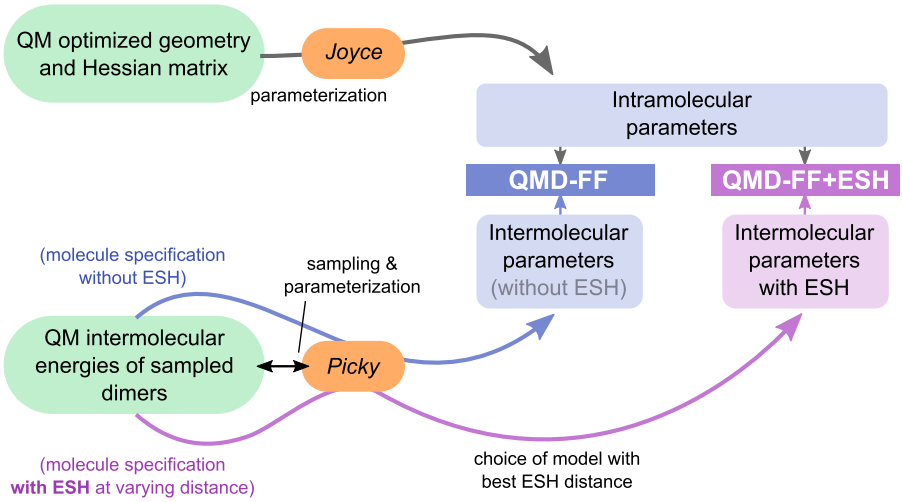

FIG. 3. Implementation of the ESH additional sites into the JOYCE/PICKY QMD-FF parameterization protocol. The standard QMD-FF parameterization[42,67] is employed to find the intramolecular parameters (JOYCE route, gray arrows) and the charges and LJ parameters in the absence of σ-holes (PICKY route, blue arrows); the resulting database of dimer QM interaction energies is then used as reference to assess the best ESH-X distance, **d**, by performing a single Picky fitting step at different **d** separations; once **d** is fixed, the PICKY protocol is repeated (violet arrows) by parameterizing point charge, ESH, and LJ parameters at once.

and, more in general, virtual sites, as described in Eqs. (1) and (2). The final QMD-FF+ESH intermolecular parameters are obtained at the end of the PICKY iterative procedure, when the ΔP convergence criterion is met [see Eq. (S11) of the supplementary material].

C. Validation

For each considered species, the accuracy and reliability of the parameterized QMD-FFs, with and without ESH, and of a transferable, general purpose FF (OPLS[51]) was assessed by comparing all FF predictions against either experimental or high level QM computed data, both in the gas and in the liquid bulk phase. As far as the former phase is concerned, we exploit the large number of QM interaction energies, computed on hundreds of representative dimer geometries selected by PICKY. First, the representativity of the sample was assessed by monitoring the distribution, across the collected dimers, of selected geometrical descriptors such as the inter-atomic halogen–donor (d_{X-D}) distance or the carbon–halogen···donor angle (θ_X) displayed in Figs. S3–S13 of the supplementary material. Next, the OPLS, QMD-FF, and QMD-FF+ESH models were benchmarked with respect to the QM reference dimer interaction energy across the large portions of the sampled configurational space.

Turning to the liquid phase, for each investigated species, three sets of classical MD simulations are carried out on systems composed of 1000 molecules in the NPT ensemble at standard conditions, separately using each one of the investigated FFs. The resulting condensed phases are compared in terms of liquid structure, described by atom–atom pair correlation functions [see Eq. (S12) of the supplementary material] and several other geometrical descriptors (*vide infra*), and eventually validated with respect to the structural patterns emerging from *ab initio* MD (AIMD) trajectories, purposely carried out on each investigated system. Concretely, the spatial arrangement of molecular dimers in the liquid phase has been characterized by analyzing three pair distributions, computed for all XB pairs within the first interaction shell, as defined by the previously mentioned pair correlation functions. The θ_X normalized distribution,

$$\hat{p}(\theta_X) = \frac{p(\theta_X)}{\sin(\theta_X)}, \tag{4}$$

was computed to measure the distribution of X···D contacts per unit surface across the spherical cap around the halogen nucleus. Next, the distribution of the X···D–C angle (θ_D) for all contacts with a given θ_X, $p(\theta_D)|_{\theta_X}$, was evaluated for each model and compound to determine the region of the electron donor's van der Waals surface that interacts with each region of the halogen surface. Finally, a series of pair distribution functions for the XB contact, taken at given intervals of θ_X, $p(d_{X\text{-}D})|_{\theta_X}$, was also computed along each MD trajectory. It might be worth noting that for a fully isotropic interaction, all these functions would be equal to the non-conditional pair distribution function, $p(d_{X-D})$.

III. COMPUTATIONAL DETAILS

A. QM calculations

All QM calculations were performed with the GAUSSIAN09 package.[79] The QM data required for the intramolecular parameterizations, namely, one geometry optimization and one Hessian matrix calculation at the resulting optimized geometry, were obtained for each isolated monomer at the DFT level, employing the standard B3LYP functional with the Dunning's correlation consistent cc-pvDz basis set, in line with our previous work on halogenated hydrocarbons.[42] On the same foot, as far as the intermolecular term is concerned, the reference QM dimer energies were computed at the DFT level, by means of the popular B3LYP functional, using a 6-311G(2d,p) basis set and the D3GBJ[80,81] correction for dispersion. In fact, in Ref. 42, this level of theory, employed to describe the QM interaction potential energy surface (IPES) required for PICKY parameterizations, was validated with respect to CCSD(T)/CBS calculations for the halogenated species considered therein and was proven to be an excellent compromise between accuracy and computational cost.

B. MD simulations

All classical MD simulations were performed with the GROMACS5.1 engine[82] on systems composed of 1000 molecules. During PICKY parameterization cycles, each system was simulated for 500 ps in the NPT ensemble, keeping the temperature and pressure constant through the Berendsen thermostat and barostat, with τ_T and τ_P equal to 0.1 ps and 5.0 ps, respectively. Once the best fit parameters were obtained, each QMD-FF was employed in a 10 ns production run, carried out again in the NPT ensemble, but through a velocity-rescale thermostat[83] and the Parrinello–Rahman scheme.[84] The same equilibration/production procedure was adopted also in the preliminary simulations carried out, for comparison purposes, with a transferable FF, assembled adopting the OPLS[51] LJ parameters and the point charges that were derived from QM calculations using the CM5[85] scheme at the B3LYP/cc-pvDz level. In all simulations, a cutoff of 14 Å was employed for both charge–charge and LJ terms, whereas long-range electrostatics was accounted through the particle mesh Ewald (PME) procedure. Bond lengths are constrained at their equilibrium value using the LINCS algorithm, allowing for a 1 fs time step.

Due to their high computational cost, all the AIMD dynamics were performed on smaller simulation boxes, composed of 40 molecules each. For all the systems, pre-equilibration was performed with classical MD, using the QMD-FF previously obtained. The final configurations of the classical trajectory were used as starting points for the AIMD simulations, which were accomplished with the program package CP2K,[86] using the Quickstep module[87] and the orbital transformation[88] for faster convergence. Although the B3LYP functional could not be used in AIMD in periodic boundary conditions, the electronic structure was calculated still in the DFT framework,[89,90] but using the PBE[91] functional, with the explicit van der Waals terms according to the empirical dispersion correction (D3) by Grimme.[80] Indeed, in our previous work, this functional was found to yield a similar standard deviation with respect to B3LYP, over a set of 50 higher level QM interaction energies.[42] Basis sets belonging to the MOLOPT-DZVP-SR-GTH[92] family and GTH pseudopotentials[93,94] were applied; the time step chosen was 0.5 fs, and the target temperature was set at 300 K by a Nosé–Hoover chain thermostat. After 7 ps of QM-equilibration, NVT trajectories of 32 ps were obtained. The parameters for the thermostat and SCF convergence criteria have been set in order to have a stable dynamics, as shown in previous works.[95–97] Finally, to further assess the

negligible dependence of the AIMD results on the starting simulation box,[98,99] preliminary tests on **dcm** were carried out in the NVT ensemble at different initial densities (see, for instance, Fig. S15 of the supplementary material).

IV. RESULTS

A. QMD-FF+ESH parameterization

1. Setting distance d

The first step of the QMD-FF+ESH parameterization consists in finding a reliable distance **d** between the halogen atoms in the considered species and the ESH virtual site, as displayed in Fig. 2. To this end, a virtual ESH is introduced at a fixed distance **d** from each halogen atom, and a single PICKY fitting step is performed, for each considered species, by minimizing the objective function (2) with respect to the QM reference. As described in Sec. II, we exploit the QM interaction energy databases acquired during the preliminary QMD-FF parameterizations without virtual sites (see the supplementary material for details), which consist in 300–750 points, depending on the considered species. The main features of selected QMD-FF+ESH parameterizations are summarized in Fig. 4 for **mca**, **mba**, and **dba**, where selected quantities are monitored with respect to selected **d** values. Similar results were obtained for all other species.

By looking at the top panels, it is evident that the introduction of the ESH virtual site contributes to the reduction of the deviation between the QM and FF descriptions for all compounds. The main indication for setting the ESH-X distance is that σ_P slightly decreases with increasing **d**. Yet, placing the ESH outside the LJ sphere of halogen bearing it may introduce overlap artifacts when considering bulk systems because no LJ parameters are employed for such a virtual site. For this reason, and considering also the fact that all the monitored indicators agree in showing a rather converged value around 1.6 Å–1.7 Å, the chosen **d** value was fixed at ~0.1 Å below the LJ radius ($\frac{1}{2}\sigma_X$) of the connected halogen. The resulting distances **d** for each considered species are collected in Table I.

TABLE I. Distances **d**, between the ESH site and the halogen center, obtained through the PICKY fitting of the QM dimer energy database.

Mol.	**d** (Å)	Mol.	**d** (Å)	Mol.	**d** (Å)	Mol.	**d** (Å)
dcm	1.60	**mca**	1.50	**dca**	1.50	**tca**	1.50
dbm	1.65	**mba**	1.70	**dba**	1.70	**tba**	1.70
dim	1.75	**mia**	1.70	**dia**	1.70	**tia**	1.70

2. Global parameterization

An iterative PICKY parameterization was then performed on each species displayed in Fig. 1 by introducing the ESH site at the chosen distance **d** and obtaining all LJ, ESH, and charge parameters by again fitting the objective function (2) during each cycle, until convergence was reached. In Fig. S1 of the supplementary material, the behavior of selected convergence indicators is shown for the series of bromo-substituted acetonitriles. Very similar trends were found during parameterization for all investigated species, leading to a standard deviation reduced with respect to the previous QMD-FF without the ESH model, with the average QMD-FF+ESH standard deviation being below 2.5 kJ/mol for all compounds. In all cases, the introduction of the ESH contributed to lower deviations with respect to the QM reference, but for the Br or I substituted compounds, its effect appears to be more impactful.

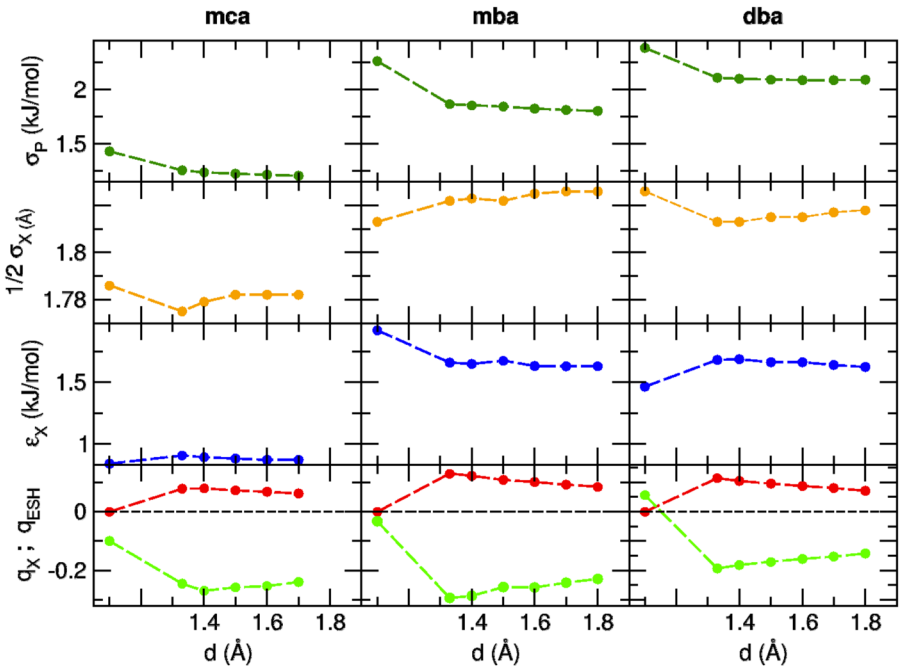

FIG. 4. Selected properties monitored as a function of the halogen-ESH distance **d**. From top to bottom: the standard PICKY deviation σ_P, given in Eq. (3); the halogen LJ radius ($\frac{1}{2}\sigma_X$); the halogen LJ well depth (ϵ_X); and the final charges paced on the halogen (q_X) and on the virtual site (q_{ESH}). The results are shown for the **mca**, **mba**, and **dba** species.

All final QMD-FF and QMD-FF+ESH parameters are reported in detail in Tables A–K of the supplementary material, whereas some relevant features are summarized in Tables II and III. As far as the smaller halomethanes are concerned, Table II highlights the point charge variation on the halogen atoms (X) and on the carbon bearing them, in going from the OPLS to the QMD-FF+ESH parameters. A complete list of all the parameters can be found in Table A of the supplementary material. Furthermore, in the last rows of Table II, we also show the molecular dipole of each species, computed at the QM level on the isolated molecule or through the atomic charges of the investigated FFs. As expected, the simple atomic charge schemes (i.e., OPLS and QMD-FF) do not account even qualitatively for the trends shown by the QM reference, where the dipole is significantly reduced in going from chlorine to iodine. Indeed, in OPLS, where the CM5 charges were employed, the dipole computed with the FF shows a non-monotonic behavior, while in QMD-FF, μ is almost constant along the series. The qualitative agreement is recovered upon introduction of the ESH virtual site. It might be worth to mention that a quantitative agreement between the QM dipole and the one computed with the QMD-FF+ESH parameters should not be expected as the former is a value computed for isolated CH_2X_2 molecules, whereas the latter is obtained through a fitting procedure based on a two-body QM interaction energy surface, hence accounting, on average, for polarization effects. Turning to the halo-acetonitriles series, a similar point charge analysis performed for halo-methanes is presented for the halogen atoms in Table III, while a complete list of all the parameters can be found in Tables C–K of the supplementary material. On the one hand, by looking at the di-substituted compounds, very similar trends to the **dcm**-**dbm**-**dim** series appear: upon ESH insertion, the charge on the halogen atoms becomes more negative; such an effect increases along the halogen series. On the other hand, it is clear that the effectiveness of the ESH site increases with the number of halogen substituents because the value of the ESH charge becomes larger, independently from the kind of halogen employed. Yet, the optimal q_{ESH} value appears to markedly depend on both the considered halogen and the number of substituents, ranging from a minimum value of 0.045, found for **dca**, to a maximum of almost 0.2, in the case of **tia**.

As reported in our previous work on halogenated hydrocarbons,[42] the standard QMD-FF parameterization, i.e., without any additional virtual site, already delivered rather accurate results, given that the XB relevance on the interaction patterns plays only a minor effect. Such a result was possible, thanks to the re-optimization of all LJ parameters, which allowed, together with the parameterized point charges, to achieve a partial description of the XB patterns.[42] The advantage of a global optimization, which optimizes all LJ parameters and point charges at once, is evident by looking at Table III, where the LJ ϵ parameter is reported for selected atoms, namely, the halogens and the carbon bearing them, of all investigated acetonitriles. Indeed, by comparing the QMD-FF and QMD-FF+ESH values to the OPLS ones, the specificity of the QM derived parameters emerges: the halogen LJ parameters in the OPLS description do not depend on the number of substituents, and only one value is employed for carbon, independently from the nature or the number of substitutions. Conversely, rather large and systematic variations are registered for QMD parameters, which appear also to be sensible to the introduction of the virtual site, even if their variation is considerably less than the one observed in going from OPLS to QMD-FF.

TABLE II. Comparison of the point charges and molecular dipoles (μ, debyes) obtained with different models for the investigated dihalomethanes. Top two tables: final PICKY point charges for the halogen atoms (X) and the central carbon (C) obtained in the QMD-FF and QMD-FF+ESH parameterizations, with the CM5 charges employed with OPLS reported for comparison. For X atoms in the QMD-FF+ESH model, the charge q^{ESH} on the virtual site is also shown. Bottom table: total molecular dipoles of the dihalomethane molecules described with the three investigated FFs, compared with the dipole moment obtained at the DFT level (QM).

		dcm		**dbm**		**dim**	
FF	Atom	q		q		q	
OPLS	C	−0.006		0.002		0.000	
QMD-FF	C	−0.026		−0.632		−0.325	
QMD-FF+ESH	C	0.251		0.301		0.263	
FF	Atom	q	q^{ESH}	q	q^{ESH}	q	q^{ESH}
OPLS	X	−0.100	...	−0.110	...	−0.063	...
QMD-FF	X	−0.193	...	−0.100	...	−0.056	...
QMD-FF+ESH	X	−0.257	0.055	−0.344	0.121	−0.356	0.147
Method		μ		μ		μ	
QM		1.71		1.36		1.15	
OPLS		1.56		1.71		1.06	
QMD-FF		1.99		2.01		1.90	
QMD-FF+ESH		1.93		1.71		1.45	

TABLE III. Final LJ well depth parameters (ϵ^{LJ}, kJ/mol) obtained in PICKY QMD-FF and QMD-FF+ESH parameterizations for the halogen atoms and the carbon atom bearing them in the investigated halo-acetonitrile series. In the first rows, the OPLS parameters are shown for comparison.

FF	Atom	**mca**	**dca**	**tca**
OPLS	C	0.293	0.293	0.293
QMD-FF	C	2.764	2.870	0.005
QMD-FF+ESH	C	2.229	2.225	0.005
OPLS	Cl	1.113	1.113	1.113
QMD-FF	Cl	0.838	1.200	1.452
QMD-FF+ESH	Cl	0.740	1.164	1.208
FF	Atom	**mba**	**dba**	**tba**
OPLS	C	0.293	0.293	0.293
QMD-FF	C	1.030	2.690	0.005
QMD-FF+ESH	C	1.581	2.919	0.864
OPLS	Br	1.966	1.966	1.966
QMD-FF	Br	1.920	1.464	2.153
QMD-FF+ESH	Br	1.773	1.218	1.820
FF	Atom	**mia**	**dia**	**tia**
OPLS	C	0.293	0.293	0.293
QMD-FF	C	0.197	0.513	0.006
QMD-FF+ESH	C	1.189	0.401	0.099
OPLS	I	2.426	2.426	2.426
QMD-FF	I	2.801	3.846	3.846
QMD-FF+ESH	I	3.108	3.854	3.863

B. Gas phase

The observations made so far, based on the analysis of the final FF parameters and their behavior during parameterization, suggest that the σ-hole effect has been at least in part encoded in the QMD-FF+ESH description. Yet, the final validation of the parameterization procedure cannot overlook the capability of the parameterized force-fields to account for XB interactions in the gas and condensed phases, thus correctly describing the structural and energetic patterns triggered by the σ-hole.

The reliability of the OPLS, QMD-FF, and QMD-FF+ESH models to account for XB interactions in the gas phase was evaluated by comparing the FF dimer IPESs with those obtained through accurate QM calculations. The left panel of Fig. 5 shows the distribution across the sampled geometries of the QM interaction energy $\Delta E^{QM_{inter}}$ and some relevant geometrical descriptors (displayed in the top inset) for **dca**. Similar trends were observed for all species, as shown in the supplementary material in Figs. S3–S13. It is evident that a rather large portion of the accessible dimer IPES is represented by the selected sample, with $E^{QM_{inter}}$ being well sampled in the −18 kJ/mol–10 kJ/mol interval, while geometrical descriptors relevant to the XB, such as the X–D distance or the θ_D and θ_X angles described in Sec. II, also span a wide range. Given the robustness of the QM database, the FF accuracy in the gas phase can be evaluated by computing the standard deviation of the MM computed interaction energy with respect to the QM reference database. The final results for all compounds are summarized in Table IV, whereas the right panel of Fig. 5 shows the deviation between the FF energies and the QM reference for all molecules in the data set at varying d_{X-N} and θ_X. The better performance of the QMD-FF+ESH model is evident in both cases, yielding the lowest deviation for all compounds and reducing significantly the error spread in the interaction energy appearing in the right panel of Fig. 5.

C. Liquid phase

To validate the parameterized FFs in the condensed phase, lengthy NPT MD simulations were performed at ambient condition on systems composed by 1000 molecules for all the investigated species. A detailed report on the computed thermodynamic properties of the resulting liquid phases can be found in Tables L and M of the supplementary material. As far as the bulk density is concerned, the standard deviations with respect to the experiment, averaged over all investigated compounds and achieved with the QMD-FF (132.9 kg/m^3) or QMD-FF+ESH (133.2 kg/m^3) models, are slightly larger than the one found for OPLS (101.7 kg/m^3). The OPLS model generally gives better bulk density estimates, with the exception of dcm, where the OPLS error of ~100 kg/m^3 is about halved by QMD-FF+ESH. Both QMD-FF and QMD-FF+ESH seem to suffer from a systematic overestimation of densities (see also Fig. S14 of the supplementary material): this is in line with a ~5% error already found in past PICKY parameterizations,[42,67,74] and previously linked to the absence of explicit three-body interactions.[67,74] In fact, the better OPLS performance in reproducing the experimental density should not be surprising, considering that, as in most of the general purpose FFs, OPLS LJ parameters are tuned to reproduce few selected experimental macroscopic properties, among which is the bulk density. Yet, in terms of detecting the peculiar features induced by halogen bonds, it is mostly important to assess the FF performances in describing, at the microscopic level, the structural and geometrical patterns established between the molecules forming the bulk.

A deeper insight into the liquid structure can then be gained through the analysis of the atom–atom pair correlation functions, $g_{\alpha\beta}(r)$. Since no experimental data, such as liquid structure factors, are to our knowledge readily available for all investigated species to validate such microscopic descriptions, AIMD simulations were purposely carried out to deliver a reference picture. Figure 6 shows the distributions $g_{CC}(r)$ of the distance between carbon atoms of a pair of neighboring di-halomethane molecules. With respect to both the OPLS and the QMD-FF models, the effect of ESH insertion is clearly visible and increases along the series: the first neighbor peak is shifted to shorter distances by about 0.7 Å in **dcm**, by 0.9 Å in **dbm**, and by 1.2 Å in **dim**. More importantly, this is in agreement with the indications gained by the AIMD runs, which seem to confirm both the shift and the new position of the first neighbor peak. Similar results were observed for other atom pairs, as shown in Figs. S16–S18 of the supplementary material. For instance, Fig. S19 displays the pair correlation between halogen atoms, $g_{XX}(r)$, along the di-halomethane series, confirming that, in agreement with the gas phase results, the implementation of the ESH model enforces the agreement with the reference QM predictions, allowing the first neighbor peak to significantly increase its intensity, as could be expected considering an XB behavior.

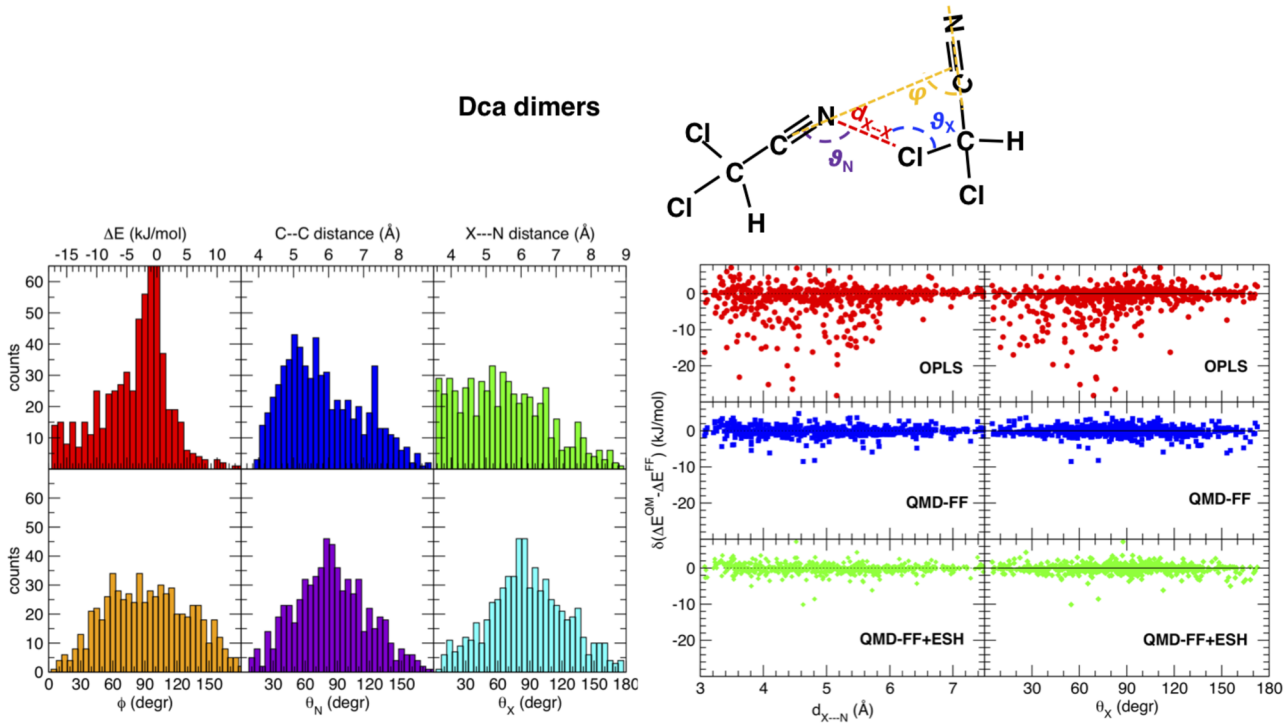

FIG. 5. Gas phase results for **dca**. Left: distribution over the PICKY sample set of 650 dimer geometries of selected properties: the QM interaction energy $\Delta E^{QM_{inter}}$ (simply labeled as ΔE, for the sake of brevity), the distance between the tetrahedral carbon atoms of the two molecules, the interaction distance X⋯N, the C–C⋯N angle ϕ, and the θ_D and θ_X angles described in the text (see also the scheme at the top). Right: deviations of OPLS, QMD-FF, and QMD-FF+ESH energies with respect to $\Delta E^{QM_{inter}}$ along the selected coordinates, d_{X-N} and θ_X.

The same g(r) analysis was performed also for the second investigated class of compounds, the halogenated acetonitriles. Given the nature of the cyano group, more marked hints of a XB network were expected to be revealed by the different correlated pairs, in particular, for the more substituted species. Figure 7 collects some of the most relevant pair correlation functions: the N–Cl distance distribution is shown at increasing numbers of chlorine substituents (top panels), while the pair correlation between the substituted carbons is reported for the series of mono-substituted acetonitriles with different attached halogens (bottom panels). As far as **mca** is concerned, the liquid structure with respect to the AIMD reference is again well reproduced by QMD-FF+ESH, considering both N–X and C–C distances. Changing the substituent produces similar effects as in the di-halomethane series, and the shorter neighbor distance found by AIMD is again well accounted for by QMD-FF+ESH. Apparently, the picture is much less clear when increasing the number of substituents: in **tca**, the signal seems to be almost unaffected by the differences in the models, and in **dca**, the most significant differences are found when looking at the second neighbor shell. More in general, as appears from an overview of all pair correlation functions displayed in the supplementary material, the indications gained from the reference AIMD runs are in better agreement with the FF when the ESH is explicitly included, although in some cases, the QMD-FF model without ESH seems to offer some increased accuracy. Finally, as displayed in Figs. S26–S28 of the supplementary material, it is worth noting that the worse agreement with AIMD prediction was registered in the iodine substituted species. In fact, by looking at $g_{N-I}(r)$ functions, AIMD predicts the appearance of a small peak at very short distances, reinforced along the **mia**, **dia**, **tia** series, which is not or only partially accounted by the QMD-FF+ESH

TABLE IV. Standard deviation (σ^P, kJ/mol) with respect to the QM reference of the FF interaction energies, obtained with the OPLS, QMD-FF, and QMD-FF+ESH models on the database of N_{geom} dimer arrangements collected during PICKY parameterizations.

Mol.	N_{geom}	σ^P_{OPLS}	σ^P_{QMD-FF}	$\sigma^P_{QMD-FF*ESH}$
dcm	500	4.4	2.0	1.5
dbm	350	6.0	2.6	1.8
dim	650	5.6	3.3	2.4
mca	500	8.6	1.7	1.4
mba	600	6.5	2.0	1.7
mia	400	7.9	3.3	2.2
dca	650	6.1	1.2	1.2
dba	400	9.8	2.9	1.9
dia	750	7.9	3.4	2.4
tca	300	3.6	2.0	1.1
tba	600	7.6	4.3	2.2
tia	450	4.4	2.8	2.4

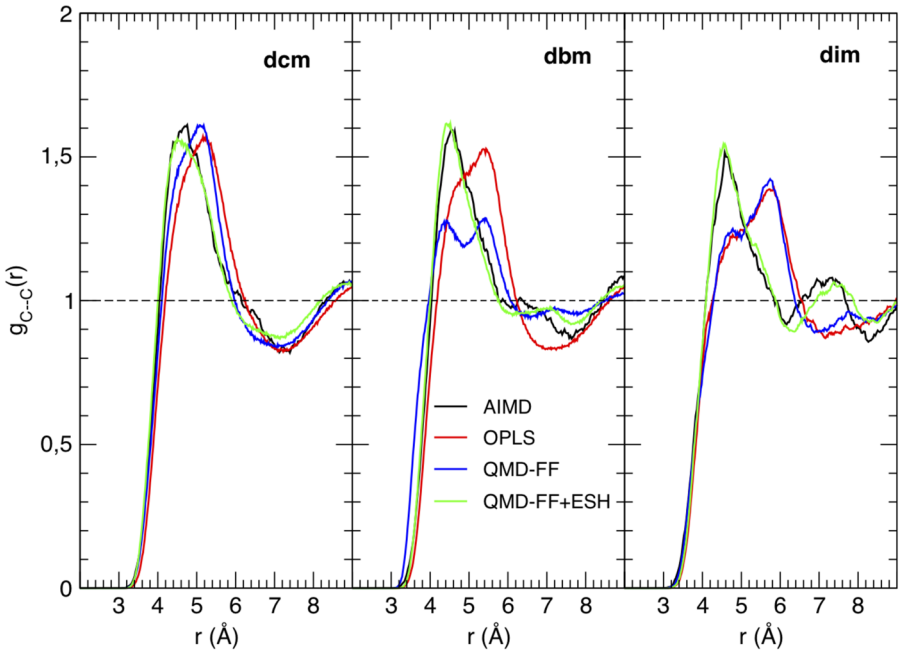

FIG. 6. Atomic pair correlation functions $g_{CC}(r)$ between carbon atoms of **dcm**, **dbm**, or **dim** computed over MD runs carried out with AIMD (black) or with different FF models (OPLS, QMD-FF, and QMD-FF+ESH in red, blue, and green, respectively).

model. Yet, this lack is most likely connected to the well-known limits of the R^{12} LJ repulsive term, which could be emphasized in larger atoms as iodine, but an investigation of alternative model potential functions for the repulsive branch is beyond the aims of the present work.

On average, the pair correlation functions strongly support the adoption of a specific QMD-FF+ESH parameterization and give additional hints of the presence of XB interactions, which lead to shorter intermolecular X–D distances; nonetheless, it is not trivial to extract the geometrical patterns that are expected to characterize XB from a combined analysis of the atom–atom $g_{\alpha\beta}(r)$. To better understand the structural and geometrical features that characterize the XB networks established within neighboring molecules in the liquid, we can exploit the radii of the first-neighbor sphere retrieved with

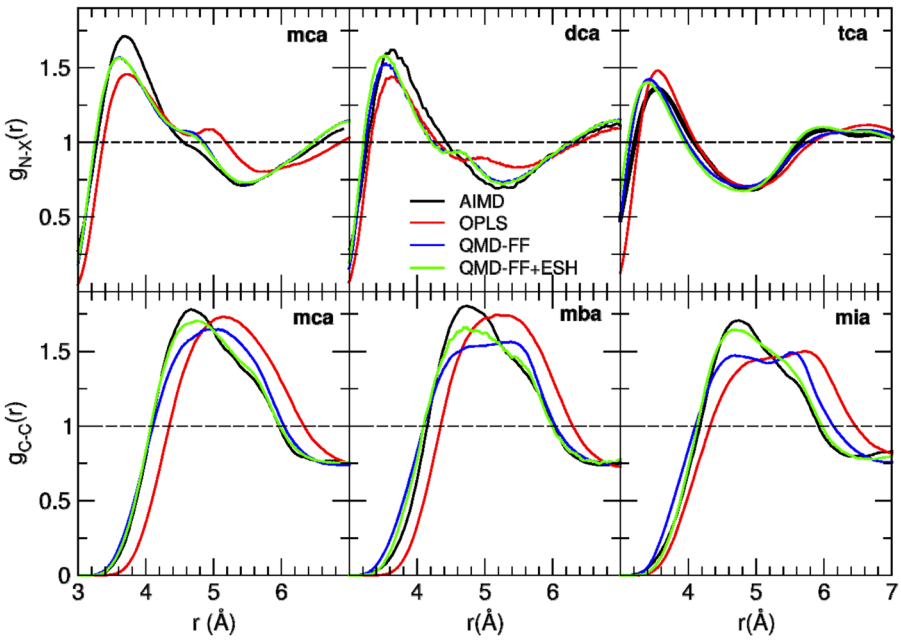

FIG. 7. Selected atomic pair correlation functions $g_{\alpha\beta}(r)$, computed over MD runs carried out with AIMD (black) or with different FF models (OPLS, QMD-FF, and QMD-FF+ESH in red, blue, and green, respectively). Top: $g_{N-X}(r)$, computed between the N atom of the cyano group and the chlorine atom(s) of neighboring molecules in the **mca**, **dca**, **tca** series. Bottom: $g_{C-C}(r)$, computed between the C atoms bearing the halogen in two neighboring molecules in the **mca**, **mba**, **mia** series.

the $g_{\alpha\beta}(r)$ analysis and compute the distribution functions, defined in Sec. III, to investigate the reciprocal orientation and disposition of molecular pairs within the first shell. A complete collection of such pair orientational functions can be found in the supplementary material. Here, for example, Fig. 8 displays the $\hat{p}(\theta_X)$ function computed over the di-substituted acetonitrile series, whereas Fig. 9 shows the effect of multiple iodine substitution on $p(d_{X\text{-}N})|_{\theta_X}$.

By looking at the top panels of Fig. 8, where the $\hat{p}(\theta_X)$ orientational distribution computed over the AIMD trajectories is displayed, it is evident that the XB has an increasing effect in going from chlorine to bromine and iodine. Indeed, while in **dca**, the number of pairs interacting, per unit surface, at a C-X···N angle close to 180° is similar to the one found at ~90°, with a slight decrease in intermediate values, for **dba** and **dia**, $\hat{p}(\theta_X)$ remarkably increases in

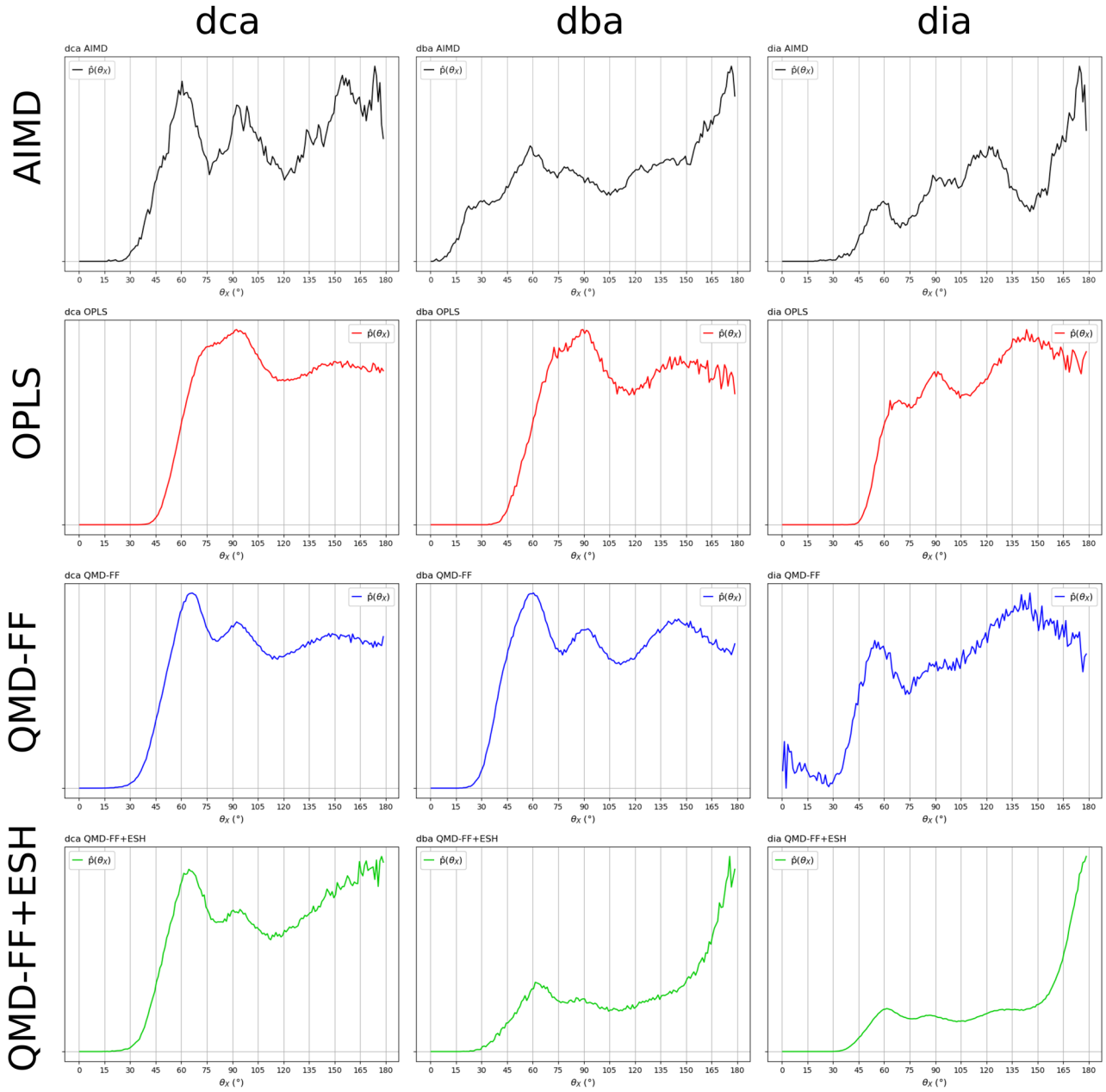

FIG. 8. Orientational $\hat{p}(\theta_X)$ distributions computed for **dca**, **dba**, and **dia** (in first, second, and third columns, respectively) on AIMD (first row) and MD (OPLS, QMD-FF, and QMD-FF+ESH in the second, third, and fourth rows) trajectories.

the [150°–180°] interval, a clear indication of the achieved head-on directionality of the XB. OPLS and QMD-FF models are able to reproduce this picture only partially: in both models, the maximum of the distribution function is placed around 90° for both **dca** and **dba**; then, it decreases to a constant plateau (**dca**) or with oscillations (**dba**) but without showing any maximum at 180°. For **dia**, two peaks appear at ~70° and 90°, whereas the maximum is instead found around 140°, while it decreases for more markedly head-on configurations. Upon ESH insertion, the AIMD picture is clearly better reproduced, in particular, for **dba** and **dia**, where the XB effects are stronger. In fact, as found for the reference trajectories, after slight oscillations at smaller angles, $\hat{p}(\theta_X)$ monotonically increases, reaching its maximum at 180°. For **dca**, conversely, the population at 180° seems slightly overestimated with respect

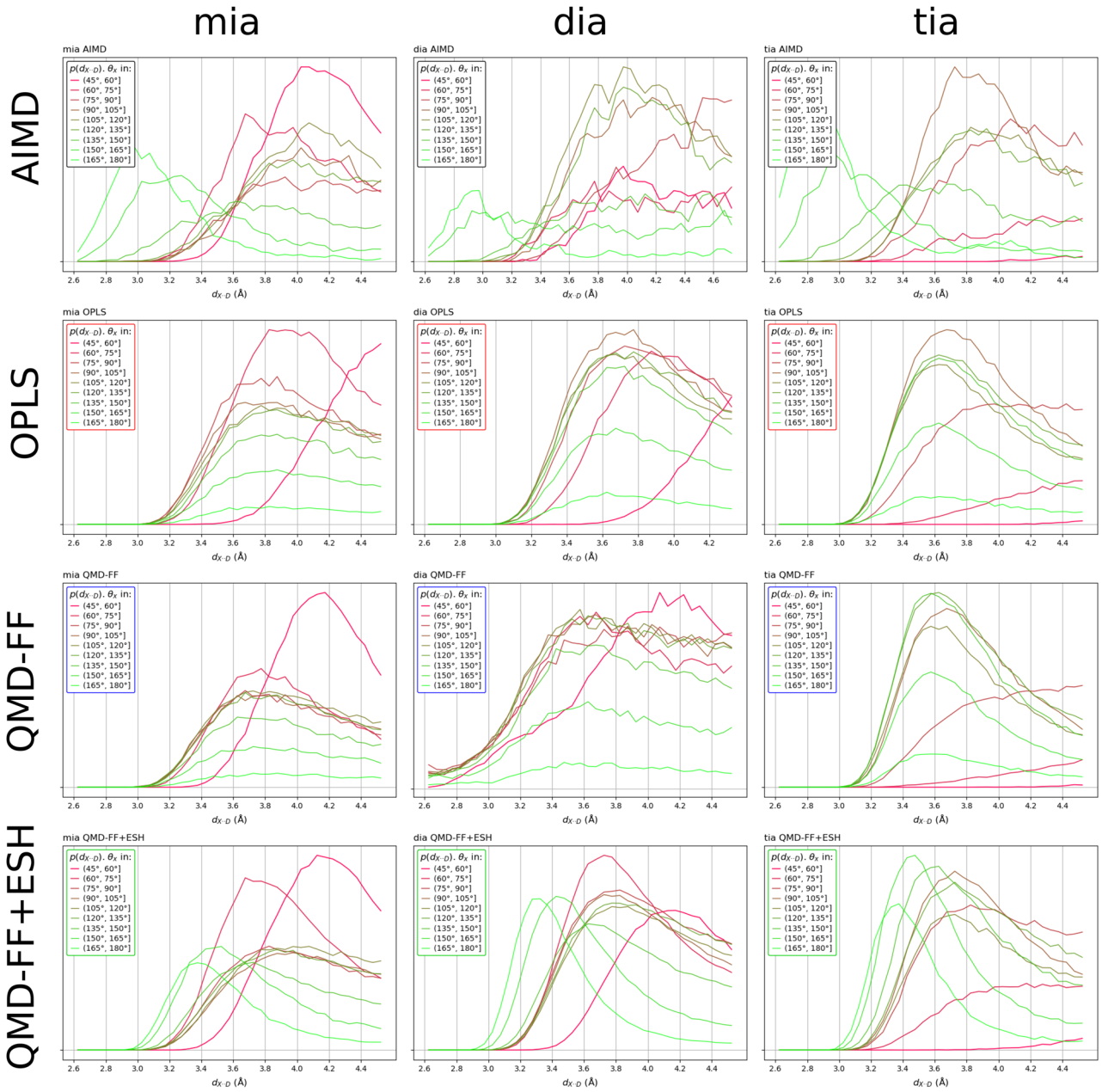

FIG. 9. Distribution of the X···N distance at different θ_X angles computed for **mia**, **dia**, and **tia** (in the first, second, and third columns, respectively) on AIMD (first row) and MD (OPLS, QMD-FF, and QMD-FF+ESH in the second, third, and fourth rows) trajectories.

to AIMD and the first peak is shifted to smaller angles, even if the intensity of the two peaks is similar, as found in the reference curves.

The ability of the QMD-FF+ESH model to better reproduce the microscopic structure described by reference AIMD runs is again confirmed in Fig. 9, where the distribution of the X···N distance at different θ_I angles is displayed for all models in the iodine substituted acetonitrile series. As expected for XBs, the C–I···N interaction exhibits a head-on character, as indicated by the clear increasing shift toward smaller distances of the $p(d_{I-N})$ peaks at increasing θ_I angle, meaning that the first neighboring molecules may closely approach the halogen atoms when the C–X···N angle is linear. It is worth recalling that for a purely isotropic distribution, $p(d_{X-N})$ should not depend on θ_X. As a matter of fact, this is what can be observed by looking at the OPLS and QMD-FF distributions for all species: even with the **tia** compound, the models without virtual sites are not able to find any difference in the position of the $p(d_{I-N})$ peaks for any θ_I larger than 90°. Furthermore, all distributions indicate the most probable X···N distance to be between 3.4 Å and 3.6 Å, while in AIMD, the nitrogen atom is found at less than 3 Å from an iodine atom, when the approaching angle θ_I is almost linear. Conversely, the AIMD description is again better reproduced when the ESH is explicitly accounted for. It is evident from the last row of Fig. 9 that the QMD-FF+ESH model well accounts for contact anisotropy, and the $p(d_{I-N})$ peaks at θ_I = 150°–180° are shifted toward smaller distance by ~0.3 Å. As appears from the complete analysis displayed in Figs. S29–S33 of the supplementary material, the agreement in the $p(d_{X-N})$ distribution between AIMD and QMD-FF+ESH is almost quantitative for most species, with the exception of iodine substituted compounds, where, consistent with the g(r) analysis, AIMD indications reveal, for almost linear approaching angles, peaks at contact distances lower than 3 Å, whereas the classical QMD-FF+ESH model only place them around 3.2 Å.

V. CONCLUSIONS

In this work, we explored the possibility of including virtual sites in an automated FF parameterization, based on the JOYCE/PICKY protocols, by introducing supplementary point charges in QMD-FFs with a consistent, robust, and general procedure, solely based on QM data, thus not relying on any experimental data. As a test case, we focused on target species, namely di-substituted halomethanes and substituted acetonitriles, whose intermolecular interactions are expected to involve weak and strong XBs respectively, and can therefore benefit from the introduction of supplementary point charges representing the ESHs. With this goal, a novel parameterization route was implemented in a development version of the PICKY code[78] and applied to all considered species. For the sake of completeness, it might be important to note that the current implementation, here applied to XB and ESH, can be straightforwardly applied to other systems, which are known to benefit from additional point charges or virtual sites, such as the lone pairs in Lewis donors and radicals.

The analysis of the resulting parameters in the present work suggests that the proposed simultaneous optimization of LJ and point charges has been successfully implemented in the QMD-FF protocol. In fact, the ESH is found, in agreement with the indications reported in the literature,[45] to yield satisfactory results when placed at a distance **d** between 1.6 Å and 1.8 Å from the halogen site, and a general recipe to assess its best value has been formulated. Once the distance **d** has been set, the QMD-FF+ESH can be built on top of a standard QMD-FF: the convergence is reached after few PICKY cycles, and the deviation with respect to the reference QM description was found to decrease in all cases due to a global change not only of the point charge distribution but also of the LJ parameters. Indeed, the variety of LJ and charge parameters confirms the higher specificity of the QMD-FF with respect to conventional, transferable FFs. Finally, at a difference with the tested FF models without ESH, the final point charges on the halogen atoms in QMD-FF+ESH seem to better represent the expected trends, both along the series of different halogens (Cl, Br, and I) and upon increasing the number of substituents.

The validation of the proposed protocol was further achieved by comparing the structural and energetic patterns, resulting from classical MD simulations carried out employing the parameterized QMD-FF+ESH models both in the gas and in the liquid phase, with the results achieved by employing empirical and QMD-FFs and with those delivered by high level QM techniques. A general good agreement was found, with the QMD-FF+ESH outperforming both standard and QMD-FF in reproducing the gas phase energetics and the structural patterns resulting from more accurate *ab initio* calculations. The main drawback of QMD-FF+ESH consists in a rather systematic overestimation of the bulk density. Yet, such a defect can be ascribed to the pure two-body nature of the reference QM data, rather than to the introduction of the ESH virtual site. In fact, the density overestimation was already found in previous PICKY applications[42,67,73,74] and traced back to the lack of three-body interactions, whose effect on bulk density was recently estimated by McDaniel and Schmidt to be on average about 5%,[100,101] on extended benchmark sets of small molecules. On the one hand, further implementations of many-body interaction terms within the PICKY procedure could allow for a more accurate prediction of the bulk density, but on the other hand, the computational burden required to account for such contributions in MD simulations undermines the possibility to extend the approach to large molecules and complex systems.

The second goal of the present work consisted in assessing the performances of the parameterized QMD-FF+ESH in describing the XB energetic and structural features. The interaction energy of hundreds of dimers, computed at the DFT level, previously[42] validated vs CCSD(T) reference calculations, is very well reproduced by QMD-FF+ESH, with an average standard deviation of less than 2.5 kJ/mol, thus better than standard QMD-FF (~3 kJ/mol) and OPLS (>6 kJ/mol). More importantly, characteristic XB features, such as the X···D contact distance and the C–X···D approaching angle, are accounted for upon ESH introduction, at a difference with the other tested FFs. The largest disagreement in the liquid phase structural patterns was found for the iodine substituted acetonitriles, where the correlation functions computed on the reference AIMD trajectories show the insurgence of a first I···I neighbor peak at small distances (less than 3 Å), whose intensity increases with the increase in the number of iodine substituents, which does not appear in any of the investigated FFs. This issue calls for further investigation and might be connected to the well known deficiencies of the LJ model functions to correctly represent the repulsive branch of the interaction curves.

The last aim of the present work was to assess for which species the introduction of ESH might be requisite or if a simple re-parameterization of the LJ parameters, as in standard QMD-FF, is sufficient. In this framework, the analysis of the QM references (dimer IPESs and AIMD structural analysis) revealed that some typical XB patterns already appear in the di-substituted halomethanes and are not accounted for neither by the standard OPLS nor by QMD-FF parameterization without ESH, even though in this latter case, some improvements with respect to general purpose FFs are already visible. In summary, the present results suggest to adopt the ESH strategy in the presence of halogen atoms, even when a strong donor (as the cyano group) is missing.

Besides performing a specific QMD-FF+ESH parameterization, an alternative solution to address XB features can stand in adopting very recent transferable FFs, such as the OPLS3,[59] especially when the target molecule is similar to the ones included in the original training set. Nonetheless, given the importance of XB in biomolecular processes, materials, and drug design, the possibility of parameterizing accurate and specific FFs without resorting to any experimental data is particularly appealing when investigating novel species, not yet synthesized and not similar to the ones included in the transferable FF training sets. In fact, this work can be considered a further step in the long-lasting effort to develop reliable and accurate FFs based solely on QM data to be used as a valuable tool in *in silico* design of advanced materials and complex systems. To this purpose, several points call for further investigations and development, some of which are currently in progress in our laboratory. First, the whole QMD-FF procedure (with or without virtual sites) should be extended to larger and more complex molecules, developing new protocols to maintain affordable computational costs, in particular, in the calculation of the QM dimer interaction energy. Next, the possibility of building QMD-FF for hetero-dimers should be explored, as very recently proposed by Belletti *et al.*,[102] thus paving the way to applications to mixtures, solutions, and, possibly, inhomogeneous systems.

SUPPLEMENTARY MATERIAL

See the supplementary material for additional data and several details about the reported calculations not included in this paper: further details on the standard JOYCE and PICKY parameterization procedures, convergence of the QMD-FF+ESH parameterizations, complete list of QMD-FF+ESH parameters for all investigated species, summary of computed thermodynamic properties for all compounds, additional atomic pair correlation functions, and distribution of geometrical descriptors.

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its supplementary material. Further data, as well as the development version of the PICKY code employed in this work, are available from the corresponding author upon reasonable request.

REFERENCES

[1] P. Hobza, R. Zahradník, and K. Müller-Dethlefs, "The world of non-covalent interactions: 2006," Collect. Czech. Chem. Commun. **71**, 443–531 (2006).
[2] S. Scheiner, *Noncovalent Forces* (Springer International Publishing, Switzerland, 2015).
[3] P. Hobza and J. Řezáč, "Introduction: Noncovalent interactions," Chem. Rev. **116**, 4911–4912 (2016).
[4] M. Juanes, R. T. Saragi, W. Caminati, and A. Lesarri, "The hydrogen bond and beyond: Perspectives for rotational investigations of non-covalent interactions," Chem. - Eur. J. **25**, 11402 (2019).
[5] D. Chandler, "Interfaces and the driving force of hydrophobic assembly," Nature **437**, 640–647 (2005).
[6] T. Steiner, "The hydrogen bond in the solid state," Angew. Chem., Int. Ed. **41**, 48–76 (2002).
[7] G. Gilli and P. Gilli, *The Nature of the Hydrogen Bond* (Oxford University Press, 2009).
[8] G. R. A. Desiraju, "Bond by any other name," Angew. Chem., Int. Ed. **50**, 52–59 (2011).
[9] V. A. Parsegian, *Van der Waals Forces: A Handbook for Biologists, Chemists, Engineers, and Physicists* (Cambridge University Press, 2005).
[10] P. Metrangolo, H. Neukirch, T. Pilati, and G. Resnati, "Halogen bonding based recognition processes: A world parallel to hydrogen bonding," Acc. Chem. Res. **38**, 386–395 (2005).
[11] P. Metrangolo and G. Resnati, *Halogen Bonding: Fundamentals and Applications; Structure and Bonding* (Springer Berlin Heidelberg, 2008).
[12] Y. Lu, T. Shi, Y. Wang, H. Yang, X. Yan, X. Luo, H. Jiang, and W. Zhu, "Halogen bonding—A novel interaction for rational drug design?," J. Med. Chem. **52**, 2854–2862 (2009).
[13] P. Politzer, J. S. Murray, and T. Clark, "Halogen bonding: An electrostatically-driven highly directional noncovalent interaction," Phys. Chem. Chem. Phys. **12**, 7748 (2010).
[14] M. Erdélyi, "Halogen bonding in solution," Chem. Soc. Rev. **41**, 3547 (2012).
[15] P. Politzer, J. S. Murray, and T. Clark, "Halogen bonding and other σ-hole interactions: A perspective," Phys. Chem. Chem. Phys. **15**, 11178 (2013).
[16] T. M. Beale, M. G. Chudzinski, M. G. Sarwar, and M. S. Taylor, "Halogen bonding in solution: Thermodynamics and applications," Chem. Soc. Rev. **42**, 1667–1680 (2013).
[17] M. H. Kolář and P. Hobza, "Computer modeling of halogen bonds and other σ-hole interactions," Chem. Rev. **116**, 5155–5187 (2016).
[18] T. Clark, "Halogen bonds and σ-holes," Faraday Discuss. **203**, 9–27 (2017).
[19] P. J. Costa, "The halogen bond: Nature and applications," Phys. Sci. Rev. **2**, 20170136 (2017).
[20] P. J. Costa, R. Nunes, and D. Vila-Viçosa, "Halogen bonding in halocarbon-protein complexes and computational tools for rational drug design," Expert Opin. Drug Discovery **14**, 805 (2019).
[21] L. D. A. Santos, I. G. Prandi, and T. C. Ramalho, "Could quantum mechanical properties be reflected on classical molecular dynamics? The case of halogenated organic compounds of biological interest," Front. Chem. **7**, 848 (2019).
[22] A. Bruckmann, M. Pena, and C. Bolm, "Organocatalysis through halogen-bond activation," Synlett **2008**, 900–902.
[23] F. Kniep, S. H. Jungbauer, Q. Zhang, S. M. Walter, S. Schindler, I. Schnapperelle, E. Herdtweck, and S. M. Huber, "Organocatalysis by neutral multidentate halogen-bond donors," Angew. Chem., Int. Ed. **52**, 7028–7032 (2013).
[24] P. Metrangolo and G. Resnati, "Halogen bonding: A paradigm in supramolecular chemistry," Chem. - Eur. J. **7**, 2511–2519 (2001).
[25] K. Rissanen, "Halogen bonded supramolecular complexes and networks," CrystEngComm **10**, 1107 (2008).
[26] Z. Xu, Z. Liu, T. Chen, T. Chen, Z. Wang, G. Tian, J. Shi, X. Wang, Y. Lu, X. Yan, G. Wang, H. Jiang, K. Chen, S. Wang, Y. Xu, J. Shen, and W. Zhu, "Utilization of halogen bond in lead optimization: A case study of rational design of potent phosphodiesterase type 5 (PDE5) inhibitors," J. Med. Chem. **54**, 5607–5611 (2011).

[27]R. Wilcken, M. O. Zimmermann, A. Lange, A. C. Joerger, and F. M. Boeckler, "Principles and applications of halogen bonding in medicinal chemistry and chemical biology," J. Med. Chem. **56**, 1363–1388 (2013).
[28]T. Clark, M. Hennemann, J. S. Murray, and P. Politzer, "Halogen bonding: The σ-hole," J. Mol. Model. **13**, 291–296 (2007).
[29]R. P. Feynman, "Forces in molecules," Phys. Rev. **56**, 340–343 (1939).
[30]P. Politzer, K. E. Riley, F. A. Bulat, and J. S. Murray, "Perspectives on halogen bonding and other σ-hole interactions: Lex parsimoniae (Occam's razor)," Comput. Theor. Chem. **998**, 2–8 (2012).
[31]A. El Kerdawy, J. S. Murray, P. Politzer, P. Bleiziffer, A. Heßelmann, A. Görling, and T. Clark, "Directional noncovalent interactions: Repulsion and dispersion," J. Chem. Theory Comput. **9**, 2264–2275 (2013).
[32]T. Clark, J. S. Murray, and P. Politzer, "The σ-hole coulombic interpretation of trihalide anion formation," ChemPhysChem **19**, 3044–3049 (2018).
[33]L. Brammer, "Halogen bonding, chalcogen bonding, pnictogen bonding, tetrel bonding: Origins, current status and discussion," Faraday Discuss. **203**, 485–507 (2017).
[34]T. Brinck, J. S. Murray, and P. Politzer, "Surface electrostatic potentials of halogenated methanes as indicators of directional intermolecular interactions," Int. J. Quantum Chem. **44**, 57–64 (1992).
[35]T. Clark, "σ-Holes," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **3**, 13–20 (2013).
[36]C. Esterhuysen, A. Heßelmann, and T. Clark, "Trifluoromethyl: An amphiphilic noncovalent bonding partner," ChemPhysChem **18**, 772–784 (2017).
[37]J. Řezáč, K. E. Riley, and P. Hobza, "Benchmark calculations of noncovalent interactions of halogenated molecules," J. Chem. Theory Comput. **8**, 4285–4292 (2012).
[38]S. Kozuch and J. M. L. Martin, "Halogen bonds: Benchmarks and theoretical analysis," J. Chem. Theory Comput. **9**, 1918–1931 (2013).
[39]C. B. Aakeröy, S. Alavi, L. Brammer, D. L. Bryce, T. Clark, J. E. Del Bene, A. J. Edwards, C. Esterhuysen, T. N. Guru Row, P. Kennepohl, A. C. Legon, G. O. Lloyd, J. S. Murray, W. T. Pennington, P. Politzer, K. E. Riley, S. V. Rosokha, S. Scheiner, S. Tsuzuki, and I. Vargas-Baca, "Computational approaches and sigma-hole interactions: General discussion," Faraday Discuss. **203**, 131–163 (2017).
[40]J. Rezáč and P. Hobza, "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. **9**, 2151–2155 (2013).
[41]J. Řezáč and P. Hobza, "Benchmark calculations of interaction energies in noncovalent complexes and their applications," Chem. Rev. **116**, 5038–5071 (2016).
[42]G. Prampolini, M. Campetella, N. De Mitri, P. R. Livotto, and I. Cacelli, "Systematic and automated development of quantum mechanically derived force fields: The challenging case of halogenated hydrocarbons," J. Chem. Theory Comput. **12**, 5525–5540 (2016).
[43]M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, 1987).
[44]D. Frenkel and B. Smith, *Understanding Molecular Simulations* (Academic Press, San Diego, 1996).
[45]M. Kolář and P. Hobza, "On extension of the current biomolecular empirical force field for the description of halogen bonds," J. Chem. Theory Comput. **8**, 1325–1333 (2012).
[46]W. L. Jorgensen and P. Schyman, "Treatment of halogen bonding in the OPLS-AA force field; application to potent anti-HIV agents," J. Chem. Theory Comput. **8**, 3895–3901 (2012).
[47]Z. Cournia, B. Allen, and W. Sherman, "Relative binding free energy calculations in drug discovery: Recent advances and practical considerations," J. Chem. Inf. Model. **57**, 2911–2937 (2017).
[48]L. A. Santos, E. F. F. da Cunha, and T. C. Ramalho, "Toward the classical description of halogen bonds: A quantum based generalized empirical potential for fluorine, chlorine, and bromine," J. Phys. Chem. A **121**, 2442–2451 (2017).
[49]R. Nunes, D. Vila-Viçosa, M. Machuqueiro, and P. J. Costa, "Biomolecular simulations of halogen bonds with a GROMOS force field," J. Chem. Theory Comput. **14**, 5383–5392 (2018).
[50]J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," J. Comput. Chem. **25**, 1157–1174 (2004).
[51]W. L. Jorgensen and J. Tirado-Rives, "Potential energy functions for atomic-level simulations of water and organic and biomolecular systems," Proc. Natl. Acad. Sci. U. S. A. **102**, 6665–6670 (2005).
[52]K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell, "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields," J. Comput. Chem. **31**, 671–690 (2009).
[53]C. Caleman, P. J. van Maaren, M. Hong, J. S. Hub, L. T. Costa, and D. van der Spoel, "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. **8**, 61–74 (2012).
[54]N. M. Fischer, P. J. van Maaren, J. C. Ditz, A. Yildirim, and D. van der Spoel, "Properties of organic liquids when simulated with long-range Lennard-Jones interactions," J. Chem. Theory Comput. **11**, 2938–2944 (2015).
[55]S. Nyburg, "Directional noncovalent interactions: Repulsion and dispersion," Acta Crystallogr., Sect. A **35**, 641–645 (1979).
[56]M. A. A. Ibrahim, "Molecular mechanical study of halogen bonding in drug discovery," J. Comput. Chem. **32**, 2564–2574 (2011).
[57]X. Mu, Q. Wang, L.-P. Wang, S. D. Fried, J.-P. Piquemal, K. N. Dalby, and P. Ren, "Modeling organochlorine compounds and the σ-hole effect using a polarizable multipole force field," J. Phys. Chem. B **118**, 6456–6465 (2014).
[58]A. N. S. Adluri, J. N. Murphy, T. Tozer, and C. N. Rowley, "Polarizable force field with a σ-hole for liquid and aqueous bromomethane," J. Phys. Chem. B **119**, 13422–13432 (2015).
[59]E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel, and R. A. Friesner, "OPLS3: A force field providing broad coverage of drug-like small molecules and proteins," J. Chem. Theory Comput. **12**, 281–296 (2016).
[60]B. Waldher, J. Kuta, S. Chen, N. Henson, and A. F. F. Clark, "ForceFit: A code to fit classical force fields to quantum mechanical potential energy surfaces," J. Comput. Chem. **31**, 2307–2316 (2010).
[61]O. Y. Yakovenko, Y. Y. Li, A. A. Oliferenko, G. M. Vashchenko, V. G. Bdzhola, and S. J. M. Jones, "*Ab initio* parameterization of YFF1, a universal force field for drug-design applications," J. Mol. Model. **18**, 663–673 (2012).
[62]J.-P. Piquemal and K. D. Jordan, "From quantum mechanics to force fields: New methodologies for the classical simulation of complex systems," Theor. Chem. Acc. **131**, 1207 (2012).
[63]N. De Mitri, S. Monti, G. Prampolini, and V. Barone, "Absorption and emission spectra of a flexible dye in solution: A computational time-dependent approach," J. Chem. Theory Comput. **9**, 4507–4516 (2013).
[64]S. Grimme, "A general quantum mechanically derived force field (QMDFF) for molecules and condensed phase simulations," J. Chem. Theory Comput. **10**, 4497–4514 (2014).
[65]D. J. Cole, J. Z. Vilseck, J. Tirado-Rives, M. C. Payne, and W. L. Jorgensen, "Biomolecular force field parameterization via atoms-in-molecule electron density partitioning," J. Chem. Theory Comput. **12**, 2312–2323 (2016).
[66]S. Vandenbrande, M. Waroquier, V. Van Speybroeck, and T. Verstraelen, "The monomer electron density force field (MEDFF): A physically inspired model for noncovalent interactions," J. Chem. Theory Comput. **13**, 161–179 (2017).
[67]L. Greff Da Silveira, M. Jacobs, G. Prampolini, P. R. Livotto, and I. Cacelli, "Development and validation of quantum mechanically derived force-fields: Thermodynamic, structural, and vibrational properties of aromatic heterocycles," J. Chem. Theory Comput. **14**, 4884–4900 (2018).
[68]K. Claridge and A. Troisi, "Developing consistent molecular dynamics force fields for biological chromophores via force matching," J. Phys. Chem. B **123**, 428–438 (2019).
[69]J. T. Horton, A. E. A. Allen, L. S. Dodda, and D. J. Cole, "QUBEKit: Automating the derivation of force field parameters from quantum mechanics," J. Chem. Inf. Model. **59**, 1366–1381 (2019).

[70]I. Cacelli, G. Cinacchi, G. Prampolini, and A. Tani, "Computer simulation of solid and liquid benzene with an atomistic interaction potential derived from *ab initio* calculations," J. Am. Chem. Soc. **126**, 14278–14286 (2004).
[71]I. Cacelli and G. Prampolini, "Parametrization and validation of intramolecular force fields derived from DFT calculations," J. Chem. Theory Comput. **3**, 1803–1817 (2007).
[72]I. Cacelli, C. F. Lami, and G. Prampolini, "Force-field modeling through quantum mechanical calculations: Molecular dynamics simulations of a nematogenic molecule in its condensed phases," J. Comput. Chem. **30**, 366–378 (2009).
[73]I. Cacelli, A. Cimoli, P. R. Livotto, and G. Prampolini, "An automated approach for the parameterization of accurate intermolecular force-fields: Pyridine as a case study," J. Comput. Chem. **33**, 1055 (2012).
[74]G. Prampolini, P. R. Livotto, and I. Cacelli, "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. **11**, 5182–5196 (2015).
[75]V. Barone, I. Cacelli, N. De Mitri, D. Licari, S. Monti, and G. Prampolini, "Joyce and Ulysses: Integrated and user-friendly tools for the parameterization of intramolecular force fields from quantum mechanical data," Phys. Chem. Chem. Phys. **15**, 3736–3751 (2013).
[76]J. Cerezo, G. Prampolini, and I. Cacelli, "Developing accurate intramolecular force fields for conjugated systems through explicit coupling terms," Theor. Chem. Acc. **137**, 80 (2018).
[77]I. Cacelli, J. Cerezo, N. De Mitri, and G. Prampolini, "Joyce2.10, a Fortran 77 code for intra-molecular force field parameterization," available at: http://www.pi.iccom.cnr.it/joyce, last consulted, April 2020.
[78]G. Prampolini and I. Cacelli, "Picky2.5, a Fortran 77 code for inter-molecular force field parameterization," available at: http://www.pi.iccom.cnr.it/picky, last consulted, April 2020.
[79]M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, P. V. Parandekar, N. J. Mayhall, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fo, GAUSSIAN 09, Revision C.01, Gaussian, Inc., Wallingford, CT, 2009.
[80]S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," J. Chem. Phys. **132**, 154104-1–154104-19 (2010).
[81]S. Grimme, S. Ehrlich, and L. Goerigk, "Effect of the damping function in dispersion corrected density functional theory," J. Comput. Chem. **32**, 1456–1465 (2011).
[82]S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, "GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit," Bioinformatics **29**, 845–854 (2013).
[83]G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," J. Chem. Phys. **126**, 014101 (2007).
[84]M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," J. Appl. Phys. **52**, 7182–7190 (1981).
[85]A. V. Marenich, S. V. Jerome, C. J. Cramer, and D. G. Truhlar, "Charge model 5: An extension of Hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases," J. Chem. Theory Comput. **8**, 527–541 (2012).
[86]J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele, "cp2k: Atomistic simulations of condensed matter systems," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **4**, 15–25 (2014).
[87]J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, "Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach," Comput. Phys. Commun. **167**, 103–128 (2005).
[88]J. VandeVondele and J. Hutter, "An efficient orbital transformation method for electronic structure calculations," J. Chem. Phys. **118**, 4365–4369 (2003).
[89]P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," Phys. Rev. **136**, B864 (1964).
[90]W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," Phys. Rev. **140**, A1133 (1965).
[91]J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," Phys. Rev. Lett. **77**, 3865 (1996).
[92]J. VandeVondele and J. Hutter, "Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases," J. Chem. Phys. **127**, 114105 (2007).
[93]S. Goedecker, M. Teter, and J. Hutter, "Separable dual-space Gaussian pseudopotentials," Phys. Rev. B **54**, 1703 (1996).
[94]C. Hartwigsen, S. Goedecker, and J. Hutter, "Relativistic separable dual-space Gaussian pseudopotentials from H to Rn," Phys. Rev. B **58**, 3641 (1998).
[95]L. Tanzi, F. Ramondo, R. Caminiti, M. Campetella, A. Di Luca, and L. Gontrani, "Structural studies on choline-carboxylate bio-ionic liquids by x-ray scattering and molecular dynamics," J. Chem. Phys. **143**, 114506 (2015).
[96]M. Campetella, M. Montagna, L. Gontrani, E. Scarpellini, and E. Bodo, "Unexpected proton mobility in the bulk phase of cholinium-based ionic liquids: New insights from theoretical calculations," Phys. Chem. Chem. Phys. **19**, 11869–11880 (2017).
[97]M. Campetella, A. Le Donne, M. Daniele, L. Gontrani, S. Lupi, E. Bodo, and F. Leonelli, "Hydrogen bonding features in cholinium-based protic ionic liquids from molecular dynamics simulations," J. Phys. Chem. B **122**, 2635–2645 (2018).
[98]M. Campetella, S. De Santis, R. Caminiti, P. Ballirano, C. Sadun, L. Tanzi, and L. Gontrani, "Is a medium-range order pre-peak possible for ionic liquids without an aliphatic chain?," RSC Adv. **5**, 50938–50941 (2015).
[99]L. Gontrani, R. Caminiti, U. Salma, and M. Campetella, "A structural and theoretical study of the alkylammonium nitrates forefather: Liquid methylammonium nitrate," Chem. Phys. Lett. **684**, 304–309 (2017).
[100]J. G. McDaniel and J. R. Schmidt, "First-principles many-body force fields from the gas phase to liquid: A "universal" approach," J. Phys. Chem. B **118**, 8042–8053 (2014).
[101]J. R. Schmidt, K. Yu, and J. G. McDaniel, "Transferable next-generation force fields from simple liquids to complex materials," Acc. Chem. Res. **48**, 548–556 (2015).
[102]G. Belletti, E. Schulte, E. Colombo, W. Schmickler, and P. Quaino, "Development of force fields for binary systems: Application to a dimethylsulfoxide (DMSO)–oxygen mixture," Chem. Phys. Lett. **735**, 136778 (2019).