

# Development and Validation of Quantum Mechanically Derived Force-Fields: Thermodynamic, Structural and Vibrational Properties of Aromatic Heterocycles

Leandro Greff da Silveira,<sup>a</sup> Matheus Jacobs,<sup>a,b,c</sup> Giacomo Prampolini,<sup>d,\*</sup>  
Paolo Roberto Livotto,<sup>a</sup> and Ivo Cacelli<sup>d,e</sup>

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

<sup>b</sup>*Institut für Physik, Humboldt-Universität zu Berlin,  
Newtonstr. 15, 12489, Berlin, Germany*

<sup>c</sup>*IRIS Adelsbrunn, Humboldt-Universität zu Berlin,  
Zum Großen Windkanal 6, 12489, Berlin, Germany*

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

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

June 5, 2018

---

\*Corresponding author: giacomo.prampolini@pi.iccom.cnr.it

## Abstract

A selection of several aromatic molecules, representative of the important class of heterocyclic compounds, has been considered for testing and validating an automated Force Field (FF) parameterization protocol, based only on Quantum Mechanical data. The parameterization is carried out separately for the intra- and inter-molecular contributions, employing respectively the JOYCE and PICKY software packages, previously implemented and refined in our research group. The whole approach is here automated and integrated with a computationally effective yet accurate method, devised very recently (*J. Chem. Theory. Comput.*, 2018, **14**, 543-556) to evaluate a large number of dimer interaction energies. The resulting quantum mechanically derived FFs are then used in extensive molecular dynamics simulations, in order to evaluate a number of thermodynamic, structural and dynamic properties of the heterocycle's gas and liquid phases. The comparison with the available experimental data is good and furnishes a validation of the presented approach, which can be confidently exploited for the design of novel and more complex materials.

# 1 Introduction

Many-body quantum mechanics (QM) is in principle the method of election for treating single molecules, clusters, solute-solvent systems as well as bulk phases formed by a large number of sub-units. In the latter case, however, it is clear that the current computational resources actually prevent accurate calculation at QM level, and one is forced to use cheaper methods,<sup>1,2</sup> based on classical physics, like Molecular Dynamics (MD) or Monte Carlo (MC). The main approximation of such methods involves a drastic simplification on the lowest eigenvalue of the electronic Hamiltonian, whose behavior with respect to the nuclear geometry is represented by an effective potential, called the Force Field (FF).<sup>1-3</sup> A FF consists in a collection of analytical functions of the positions of the nuclei, aimed to describe the energy of the simulated system, whose chemical identity is encoded into the FF parameters. Consequently, the reliability of the description achieved by classical simulations essentially relies on the FF quality.<sup>4-9</sup> Within this framework, an effective way to exploit both the accuracy of the QM techniques in describing molecular features and the efficiency of the classical MC and MD techniques, might consist in deriving the FF parameters from data retrieved through QM calculations.

As a matter of fact, following the pioneering work of Clementi's<sup>10-13</sup> and Linse's<sup>14,15</sup> groups, it is only in the past decade that, thanks to the massive increase of computational resources, there has been a renewed interest for quantum mechanically derived FFs (QMD-FFs), and several approaches have been proposed<sup>8,16-39</sup> and recently reviewed.<sup>40-42</sup> The several strategies that have been presented can be classified according to different criteria. First, these protocols concern either with the description of a single molecule flexibility (intramolecular FFs),<sup>16,17,20,21,24,25,29,31,38</sup> the interaction between molecular pairs (intermolecular FFs)<sup>18,19,27,27,30,32,34,36,37,39</sup> or both.<sup>8,22,23,26,28,33,35</sup> Next, the parameterization procedure can be applied to a particular class of molecules (aromatic, aliphatic, etc.), hence including a partial degree of transferability,<sup>27,30,39</sup> or specifically carried out on selected targets, abandoning the idea of transferability in favor of an increased accuracy.<sup>8,18-20,25,28,31-38</sup> Finally, the high level QM description can be exploited to incorporate into the QMD-FF a more detailed model, through the use of complex though computationally expensive functions, which allow for accounting, for instance, three body terms or

polarization effects.<sup>28,30,35,39</sup> Yet, in the case of simulations of large and complex systems, the QMD-FF can be conveniently built<sup>8,16,17,20,22,25,33,34,36–38</sup> with the standard simple potential functions (Lennard-Jones (LJ), point charges, harmonic potentials and Fourier series) implemented in all the most popular MD simulation engines.

Beside the details underlying the different procedures, most system specific QMD-FF show<sup>8,28,34,40–42</sup> a higher accuracy with respect to standard empirical FFs and have an intrinsic predictive capability,<sup>36</sup> as no experimental information is required for their parameterization. This latter feature is indeed appealing and certainly connected to the recent success of such approaches, especially in the search of novel materials or in the characterization of significant observables, hard to be measured experimentally. Despite these attractive features, the parameterization of QMD-FFs is a challenging and delicate process. For instance, three possible drawbacks, especially concerning with the intermolecular term, have been very recently pointed out by Vandenbrande *et al.*:<sup>39</sup>

- (a) due to the rather high number of parameters involved, the fitting procedure in such high-dimensional parameter space can turn out to be ill-conditioned;
- (b) in the hypothesis that the total interaction energy can be partitioned in different well-defined contributions (*e.g.* induction, dispersion, *etc.*), simple model functions entering the standard FF expressions as LJ or point charges are rarely able to separately account for the physical origin of the distinct energy contributions;
- (c) the computational cost associated with sampling the interaction energy potential surface (IPES) at a QM level is often limiting the extent of the QM database and the choice of the QM method, with a consequent loss of accuracy in the QMD-FF description.

In the past decade, a robust and completely automated QMD-FF parameterization protocol has been proposed and refined by our group.<sup>8,18,20,23,26</sup> The whole protocol is system-specific, *i.e.* the FF is parameterized on QM data purposely carried out on the system under investigation, hence achieving a higher accuracy with respect to most transferable FFs. The parameterization procedure is based on the classical partition of the system total energy in an intramolecular and intermolecular term,<sup>1,2</sup> and delivers a complete QMD-FF,<sup>8,33</sup> able to describe both single molecule flexibility and molecular interactions.

The intramolecular parameterization is carried out through the `JOYCE` code,<sup>20,25</sup> and has been tested previously in several applications.<sup>43–52</sup> Through this procedure, the intramolecular QMD-FF parameters are obtained by a least-square minimization, against few single-molecule QM data, specifically computed for the system under study. The `JOYCE` parameterization was shown to circumvent the two aforementioned drawbacks that affect also intramolecular FFs (namely, the ill-conditioning problem (a) and the computational cost (c), considering that point (b) essentially concerns with the intermolecular term only). In fact, regarding point (a), the linear fitting procedure underlying `JOYCE` parameterization is a robust method, with no appreciable ill-conditioning suffering.<sup>20,25</sup> On the same foot, concerning with point (c), the high quality/cost ratio of density functional theory (DFT) calculations for single molecule geometry optimizations can be exploited, maintaining the computational burden for intramolecular parameterization very limited, even for molecules of large dimensions.<sup>44,45,49–52</sup>

The intermolecular term of the QMD-FF is obtained through the `PICKY` software package.<sup>8,26</sup> In this case, the criticisms raised in Ref. [39] should be discussed point-by-point, paying particular attention to the last of such observations, concerning with the QM computational cost. As far as the ill-conditioning problem (a) is concerned, it was observed<sup>26</sup> that, notwithstanding different QMD-FF parameters could be obtained by different choices of the starting conditions in the non-linear fitting procedure, all the resulting QMD-FFs share the same adherence to the QM reference data and yielded the same level of quality in terms of both micro- and macroscopic observables. Owing to a certain degree of redundancy, the individual value of each parameter can only be determined with some uncertainties, but the final IPES is practically free from the detail of the calculation.<sup>26</sup> Turning to criticism (b), we point out that there is no limit to the complexity of the FF in the `PICKY` protocol, as for instance those proposed in the so-called physically motivated QMD-FFs,<sup>28,30,34,35,39</sup> to account for the different contributions to the intermolecular energy. However in this paper we prefer to adopt the standard LJ + charges form, both because they can be promptly employed (as already implemented in the most popular MD engines) and because computationally very effective. Indeed, an important perspective of the QMD-FF here presented concerns with the simulation of complex condensed phases

(*e.g.* liquid crystals,<sup>23</sup> embedded dyes,<sup>44</sup> bulk hetero-junctions,<sup>51</sup> biologically relevant systems,<sup>52</sup> *etc.*), where the QMD-FF computational convenience may become decisive.

The third issue discussed in Ref. [39] concerned with the computational cost of the required QM database. Recently, QMD-FFs derived according to our protocol were successfully applied in the simulation of liquid benzene,<sup>8</sup> the condensed phases of several halogenated organic compounds.<sup>33</sup> and organic triads suitable for organic photovoltaic cells.<sup>51</sup> In all these cases the whole intermolecular QM database was retrieved by density functional theory (DFT) calculations, performed over selected dimer geometries and accounting for dispersion through an empirical correction term (*e.g.* DFT-D<sup>53-55</sup>). Although DFT-D methods are in general computationally more convenient than post SCF ones, it was shown<sup>8,33</sup> that the choice of the functional has a remarkable impact on the resulting QMD-FF, thus requiring extensive benchmarks against high-level post-SCF data, with a consequent significant increase of the computational cost. In this work an alternative route will be attempted, abandoning the DFT-D description in favor of a computationally more feasible and possibly more accurate method: the recently proposed MP2<sup>mod</sup> procedure.<sup>56</sup> This method essentially consists in a MP2 calculation, carried out with a modified basis set, where the exponents of the polarization functions are optimized in order to mimic a reference high-level calculation, specifically performed on the target species.

The aim of this work is therefore twofold. The first goal is a further validation of the JOYCE/PICKY protocol in producing QMD-FFs, whose performances for the target system should be better or at least comparable to those obtained with standard FFs. The second goal is to explore the possibility of going beyond dispersion corrected DFT, resorting to more accurate postHF methods, in particular coupling the intermolecular parameterization PICKY protocol with the MP2<sup>mod</sup> procedure.

To accomplish these two goals, we have chosen as test case a challenging class of molecules: the aromatic heterocycles. Given the ubiquitous presence of aromatic compounds bearing heteroatoms in many biologically relevant processes, the possibility of building specific and accurate QMD-FFs for such species should in fact be carefully explored. Moreover, (heterocyclic) aromatic compounds constitute a significant challenge both in terms of the QM calculation of their interaction energy<sup>57-64</sup> and the FF repre-

sensation of their IPES.<sup>8,26,65–69</sup> As far as the former issue is concerned, we have recently shown<sup>56</sup> that very accurate results for aromatic heterocycles can be obtained resorting to the MP2<sup>mod</sup> method. In fact the standard deviation with respect to the so called gold standard of quantum chemistry,<sup>70,71</sup> *i.e.* CCSD(T)/CBS, resulted<sup>56</sup> to be less than 0.3 kcal/mol, in average. Furthermore, each MP2<sup>mod</sup> calculation was shown<sup>56</sup> to be more than one thousand time faster with respect to a CCSD(T)/CBS one. Turning to the second of the aforementioned challenges, *i.e.* regarding the FF quality for aromatic heterocycles, it has been for instance reported that standard FF performances in describing pyridine’s microscopic structure have to be improved, by refining the intermolecular model potential with the addition of supplementary interaction sites, thus enhancing the agreement with QM derived data.<sup>66–69</sup> A more general trend can be observed by looking at the <http://virtualchemistry.org> website, an extremely useful reference database for the development of new FFs, where the MD outcomes obtained<sup>4,9</sup> with three popular empirical transferable FFs (namely, GAFF,<sup>72</sup> OPLS<sup>3</sup> and CGenFF<sup>73</sup>) are reported in detail for many organic compounds. Therein, very different performances result for heterocycles, with respect to the average behavior, depending on the choice of the FF. For instance, for furan or pyridine, the error on bulk density and vaporization enthalpy, registered with different FFs, ranges from less than 1% up to 10%. The situation is definitively worst for those molecules which were not included in the original training sets considered for FF empirical parameterization, *e.g.* pyrazine, oxazole or isoxazole, where the reported error for the vaporization enthalpy with respect to the experimental values is always larger than 20%. Furthermore, there is no FF that seems to be systematically better than the others: CGenFF, for instance, yields the best performances for pyrimidine and thiophene, but it is the most inaccurate for pyridine and pyrrole. This last observation can become particularly critical if one has to make some "blind choice" of a FF to describe a molecular target with a novel structure, or bearing substituents, which were not included in the original empirical FF parameterization.

In this work, QMD-FFs will be set up for a set of selected compounds, whose structures are displayed in Figure 1, intended to be representative of the aromatic heterocycles class. The quality of the obtained QMD-FFs will be then assessed through the comparison of

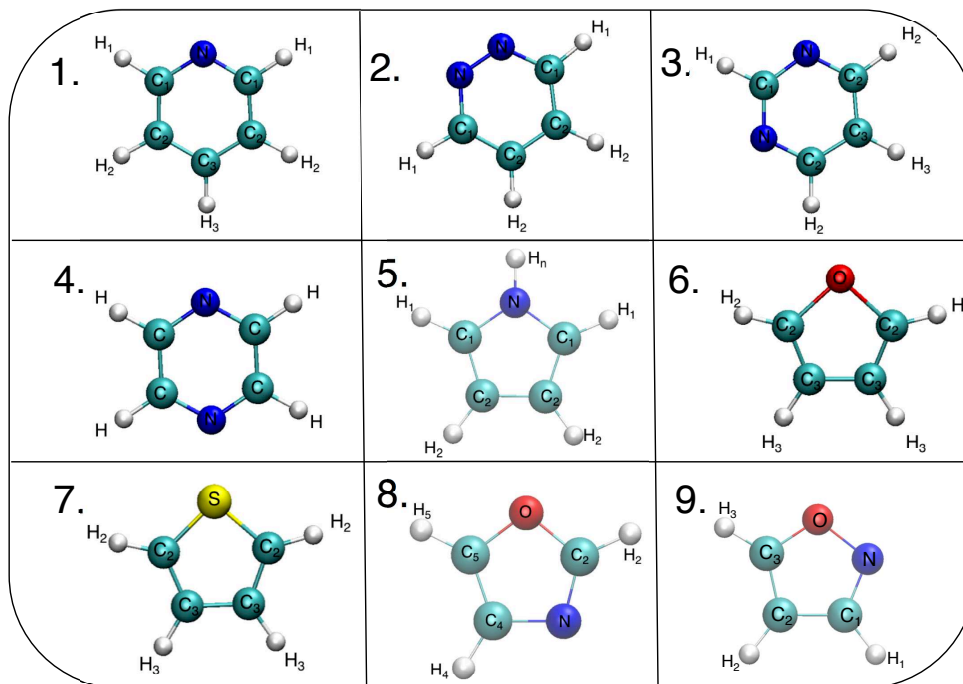


Figure 1: Sample hetero-cyclic aromatic molecules considered in this work: **1**: pyridine, **2**: pyridazine, **3**: pyrimidine, **4**: pyrazine, **5**: pyrrole, **6**: furan, **7**: thiophene, **8**: oxazole and **9**: isoxazole. In all pictures, Carbon atoms are displayed with cyan spheres, Nitrogen blue, Oxygen red and Hydrogen white. The different atom types employed in the FF parameterization are also shown within each panel.

the resulting macroscopic bulk properties, computed through lengthy MD simulations for each system, with the relevant experimental measures and, when possible, with the results reported in literature, yielded by empirical or specifically tuned FFs.

## 2 Methods

A specific QMD-FF is here developed for each of the selected benchmark molecules displayed in Figure 1, making use of the `JOYCE` and `PICKY` parameterization software, separately proposed in our group<sup>20,25,26</sup> and recently improved, automated and refined.<sup>8,33</sup> According to the procedure, the system-specific FF parameters are retrieved by minimizing selected objective functions, devised to optimize a classical based description against a reference representation obtained at QM level. To this end, the classical partition of



the total FF energy ( $E_{AB}^{FF_{tot}}$ ) of a pair of molecules  $A$  and  $B$  can be exploited:

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

where  $E_{A/B}^{FF_{intra}}$  and  $E_{AB}^{FF_{inter}}$  are the intramolecular and intermolecular FF contributions, for the  $A/B$  molecules and the  $AB$  interacting pair, respectively. The former drives the conformational behavior of each single molecule, and depends only on a collection of internal coordinates (*i.e.* bonds ( $\bar{b}$ ), angles ( $\bar{\theta}$ ) and dihedrals ( $\bar{\phi}$ )), while  $E_{AB}^{FF_{inter}}$  accounts for the interaction energy between the  $AB$  pair, and can hence be expressed as a function of the distances  $\bar{r}_{AB}$  between the atoms belonging to  $A$  and those of  $B$ .

## 2.1 Intramolecular Parameterization

For each molecule, the intramolecular part of the QMD-FF has the standard expression

$$E^{FF_{intra}} = E_{stretch} + E_{bend} + E_{tors} \quad (2)$$

where

$$E_{stretch} = \frac{1}{2} \sum_{\mu}^{N_{bonds}} k_{\mu}^s (b_{\mu} - b_{\mu}^0)^2 \quad (3)$$

$$E_{bend} = \frac{1}{2} \sum_{\mu}^{N_{angles}} k_{\mu}^b (\theta_{\mu} - \theta_{\mu}^0)^2 \quad (4)$$

$$E_{tors} = \frac{1}{2} \sum_{\mu}^{N_{dihedrals}} k_{\mu}^t (\phi_{\mu} - \phi_{\mu}^0)^2 \quad (5)$$

It is worth noticing that an harmonic form, rather than the more usual Fourier sum, is employed for the torsions in equation (5), in consideration of the stiff nature of the aromatic dihedrals.

The whole intramolecular parameterization is carried out with the JOYCE program,<sup>20,25</sup> freely available at <http://www.pi.iccom.cnr.it/joyce>. Following the standard protocol,<sup>20,25</sup> the equilibrium internal coordinates  $q^0$  (with  $q = \bar{b}, \bar{\theta}, \bar{\phi}$ ) entering equations (3)-(5) were taken from the QM equilibrium geometry of the isolated molecule. The force constants

$k$  were instead obtained at once for each molecule through the JOYCE code, minimizing the objective function:

$$I^{intra} = \sum_{K \leq L}^{3N-6} W_{KL} \left[ H_{KL} - \left( \frac{\partial^2 E^{FF_{intra}}}{\partial Q_K \partial Q_L} \right) \right]_0^2 \quad (6)$$

where  $Q_K$  is the  $K^{th}$  normal coordinate,  $H_{KL}$  is a QM Hessian matrix evaluated in the minimum energy geometry and  $W_{KL}$  are selected weights. Further details on JOYCE parameterization can be found in previous applications<sup>43-52</sup> or in the original papers.<sup>20,25</sup>

## 2.2 Intermolecular Parameterization

The intermolecular contribution,  $E^{FF_{inter}}$ , is expressed by sums of LJ potentials ( $E_{ij}^{LJ}$ ) and charge-charge ( $E_{ij}^{Coul}$ ) pairwise interactions:

$$E_{AB}^{FF_{inter}}(\bar{r}_{AB}) = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} [E_{ij}^{LJ}(r_{ij}) + E_{ij}^{Coul}(r_{ij})] \quad (7)$$

where  $N_{A/B}$  is the number of interaction sites of the  $A/B$  molecule and  $i/j$  is the  $i^{th}/j^{th}$  site of the molecule  $A/B$ .  $E_{ij}^{LJ}$  and  $E_{ij}^{Coul}$  take the standard expressions

$$E_{ij}^{LJ}(r_{ij}) = 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (8)$$

and

$$E_{ij}^{Coul}(r_{ij}) = \frac{q_i q_j}{r_{ij}} \quad (9)$$

The best parameters for both  $E_{ij}^{LJ}$  and  $E_{ij}^{Coul}$  are found with the PICKY approach,<sup>8,26,33</sup> namely through a non linear fitting, in order to minimize the objective function

$$I^{inter} = \frac{\sum_{k=1}^{N_{geom}} [(\Delta U_k^{inter} - E_k^{FF_{inter}})^2] e^{-\alpha \Delta U_k^{inter}}}{\sum_{k=1}^{N_{geom}} e^{-\alpha \Delta U_k^{inter}}} \quad (10)$$

where  $N_{geom}$  is the number of geometries considered for the target dimer,  $\alpha$  a Boltzmann-like weight and  $\Delta U_k^{inter}$  the QM intermolecular energy for dimer  $k$ .

According to PICKY's procedure, equation (10) is optimized iteratively, over a sample of  $N_{geom}$  dimer geometries, whose dimensions increase at every cycle. The convergence

is monitored by means of the quantity  $\Delta P$ , which measures the difference between two IPESs, obtained with the FF parameters of two consecutive cycles, and is defined as

$$\Delta P = \left( \frac{1}{N_{points}} \sum_i^{N_{points}} [(E_i^{FF_{inter}})_c - (E_i^{FF_{inter}})_{c-1}]^2 \right)^{\frac{1}{2}} \quad (11)$$

where  $(E_i^{FF_{inter}})_c$  is the intermolecular energy computed for dimer  $i$  with the QMD-FF parameters obtained at cycle  $c$ , and  $N_{points}$  is the number of dimer arrangements ( $> 10^6$ ) over which the difference is evaluated. Further details for PICKY parameterization can be found in the original papers.<sup>8,26</sup> As JOYCE, the PICKY package is also freely available at <http://www.pi.iccom.cnr.it/picky>.

### 2.3 QMD-FF assembling and validation

Concretely, the QMD-FF best parameters for each investigated compound were computed

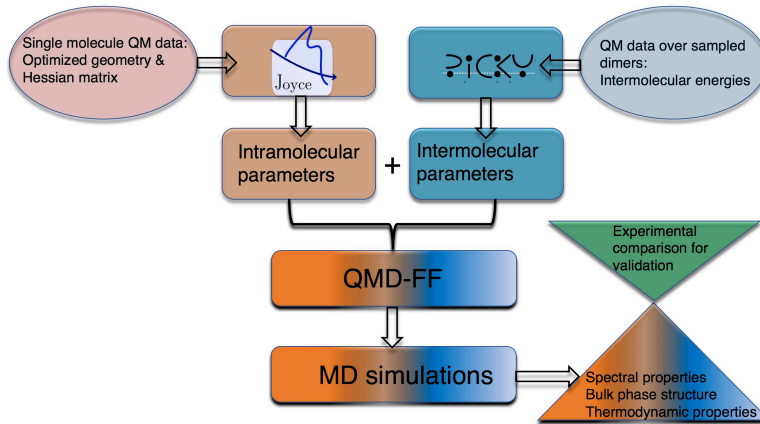


Figure 2: General scheme for QMD-FF derivation and validation employed in this work. Intra- and inter-molecular parameterization are evidenced with orange and blue panels respectively. Note that experimental data (green panel) are only employed for validation purposes, and do not enter the parameterization process.

according to the scheme displayed in Figure 2 and briefly commented in the following:

- i) For each compound,  $E^{FF_{intra}}$  is parameterized once and for all by means of the JOYCE protocol,<sup>20,25</sup>

- ii) The parameterization of  $E^{FF_{inter}}$  is performed following PICKY route I,<sup>8</sup> *i.e.* based on flexible MD simulations. Briefly, a selected number of dimer is extracted from an equilibrated MD configuration, their interaction energy computed at MP2<sup>mod</sup> level and used as reference data in equation (10), to retrieve a new set of FF parameters. The latter are then used for a new MD simulation, from where several additional dimers are again extracted. The final QMD-FF intermolecular parameters are obtained at the end of this iterative procedure, when the  $\Delta P$  convergence criterion (11) is met.
- iii) Extended MD simulations are performed with the resulting QMD-FF, on systems composed of 1000 molecules, in the NPT ensemble, in standard conditions. 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 obtained through the adoption of literature FFs and with their experimental counterparts.

## 3 Computational details

### 3.1 QM calculations

All QM calculations were performed with the GAUSSIAN09 package.<sup>74</sup> The QM data required for the intramolecular parameterizations, one geometry optimization and one vibrational frequency calculation, were obtained for each isolated monomer at DFT level, employing the standard B3LYP functional with the Dunning’s correlation consistent cc-pv-Dz basis set. *From such calculations, the molar volume and the isotropic polarizability were also retrieved for each considered monomer.*

As far as the intermolecular term is concerned, the reference QM dimer energies were computed by means of the recently proposed MP2<sup>mod</sup> method.<sup>56</sup> In fact, in view of the large number of dimer energies to be calculated, this approach was proven to be an excellent compromise between accuracy (less than 0.3 kcal/mol on average with respect to CCSD(T)/CBS) and computational cost (few minutes for each dimer). For all molecules except oxazole and isoxazole, all MP2<sup>mod</sup> calculations were carried out making use of the

specifically tuned basis sets reported in Ref. [56]. These basis sets were derived from a standard 6-31G\*\* basis set, in which the exponent of the polarization functions of each atom was optimized against a CCSD(T)/CBS database, specifically for each molecule. Since neither the oxazole nor the isoxazole molecule were included in the benchmark set considered in Ref. [56], no specific exponent is readily available to modify the original 6-31G\*\* basis set for these molecules. Here, to avoid the computational burden connected with the reference CCSD(T)/CBS calculations, an alternative route will be attempted, consisting in transferring all the exponents retrieved for furan, plus the exponent found in pyrrole for Nitrogen, to the atoms of oxazole and isoxazole. Concretely, the exponents employed in the modified 6-31G\*\* basis set adopted for oxazole and isoxazole are 0.36, 0.37, 0.21, for the  $d$  polarization functions on Oxygen, Nitrogen and Carbon, respectively, and 0.27 for the  $p$  functions on Hydrogen.

## 3.2 Molecular dynamics

All classical MD simulations were performed with the GROMACS5.1 engine,<sup>75</sup> on systems composed of 1000 molecules. During PICKY parameterization cycles,<sup>8,26</sup> each system was simulated for 500 ps, in the NPT ensemble, keeping the temperature and pressure constant through the Berendsen thermostat and barostat, with  $\tau_T$  and  $\tau_P$  to 0.1 and 5.0 ps, respectively. Once the best fit parameters were obtained, each QMD-FF was employed in a 525 to 50 ns production runs, carried out again in the NPT ensemble, but through velocity-rescale thermostat<sup>76</sup> and the Parrinello-Rahman scheme.<sup>77</sup> In all simulations, a cut-off 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. Unless otherwise stated, bond lengths are constrained at their equilibrium value using the LINCS algorithm, allowing for a 1 fs time step. Conversely, two additional 1 ns runs were carried out for each compound, in the NPT and NVE ensembles, saving the trajectories every 4 fs or 2 fs, respectively, as detailed hereafter. In these simulation, the constraints on the bond lengths were removed, and the time step was decreased to 0.1 fs.

### 3.3 Properties calculation

The MD trajectories were employed for the calculation of selected macroscopic properties. Unless otherwise stated, 5 ns NPT runs were used to monitor the following thermodynamic observables:

- bulk density  $\rho$ : computed by averaging the simulation box volume  $V$

$$\rho = \frac{M}{\langle V \rangle} \quad (12)$$

being  $M$  the total mass of the system.

- vaporization enthalpy  $\Delta H^{vap}$ : calculated as

$$\Delta H^{vap}(T) = H_g(T) - H_l(T) \simeq RT + \Delta U^{intra} - U_l^{inter}(T) \quad (13)$$

where  $H_g$  and  $H_l$  are the enthalpies of the gas and liquid phase,  $T$  is the temperature,  $R$  the gas constant,  $\Delta U^{intra}$  is the difference between the intra-molecular energy in the gas and liquid phase and  $U_l^{inter}$  the total intermolecular energy of the liquid.

- molecular dipole moment  $\bar{\mathbf{M}}$ : computed by double averaging the classical dipole moment of each monomer  $\bar{\mu}_i(t)$ , *i.e.*

$$\bar{\mathbf{M}} = \left\langle \frac{1}{N_{mol}} \sum_{i=1}^{N_{mol}} \bar{\mu}_i(t) \right\rangle \quad (14)$$

where  $N_{mol}$  is the total number of molecules composing the system and the  $\langle \rangle$  brackets stand for the time average.  $\bar{\mu}_i(t)$  is computed for each monomer  $i$  from the QMD-FF point charges entering equation (9) as

$$\bar{\mu}_i(t) = \sum_{\alpha=1}^{N_{atoms}} q_{\alpha} r_{i\alpha}^*(t) \quad (15)$$

where  $r_{i\alpha}^*(t)$  is the displacement of atom  $\alpha$  with respect to the center of mass of molecule  $i$  and  $N_{atoms}$  is the total number of atoms for each molecule.

- heat capacity at constant pressure  $c_P$ : determined (from the 1 ns NPT trajectories saved every 4 fs) through the two phase thermodynamic method,<sup>78-81</sup> implemented in GROMACS5.1 as described in detail in Ref. [4].

- volumetric thermal expansion coefficient  $\alpha_P$ : determined, as in previous work,<sup>8,33</sup> from the slope of the curve obtained by plotting  $\ln[\rho]$  as a function of  $T$  for the three 5 ns NPT runs, purposely carried out at different temperatures.
- isothermal compressibility  $\kappa_T$ : computed from the volume fluctuations as

$$\kappa_T = \frac{\langle \delta V^2 \rangle}{k_B T \langle V \rangle} \quad (16)$$

- *static dielectric constant  $\epsilon_0$ : obtained according to two different routes,<sup>82</sup> i.e. either as<sup>4</sup>*

$$\epsilon_0 = 1 + \frac{4\pi}{3} \frac{\langle \bar{M}^2 \rangle - \langle \bar{M} \rangle^2}{\epsilon_0 3V k_B T} \quad (17)$$

*or<sup>82-84</sup>*

$$\epsilon_0 = \epsilon_\infty + \frac{4\pi}{3} \frac{\langle \bar{M}^2 \rangle - \langle \bar{M} \rangle^2}{\epsilon_0 3V k_B T} \quad (18)$$

*In the above equations  $\bar{M}$  is the total dipole moment defined in equation (14),  $V$  is the simulation box volume,  $T$  the temperature and  $k_B$  the Boltzmann constant. In equation (18),  $\epsilon_\infty$  is estimated, as suggested in Ref. [<sup>82</sup>], from the Clausius-Mossotti relation*

$$\frac{\epsilon_\infty - 1}{\epsilon_\infty + 2} = \frac{4\pi\alpha}{3 \langle v \rangle} \quad (19)$$

*where for  $\alpha$  and  $v$ , were employed the isotropic polarizability and the molar volume, computed on each isolated monomer.*

The microscopic structure of each resulting liquid bulk phase was analyzed, again over the 5 ns trajectories, in terms of center of mass and atomic pair correlation functions, computed as

$$g_{\alpha\beta}(r) = \langle \delta(r - |r_i^\alpha - r_j^\beta|) \rangle ; \quad i \neq j \quad (20)$$

where the mean value  $\langle \dots \rangle$  is a double average, over all frames and molecules,  $r_k^\alpha$  is the position of the  $\alpha$  site in molecule  $k$ .

Finally, the vibrational frequencies were recovered, over the 1 ns NVE runs, from the classical dipole moment correlation function according<sup>85</sup> to

$$I(\nu) \propto Q(\nu)\nu \int_0^\infty e^{i2\pi\nu t} \langle \bar{\mu}(t+t_0)\bar{\mu}(t) \rangle dt \quad (21)$$

where  $\bar{\mu}(t)$  is computed according to equation (15) and the function  $Q(\nu)$  is a QM correction. Among the many different equations proposed to account for such correction,<sup>85–87</sup> following the suggestions of Ref. [88], the function here employed reads

$$Q(\nu) = \frac{\beta h\nu}{1 - e^{-\beta h\nu}} \quad (22)$$

with  $\beta = 1/k_B T$ , where  $T$  is the simulation temperature and  $k_B$  the Boltzmann constant.

## 4 Results

### 4.1 QMD-FF Parameterization

#### 4.1.1 Intramolecular term

The parameterization of the intramolecular potential was performed for all molecules displayed in Figure 1, employing the JOYCE code, according to the protocol outlined in the previous sections. In all cases, a similar strategy was followed. First, exploiting the harmonicity of the potentials entering the QMD-FF intramolecular expression (2), the only QM data required for each investigated molecule are its equilibrium geometry and Hessian matrix. Second, the only restrictions imposed on the atom-types were those dictated by symmetry. A complete list of the final parameters as well as the assignment of the different atom-types is reported in the Supporting Information (see Tables A1-A9).

In Table 1 some outcomes of the performed parameterizations are reported for the considered isolated molecules. As appears by looking at the last column, a similar standard deviation  $\sigma_J$ , less than 0.02 kJ/mol, was found in each case, regardless of the number of employed parameters. From a computational point of view, the main cost of the intramolecular parameterization resides in the QM calculations, which only takes few minutes on a standard workstation.

The resulting best-fit parameters (see Supporting Information) can be briefly commented to enlight the capability of the QMD-FF to reproduce even some subtle differ-



Compound	number of parameters	$\sigma_J$ (kJ/mol)
pyridine	42	0.015
pyridazine	36	0.015
pyrimidine	36	0.016
pyrazine	36	0.019
pyrrole	40	0.014
furan	34	0.013
thiophene	34	0.011
oxazole	31	0.014
isoxazole	29	0.017

Table 1: JOYCE parameterization overview. The standard deviation  $\sigma_J$  was obtained as  $\sqrt{I^{intra}}$ , where  $I^{intra}$  refers to equation (6). For each molecule, the CPU time required for the whole QM database was less than 15 minutes on a single Intel Xeon X5650 processor.

ences, found at QM level, among chemically similar molecules as these considered in the present work. For instance, for the C-C stretching in the investigated diazines, different force constant values were obtained, depending on both the position within the molecule (*e.g.*  $k_{C_1-C_2}^S = 3087 \text{ kJ mol}^{-1}\text{\AA}^{-2}$  and  $k_{C_2-C_3}^S = 3405 \text{ kJ mol}^{-1}\text{\AA}^{-2}$  in pyridine, see Table A1 in the Supporting Information) or the molecule bearing it (*e.g.*  $k_{C_2-C_3}^S = 3369 \text{ kJ mol}^{-1}\text{\AA}^{-2}$  and  $k_{C-C}^S = 2896 \text{ kJ mol}^{-1}\text{\AA}^{-2}$  in pyrimidine and pyrazine, Tables A3 and A4, respectively). On the same foot, the C-C stretching constants between different Carbons in furan and thiophene (see Tables A6 and A7 in the Supporting Information) reflect the different nature of such bonds resulting from the QM calculations: while  $k_{C_2-C_3}^S$  is  $3929 \text{ kJ mol}^{-1}\text{\AA}^{-2}$  for furan and  $3649 \text{ kJ mol}^{-1}\text{\AA}^{-2}$  for thiophene, the respective force constants for the  $C_3 - C_3$  stretching are  $2946$  and  $2731 \text{ kJ mol}^{-1}\text{\AA}^{-2}$ .

The adherence of the QMD-FF Hessian to the parent QM one, reflects in the very good agreement of the QM and FF vibrational normal modes. A detailed description of the accuracy of the predicted frequencies is displayed in Figure 3, again for furan and thiophene. Results of similar quality were obtained in all the considered cases. By looking at the bottom panel of Figure 3, the strong correlation between the vibrational frequencies computed at QM and QMD-FF level is apparent for both molecules. The agreement is particularly good for the higher frequencies, which correspond to the most localized normal modes (*CH* stretchings). In fact, as shown in the top panels, when the normal mode

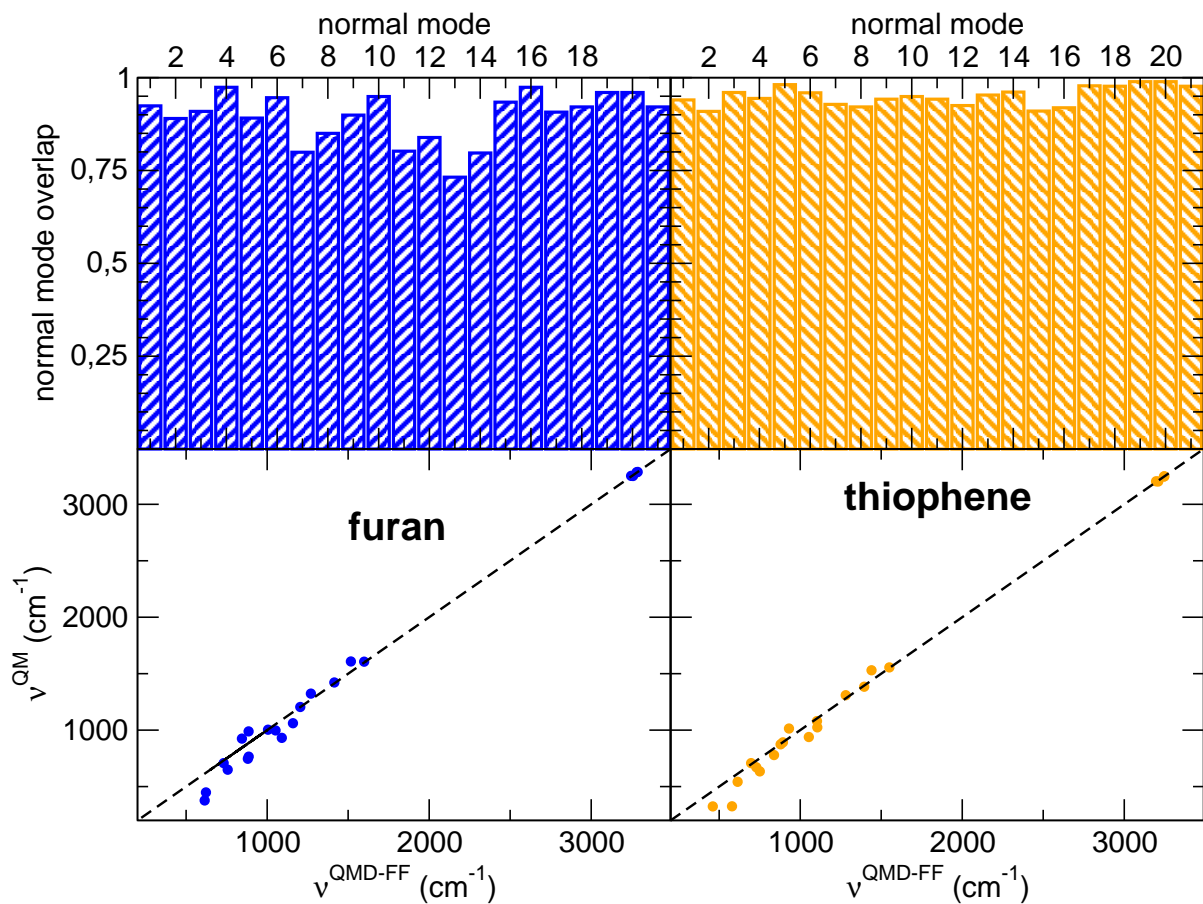


Figure 3: Comparison of JOYCE results obtained for furan (blue bars and symbols) and thiophene (orange bars and symbols). Top: overlap between QM and QMD-FF normal modes. Bottom: correlation plots between QM and QMD-FF computed vibrational frequencies.

is highly localized, the FF description based on internal coordinates is more accurate, and the overlap between the QM and FF normal modes is very close to unity. On the contrary, modes corresponding to lower frequencies, such as those involving the distortion of the planarity of the ring, are more delocalized in the internal coordinates set, and the vibrational description given by a diagonal FF becomes less accurate. As discussed by some of us in a recent paper,<sup>52</sup> this shortcoming could be circumvented by the addition of coupling FF terms, which depend on two or more internal coordinates, which were proven to be effective in improving the description of the vibrational frequencies. In other words,

the natural strategy to improve the description of the low frequency and delocalized modes is to abandon diagonal FF toward class II formulations, which explicitly include off-diagonal terms of the Hessian. Work in this direction is indeed in progress in our group, and a first step toward their complete implementation into the JOYCE package has been very recently reported.<sup>89</sup> Yet, since the present work is not focused on reaching spectroscopic accuracy, such coupling terms have been omitted, and the here employed FFs are in the standard diagonal form. Nonetheless, the intramolecular QMD-FFs obtained for furan and thiophene, are expected to be capable of accounting for the vibrational differences of these compounds, not only for the modes directly involving the oxygen or sulfur atoms, but even for the propagation of the heteroatom effects to the neighboring internal coordinates. For example, the two QMD-FFs succeed in reproducing the red shift of the *CH* stretching frequencies, in going from furan to thiophene.

#### 4.1.2 Intermolecular term

The parameterization of the intermolecular part of the FF was carried out, again specifically for each molecule, according to the PICKY procedure described in the previous sections. For all molecules except isoxazole, the initial fifty dimers were extracted from a trial configuration, obtained equilibrating a system of 1000 molecules in the NPT ensemble adopting the OPLS<sup>3</sup> description. Conversely, to test the impact of the starting configuration on the parameterization convergence, the initial isoxazole dimers were extracted from an equilibrated system of 1000 oxazole molecules, converted into isoxazole trivially exchanging the proper atom-type labels.

For each compound, in the first PICKY cycle, the intermolecular energy of all the fifty dimers sampled from the starting configuration was computed at QM level through the MP2<sup>mod</sup> method, as described in the Computational Details section. It might be worth noticing that the results achieved for oxazole and isoxazole, whose specific MP2<sup>mod</sup> basis sets was transferred rather than re-optimized, are particularly interesting in view of extending the PICKY procedure to larger molecules. Once all MP2<sup>mod</sup> interaction energies are retrieved, the system specific database is used to derive a first set of intermolecular parameters according to equation (10). Following the iterative PICKY procedure, such

parameters are used in a further MD run, to obtain a new equilibrated configuration, which is used as starting point for the subsequent cycle. All steps performed within each cycle (dimer sampling, GAUSSIAN09 input file generation for MP2<sup>mod</sup> calculations, extraction and storage of the QM energies incremental database, parameter fitting, calculation of the convergence index  $\Delta P$ , editing of the GROMACS5.1 files for MD runs) have been automated and coded into a development version of the PICKY code.

In Table 2 a summary of some relevant features of PICKY parameterization are given for all molecules. The parameterization computational cost, reported in the fourth

Compound	Cycles	Sampled dimers	CPU time (h)	$\Delta P$ (kJ/mol)	$\sigma_P$ (kJ/mol)
pyridine	5	450	7 (4)	0.16	1.26
pyridazine	6	550	8 (5)	0.41	1.77
pyrimidine	6	550	8 (5)	0.20	1.37
pyrazine	5	450	7 (4)	0.38	1.75
pyrrole	4	350	4 (2)	0.21	1.46
furan	4	350	5 (4)	0.30	0.99
thiophene	5	450	6 (4)	0.35	1.24
oxazole	6	550	5 (3)	0.49	1.45
isoxazole	6	550	5 (3)	0.27	1.00

Table 2: PICKY parameterization overview. CPU time refers to a Linux workstation with 16 Intel Xeon X5650 cores, the number in parenthesis indicates the time devoted to QM calculations. The convergence index  $\Delta P$  and the standard deviation  $\sigma_P$  were respectively obtained from equation (11) and as  $\sqrt{I^{inter}}$ , where  $I^{inter}$  refers to equation (10).

column on Table 2, is definitively larger than the one required for the intramolecular part of the FF, but never exceeds 10 hours for each compound. In particular, thanks to the convenience of the MP2<sup>mod</sup> approach, the time needed by the QM calculations is limited to few hours, despite the number of considered dimers ranges from 350 to 550 for each investigated species. All considered, the full parameterization for a target molecular species is achievable in one day on a standard workstation. Similarly to the intramolecular parameterization, the QMD-FF standard deviation ( $\sigma_P$ ) appears to be relatively independent on the system under investigation, as its value ranges from 0.99 kJ/mol for furan to 1.77 kJ/mol for pyridazine.

A more detailed description of  $\sigma_P$  and other relevant quantities along the parameterization cycles, together with a full list of the best-fit parameters obtained for each

compound, is reported in the Supporting Information (see Tables B1-B18). By looking

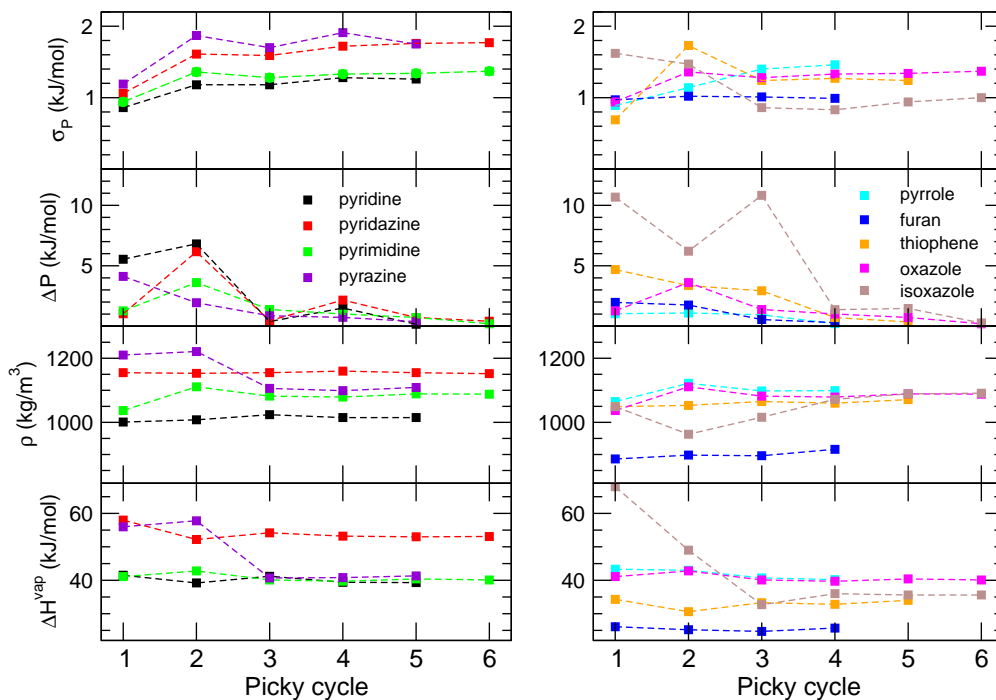


Figure 4: Monitored properties along the parameterization cycles: six-membered rings (left) and five-membered rings (right). From top to bottom: PICKY’s standard deviation,  $\sigma_P$ ; convergence index,  $\Delta P$ ; MD’s bulk density,  $\rho$ ; vaporization enthalpy,  $\Delta H^{vap}$ .

at the Tables reported therein, the capability to account for the chemical diversity of the different atom-types clearly emerges also for the intermolecular term of the QMD-FF. For instance, as far as the LJ parameters are concerned,  $\epsilon_N$  (see Tables B1-B18) is found to assume sensibly different values, depending on the molecule bearing the nitrogen atom, being  $\sigma_P \sim 2.7, 1.7, 3.4, 1.5, 2.1, 0.6, 0.5$  kJ/mol in pyridine, pyridazine, pyrimidine, pyrazine, pyrrole, oxazole and isoxazole, respectively). Similarly, the point charge  $q_O$  on oxygen atoms of different molecules ( $-0.073, -0.160$  and  $-0.057 e^-$  for furan, oxazole and isoxazole, see Tables B12, B16 and B18) reflects the different interaction patterns found at QM level. Yet, to not overwhelm the reader with excessive information, only a graphical summary of some of the contents of Tables B1 to B16 is here displayed in Figure 4.

To reach convergence (*i.e.*  $\Delta P < 0.5$  kJ/mol), all parameterizations required from four to six cycles, but it appears from Figure 4 that the  $\Delta P$  convergence index showed its larger variation during the first three cycles, meaning that only small changes on the IPES landscapes took place in the last cycles. This in turn reflects also on the monitored thermodynamic properties, the liquid bulk density ( $\rho$ ) and the vaporization enthalpy ( $\Delta H^{vap}$ ), which show significant variations only in the first parameterization cycles. This behavior also holds for isoxazole, despite the starting  $\rho$  and  $\Delta H^{vap}$  were far away from their converged values, due to the quasi-random nature of the starting configuration. Nonetheless, both observables appear to reach convergence after 4 cycles, in line with the other compounds, suggesting that the choice of the starting sample configuration should not affect the final intermolecular FF.

Finally, turning to the overall adherence of the QMD-FF description of the IPES with respect to the parent QM one, the PICKY standard deviation  $\sigma_P$  is displayed in the top panel of Figure 4. The final converged value was in all cases around 1.5 kJ/mol and obtained over  $\sim 550$  dimers. In conclusion, we are confident that a significant portion of the dimer IPES has been considered and that the main features of the QM description were tailored on the QMD-FF.

## 4.2 MD simulations with QMD-FFs

### 4.2.1 Thermodynamic properties

Once both the intramolecular and the intermolecular terms were parameterized for all the investigated compounds, the QMD-FFs were assembled and employed in extensive MD simulations to characterize the resulting bulk liquid phases. To this end, ~~5 ns~~, *unless otherwise stated, 25 ns* NPT runs were performed on systems of 1000 molecules, at 1 atm and 298 K for all samples except pyrazine, where the temperature was set to 340 K, since its experimental melting point is  $\sim 320$ K.<sup>90</sup> In Table 3, the bulk density  $\rho$  and the vaporization enthalpy  $\Delta H^{vap}$  of the resulting liquid phases are reported and compared to their experimental counterparts. The statistical uncertainty of the theoretical measures is around 5 kg/m<sup>3</sup> and 0.3 kJ/mol, respectively for the density and enthalpy. In general, a very good agreement with experiment is apparent, with the exception of

Compound	$\rho$ (kg/m <sup>3</sup> )			$\Delta H^{vap}$ (kJ/mol)		
	exp	QMD-FF	$\Delta$ (%)	exp	QMD-FF	$\Delta$ (%)
pyridine	979 <sup>91</sup>	1012	3.4	41.0 <sup>92</sup>	39.1	4.4
pyridazine	1103 <sup>93</sup>	1152	4.4	53.5 <sup>94</sup>	53.2	0.6
pyrimidine	1016 <sup>95</sup>	1088	7.1	49.8 <sup>93</sup> - 50.0 <sup>92</sup>	40.1	19.4
pyrazine <sup>(a)</sup>	1031 <sup>93</sup>	1059	2.7	40.5 <sup>92</sup>	38.8	4.7
pyrrole	965 <sup>95</sup>	1093	13.2	45.1 <sup>93,95</sup>	40.6	10.9
furan	931 <sup>93</sup>	920	1.2	27.1 <sup>93</sup>	26.2	3.3
thiophene	1059 <sup>93</sup>	1071	1.1	34.7 <sup>93</sup>	34.0	2.0
oxazole	1050 <sup>93</sup>	1068	1.7	31.1 <sup>91</sup> - 32.5 <sup>96</sup>	31.3	0.6
isoxazole	1078 <sup>93</sup>	1090	1.1	36.5 <sup>96</sup> - 37.2 <sup>97</sup>	35.6	3.8

<sup>(a)</sup> All data for pyrazine refer to T = 340 K

Table 3: Comparison between experimental and QMD-FF computed thermodynamic properties for the investigated samples: bulk density ( $\rho$ ) and vaporization enthalpy ( $\Delta H^{vap}$ ). Unless otherwise stated all data refer to 298 K.

pyrrole and pyrimidine, which show more marked deviations. In fact, the accuracy of the results reported in Table 3, is comparable or, as for the vaporization enthalpy of pyrazine, isoxazole and oxazole, better than most of the empirical FFs. The mean unsigned error, averaged over all investigated heterocycles, amounts to  $\sim 4\%$  and  $\sim 5\%$ , for density and enthalpy, respectively, and drops to  $\sim 4\%$  and  $\sim 3\%$  if pyrimidine and pyrrole are not included. Considering that the aforementioned properties are often the reference quantities for empirical FF tuning, the current results are encouraging. However the poor results of pyrrole and pyrimidine deserve some further investigations which are currently in progress in our laboratory. One possible explanation for pyrrole arise from the possibility that the  $MP2^{mod}$  basis set may be inadequate to represent with sufficient accuracy the hydrogen bond. This hypothesis is perhaps supported by the results of Ref [56] where the  $MP2^{mod}$  energies are in general less attractive than the CCSD(T)/CBS ones, even in the short range region. This could be consistent the underestimated density and vaporization enthalpy obtained in the present work.

As suggested by Caleman and co-workers,<sup>4</sup> supplementary thermodynamic properties should be used for a more sound validation of a FF, as for instance the specific heat at constant pressure ( $c_P$ ), the volumetric expansion coefficient ( $\alpha_P$ ) and the isothermal compressibility ( $\kappa_T$ ). All these quantities have been computed according to the details

Compound	$c_p$ (J K <sup>-1</sup> mol <sup>-1</sup> )		$\alpha_P$ (10 <sup>3</sup> K <sup>-1</sup> )		$\kappa_T$ (GPa <sup>-1</sup> )	
	exp	QMD-FF	exp	QMD-FF	exp	QMD-FF
pyridine	135 <sup>98</sup>	127	1.10 <sup>91</sup>	1.19	0.72 <sup>95</sup>	0.74
pyridazine	n.a.	126	n.a.	0.88	n.a.	0.44
pyrimidine	n.a.	120	0.89 <sup>95</sup>	1.17	n.a.	0.66
pyrazine <sup>(a)</sup>	129 <sup>92</sup>	135	n.a.	1.14	n.a.	1.14
pyrrole	128 <sup>91</sup>	128	0.87 <sup>91</sup>	1.18	0.65 <sup>95</sup>	1.04
furan	114 <sup>99</sup>	99	0.79 <sup>95</sup> - 1.49 <sup>91</sup>	1.72	1.49 <sup>93</sup>	1.64
thiophene	122 <sup>100</sup>	113	1.11 <sup>91</sup>	1.22	n.a.	0.89
oxazole	96 <sup>96</sup> - 112 <sup>91</sup>	92	1.0 <sup>95</sup>	1.50	n.a.	0.92
isoxazole	93 <sup>96</sup> - 108 <sup>97</sup>	100	n.a.	1.50	n.a.	0.76

<sup>(a)</sup> All data for pyrazine refer to T = 340 K

Table 4: Comparison between experimental and QMD-FF computed thermodynamic properties for the investigated systems: specific heat at constant pressure ( $c_p$ ), volumetric expansion coefficient ( $\alpha_P$ ) and isothermal compressibility ( $\kappa_T$ ). Unless otherwise stated all data refer to 298 K.

given in the previous section, and reported in Table 4, where they are compared to the available experimental measures. Regarding the specific heat, where many experimental data are available, an overall good agreement appears, in particular for the five membered rings. The computed volumetric expansion coefficients are also in line with the experimental data, although a slight systematic overestimation ( $\sim 0.3 \cdot 10^{-3} \text{ K}^{-1}$ ) is found for all molecules. Finally, despite the scarce availability of experimental isothermal compressibilities prevents any statistical interpretation, a good agreement is again registered for pyridine and furan, while a more significant overestimation ( $0.4 \text{ GPa}^{-1}$ ) was found for pyrrole.

A more systematic evaluation of the QMD-FF performances in predicting the thermodynamic properties of the condensed bulk phases can be gained by looking at Figure 5, where the correlation plots between the experimental and computed density, vaporization enthalpy and specific heat are displayed. By looking at the linear fits of each set of data, it appears as the  $\Delta H^{vap}$  and  $c_p$  slopes are close to unity, being the correlation coefficients 0.93 and 0.92, respectively. A somewhat worse performance is registered for the bulk density  $\rho$ , where the computed correlation coefficient is 0.77. Yet, if pyrrole is excluded from the sample, a better fit is obtained, and the correlation coefficient increases to 0.93. A similar evaluation of some popular empirical FFs can be done resorting to the



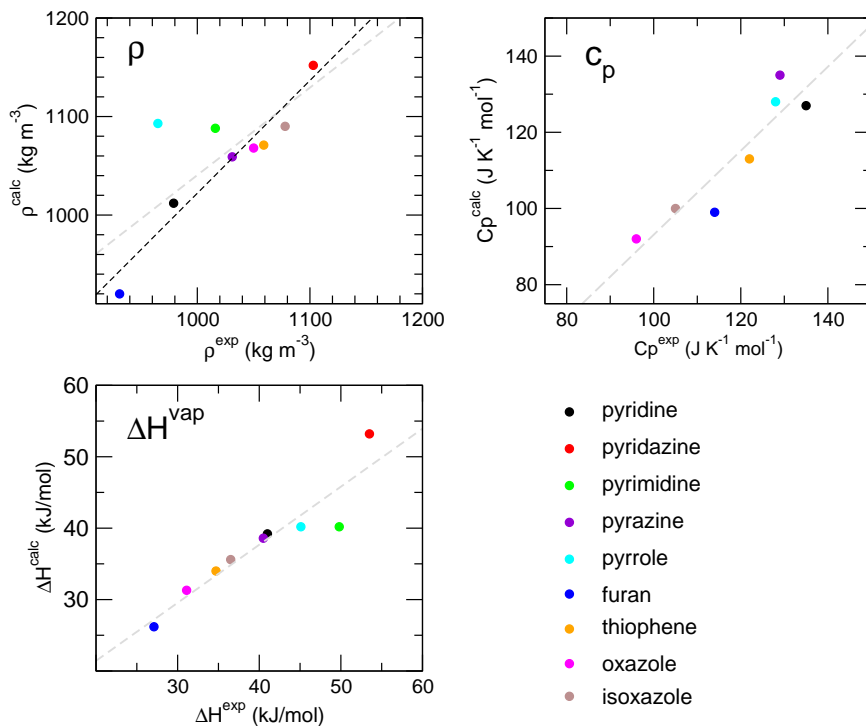


Figure 5: Correlation plots between experimental and QMD-FF computed values for bulk density ( $\rho$ ), vaporization enthalpy ( $\Delta H^{vap}$ ) and specific heat at constant pressure ( $C_p$ ). Linear regression are evidenced by dashed lines. In the first panel, the regressions obtained with or without considering pyrrole (see text) are shown by a black and a grey line, respectively.

<http://virtualchemistry.org> website. For instance, if the sample is limited to the selection of heterocyclic molecules considered therein, the correlation coefficients for  $\rho/\Delta H^{vap}$  (no data was available for  $c_p$ ) are 0.98/0.90, 0.97/0.85 and 0.98/0.96 for OPLS,<sup>3</sup> GAFF<sup>72</sup> and CgenFF,<sup>73</sup> respectively. In summary, the thermodynamic picture yielded by QMD-FF is reliable, and the results are on average in good agreement with the experimental data. It is worth recalling that although QMD-FFs were obtained without resorting to any experimental information, they yield performances comparable to empirical FFs, where instead the experimental values of  $\rho$  and  $\Delta H^{vap}$  were employed in the tuning procedure.

#### 4.2.2 Liquid phase structure

Beside reproducing the macroscopic thermodynamic behavior of the liquid phase, it is of fundamental importance that QMD-FFs are able to account also for the microscopic

structure. In Figure 6, for instance, the pair correlation function  $g(r)$  between pyridine’s centers of mass, is displayed together with its experimental counterpart, derived from recent neutron scattering experiments.<sup>101</sup> An almost quantitative agreement appears be-

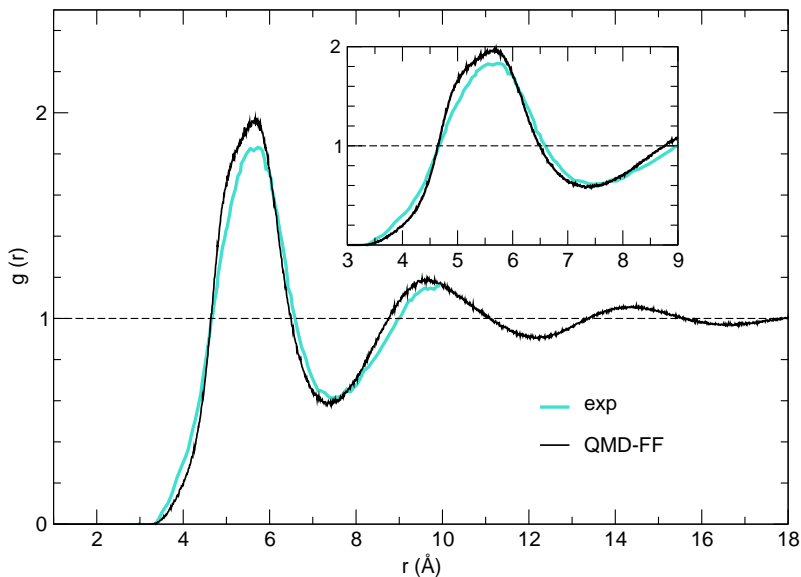


Figure 6: Comparison between pyridine’s experimental<sup>101</sup> (turquoise) and QMD-FF computed (black) pair correlation functions  $g(r)$ .

tween the two curves, both in the peak positions and in their shapes: the first neighbor peak is located by both signals near 5.6 Å, while a second shell appears at  $\sim 9.8$  Å. Furthermore, the computed results suggests the existence of a long range structure, which extends up to more than 15 Å.

The atomistic detail granted by MD simulations, allows for a deeper insight into the patterns of the microscopic intermolecular structure. In particular, the atom-atom pair correlation functions can help to elucidate the relative orientation of two neighboring molecules. In this framework, pyridine is a challenging benchmark, and several authors<sup>66–69</sup> suggested that the inclusion of additional interaction sites (*e.g.* supplementary point charges to mimic Nitrogen lone pair) is mandatory to account for the correct structure. Yet, it should be noted that the intermolecular energy registered in MD sim-

ulation of pyridine is, according to most FFs, made of  $\sim 80\%$  LJ contribution. In such cases, as recently noted by Khan and co-workers,<sup>37</sup> the modification of the LJ parameters can be more effective than tuning the point charges in improving the the FF quality. It is thus interesting to ascertain if the same accuracy found for the pyridine’s models with additional charges, can be obtained by a simple re-parameterization of the standard atomistic description, given by the present QMD-FF. To this end, the Nitrogen-Nitrogen and Nitrogen-Hydrogen atomic pair correlation functions are displayed in Figure 7 and C1 in the Supporting Information, respectively, for both the present QMD-FF and two of the most recently proposed models that make use of virtual sites.<sup>68,69</sup> In all cases, an almost

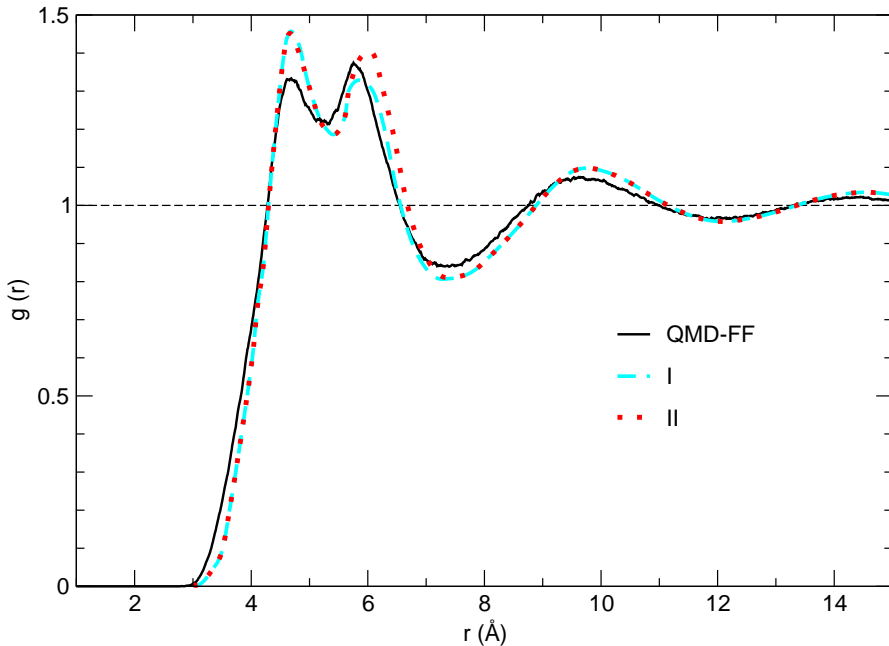


Figure 7: Atomic pair correlation functions  $g_{NN}(r)$  between pyridine’s nitrogen atoms, computed through the present QMD-FF (black line), or obtained through models which include virtual sites, as the pyridine model (**I** reported by Macchiagodena *et al.*<sup>69</sup> (cyan dashed line) or the recently proposed OPLS3 ((**II**, red dotted line).<sup>68,69</sup>

quantitative agreement among the three models is found, for both the peaks position and their width. In particular, as displayed in Figure 7, the characteristic<sup>66,68,69</sup> double peak

in  $g_{NN}(r)$  at  $\sim 4.9 \text{ \AA}$  and  $\sim 5.9 \text{ \AA}$ , in the first neighbor shell region, clearly appears in all considered models. The main differences among them are found in the relative intensity of the  $g_{NN}(r)$  double peak (Figure 7) and, to a lesser extent, in the first peak of the  $g_{NH}(r)$  functions, displayed in the Supporting Information (Figure C1). According to both Ref. [66] and [69], the  $g_{NN}(r)$  double peak is due to two classes of neighboring dimer arrangements, differing by an antiparallel (at  $4.9 \text{ \AA}$ ) and a parallel (at  $5.9 \text{ \AA}$ ) alignment of the molecular dipoles. In both conformations, the most important source of interaction is found to stem essentially from  $CH \cdots N$  hydrogen bond interaction, complemented with  $CH \cdots \pi$  and stacking forces, where the molecular dipoles are antiparallel (at  $4.9 \text{ \AA}$ ) or parallel (at  $5.9 \text{ \AA}$ ). On the same foot, the importance of the  $CH \cdots N$  network is also suggested by the  $g_{NH}(r)$  functions, which all show (see Figure C1) a rather wide peak at significantly ( $< 3 \text{ \AA}$ ) shorter distances with respect to  $g_{NN}(r)$ . The present results completely agree with this picture, and the scarce importance, at least in the first neighbor shell, of the stacking arrangements is further confirmed by the comparison of the first peak position ( $\sim 5.6 \text{ \AA}$ ) in the center of mass  $g(r)$  displayed in Figure 6 and the one found for  $g_{NN}(r)$  ( $4.9 \text{ \AA}$ ), which should be equal in a stacked geometry. From a modeling point of view, it is worth noticing that this description can be achieved without resorting to additional interaction sites, which have the computational drawback of increasing the computational cost of the MD runs.<sup>66</sup>

Despite a detailed analysis of the structural patterns of all the considered species is beyond the aims of the present work, the results achieved for furan and thiophene might be worth an additional comment, as a further proof of the QMD-FF accuracy in reproducing the liquid microscopic structure. To this end in Figure 8 the center of mass and atom-atom pair correlation functions are displayed for furan and thiophene. As far as the  $g(r)$  function is concerned, the two curves shown in the top panel have an almost identical shape, but the thiophene function is shifted to higher  $r$  values. A slightly larger shift ( $\sim 0.6 \text{ \AA}$ ) is also found for the atom-atom functions (bottom panel): noticeably, this value is in very good agreement with the experimental shift of  $0.5 \text{ \AA}$ , found with energy dispersive X-ray diffraction spectra.<sup>102</sup> In that work, Gontrani and co-workers<sup>102</sup> also commented on a small shoulder around  $4 \text{ \AA}$ , evident for thiophene but absent in furan, which again

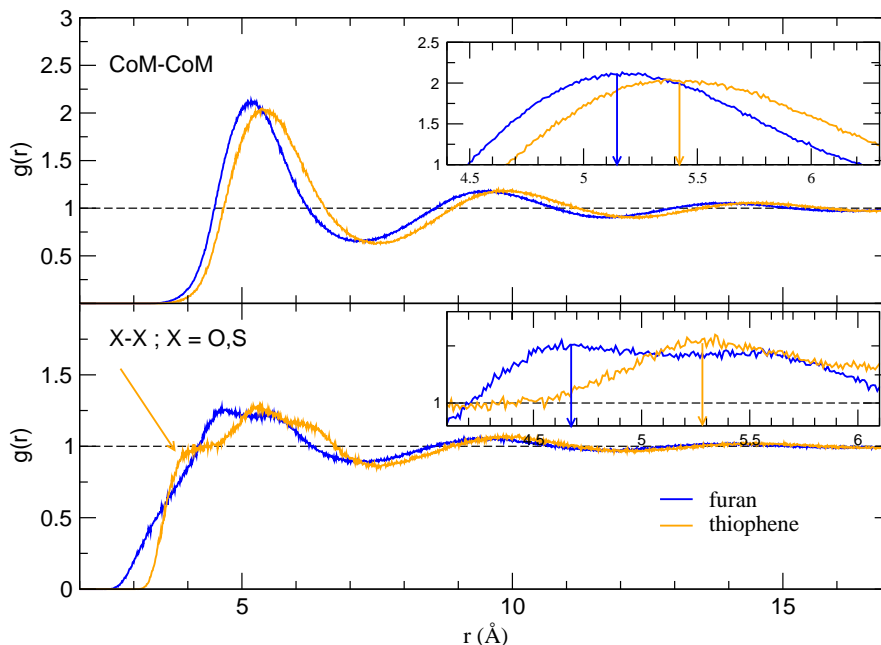


Figure 8: Center of mass (top) and atomic (bottom) pair correlation functions ( $g(r)$ ) computed for furan (blue lines) and thiophene (orange lines) from MD trajectories at 298 K and 1 atm. The position of the both the neighbor peaks is evidenced in the inset.

is accounted for by the present QMD-FF.

*A final validation on the ability of QMD-FFs to accurately represent the liquid structure of the resulting condensed phases can be achieved by computing the static dielectric constant  $\epsilon_0$ . As noted by Coleman and coworkers,<sup>4</sup>  $\epsilon_0$  is one of the most difficult properties to reproduce. As possible causes, the authors suggest<sup>4</sup> two factors: the simulation time being too short (1-5 ns) and the lack of explicit polarization in the tested empirical FFs. As far as the former issue is concerned, following the suggestions of Ref. [4], here the simulation time was here extended to values about one order of magnitude larger than the auto-correlation time*

$$\tau_M = \int_0^\infty \Phi(t) ; \Phi(t) = \left\langle \frac{\bar{\mathbf{M}}(t)\bar{\mathbf{M}}(0)}{\bar{\mathbf{M}}(0)^2} \right\rangle$$

which lead to 25 ns runs for pyridine, pyrazine, oxazole and isoxazole and 50 ns runs for

the rest of the investigated samples. Turning to the second, as for the empirical FFs tested by Caleman and co-workers,<sup>4</sup> polarization is not explicitly accounted for in our QMD-FF, hence  $\epsilon_0$  can be computed through equation (17). However, exploiting the link between the QMD-FF and the parent DFT description,  $\epsilon_\infty$  can be estimated through relation (19) by computing both the molar volume and the polarizability at QM level, and equation (18) employed to a more accurate estimate of  $\epsilon_0$ . All the computed values are reported in

Compound	exp <sup>103</sup>	QMD-FF	OPLS <sup>4</sup>	GAFF <sup>4</sup>
pyridine	12.98	7.6 ( 5.9)	6.7	6.6
pyridazine	-	38.5 (37.2)	-	-
pyrimidine	-	8.8 ( 7.7)	8.8	8.6
pyrazine <sup>(a)</sup>	2.80	2.2 ( 1.1)	-	-
pyrrole	7.72	5.4 ( 4.0)	4.0	4.2
furan	2.95	3.0 ( 1.4)	1.5	1.5
thiophene	2.72	2.7 ( 1.4)	2.6	-
oxazole	-	5.4 ( 4.4)	-	-
isoxazole	-	15.8 (14.9)	-	-

<sup>(a)</sup> All data for pyrazine refer to T = 340 K

Table 5: Comparison between experimental, QMD-FF and literature<sup>4</sup> dielectric constant  $\epsilon_0$ . QMD-FF were computed according equations (17) and (18) and the former value is reported in parenthesis. Unless otherwise stated all data refer to 298 K.

Table 5 and compared to both experimental measures and to values reported in literature for popular empirical FFs.<sup>4</sup> Despite the extended length of the MD runs, in line with the results reported for empirical FFs, the dielectric constant confirms to be a property difficult to reach, in particular when polarizability or additional virtual sites<sup>69</sup> are not explicitly included in the FF. In fact, although the use of equation (18) allows to achieve values closer to the experiment with respect to standard force-fields, the error is still larger than the one registered for most other properties.

### 4.2.3 Molecular structure and dynamics

All the above discussed validation tests concerned with thermodynamic and structural properties of the condensed liquid phase, which are known to be mostly influenced by the intermolecular term of the FF. Yet, the present QMD-FF protocol should also be able to deliver a reliable description of the single molecule structure and its flexibility. This is

a key feature whenever classical MD simulations are employed as a statistical sampling tool for successive QM calculations, as for instance in the so-called MD/QM sequential approaches (see for instance Ref. [49,50,52]). Since in these methods the geometries sampled from MD trajectories are directly used in QM calculations without any further refinement, it is mandatory to adopt a FF capable to yield an accurate representation not only of the equilibrium molecular structure, but also of the geometrical distortions that may take place during the MD run because of thermal fluctuations and intermolecular interactions.

The quality of the intramolecular description is here checked by comparing, with the available experimental data, two additional observables, namely the molecular dipole moment and the vibrational frequencies. First, considering that the QMD-FF equilibrium internal coordinates,  $q^0$ , are taken directly from the QM optimized geometry, the comparison of the QMD-FF equilibrium dipole moment

$$\mu_{QMD-FF}^0 = \left| \sum_{\alpha=1}^{N_{atoms}} q_{\alpha} r_{\alpha}^0 \right| \quad (23)$$

with its experimental gas-phase counterpart,  $\mu_{exp}^{gas}$ , can give information on the reliability of the point charge distribution. Next, the geometrical distortions sampled through the MD runs can be evaluated by comparing the average MD dipole  $\bar{\mathbf{M}}$  and its fluctuations to  $\mu_{QMD-FF}^0$ . Finally, further information on the accuracy of the distorted geometries can be retrieved from the comparison with the available experimental measures of the vibrational frequencies, computed along the MD trajectories.

The molecular dipole moment at the equilibrium geometry,  $\mu_{QMD-FF}^0$ , and its mean value in the liquid phase,  $|\bar{\mathbf{M}}|$ , were computed for each investigated compound and reported in in Table 6, together with the gas phase experiment. A nice agreement appears in general for  $\mu_{QMD-FF}^0$ , being the average error  $\sim 0.05$  D, with the maximum error found for thiophene (0.26 D). By considering that the QMD-FF point charges determining  $\mu_{QMD-FF}^0$  were obtained by PICKY with a global fitting through the simultaneous optimization of both LJ parameters and point charges, this agreement goes beyond the most optimistic forecasts. The mean value  $|\bar{\mathbf{M}}|$  is correctly near  $\mu_{QMD-FF}^0$  in all cases, and its deviation never exceeds 0.2 D, thus suggesting that no unphysical distortion took place during simulation and that the MD trajectories can be safely used in the computation of vibrational frequencies.

Compound	$\mu_{QMD-FF}^0$ (D)	$ \bar{\mathbf{M}} $ (D)	$\mu_{exp}^{gas}$ (D)	
			Ref. [104]	Ref. [93]
Pyridine	2.20	$2.20 \pm 0.12$	2.15	2.22
Pyridazine	4.23	$4.09 \pm 0.08$	4.10	4.22
Pyrimidine	2.33	$2.35 \pm 0.11$	2.42	2.33
Pyrazine	0.00	$0.27 \pm 0.18$	0.00	0.00
Pyrrole	1.96	$2.01 \pm 0.11$	1.74	1.77
Furan	0.60	$0.64 \pm 0.10$	0.72	0.66
Thiophene	0.78	$0.71 \pm 0.12$	0.52	0.55
Oxazole	1.70	$1.71 \pm 0.09$	1.50	1.50
Isoxazole	2.80	$2.82 \pm 0.08$	2.90	2.95

Table 6: Dipole moment at the QMD-FF equilibrium monomer geometry ( $\mu_{QMD-FF}^0$ ), average dipole  $\bar{\mathbf{M}}$ , see equation (14), from MD simulations, and experimental dipole in gas phase ( $\mu_{exp}^{gas}$ ).

Before discussing the QMD-FF performances in reproducing molecular vibrations, few issues should be briefly commented. First, despite the quantum correction (22), the IR spectrum was computed according to classical physics, hence a quantitative agreement on the peak intensities should not be expected. Therefore the following discussion will be focused on the peaks position, rather than on their intensity. A second issue is connected with the possible inaccuracies arising from the DFT description. As the intramolecular term of each QMD-FF here presented is built by minimizing the differences with a DFT Hessian matrix, the error of the QM calculation is unavoidably transferred to the resulting FF. This drawback can be partially circumvented by scaling the QMD-FF computed frequencies by a proper factor, *e.g.* 0.952, as suggested in the Minnesota Database of Frequency Scale Factors for Electronic Model Chemistries.<sup>105,106</sup>

The vibrational frequencies of the three diazines here investigated, were computed according to two different methods, *i.e.* directly exploiting the QMD-FF Hessian resulting from the intramolecular parameterization, and by means of equation (21), applied to the MD trajectories. The results are displayed in Figure 9 and correlated to the experimental data<sup>107</sup> obtained for the almost isolated molecule in an Argon matrix or in condensed bulk phase (either liquid, low temperature crystal or vapor). Despite most of the experimental frequencies have been found in the computed spectra, the agreement between the two sets is again found to depend on the type of normal mode considered. As suggested by the



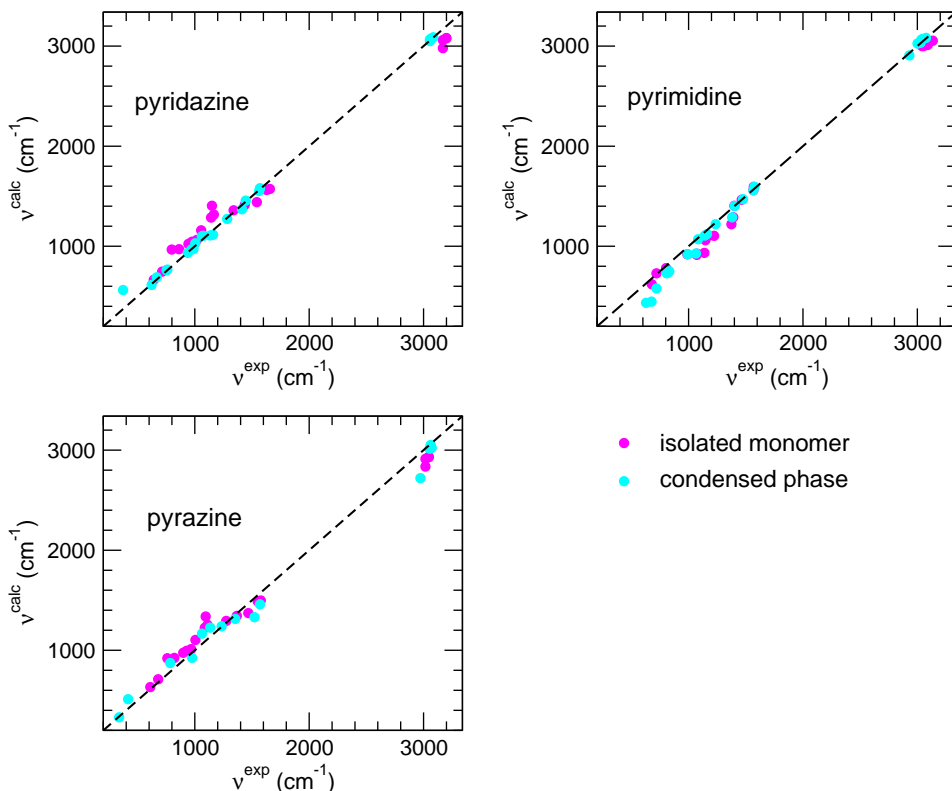


Figure 9: Correlation plots between computed and experimental<sup>107</sup> vibrational frequencies for the investigated diazines. The whole IR spectra computed from MD simulations are reported in the Supporting Information, Figures D1-D3.

comparison shown in Figure 3, the frequencies corresponding to the more localized modes agree better with their experimental counterparts. In fact, as appears by looking at Table 7, with the exception of the two *CN* stretching modes in pyrazine ( $\sim 1180$  and  $\sim 1150$   $\text{cm}^{-1}$ ) (where only a single peak (at  $\sim 1220$   $\text{cm}^{-1}$ ) is found both *in vacuo* or in the bulk phase), all the other stretching frequencies considered match fairly well the experimental peaks. Moreover, QMD-FFs appear to be able to account for the trends within the homologue series, as for instance, the experimental red shift of the *CN* stretching in going from pyridazine to pyrazine. Finally, the strong resemblance between the experimental stretching frequencies found for the isolated molecule and in condensed phase indicates that the bulk environment has only a minor influence on the position of the vibrational

Pyridazine				
stretching mode	$\nu^{QMD-FF}$ (cm <sup>-1</sup> )		$\nu^{exp}$ (cm <sup>-1</sup> )	
	<i>in vacuo</i>	liquid	Ar matrix	liquid
<i>NN</i>	968	973	967	973
<i>CC'</i>	1468	1457	1441	1445
	1087	1087	1046	1062
<i>CN</i>	1550	1551	1563	1563
<i>C'C'</i>	1557	1580	1573	1570

Pyrimidine				
stretching mode	$\nu^{QMD-FF}$ (cm <sup>-1</sup> )		$\nu^{exp}$ (cm <sup>-1</sup> )	
	<i>in vacuo</i>	liquid	Ar matrix	crystal
<i>CN</i>	1404	1402	1401	1403
	1221	1224	1223	1236
<i>C'N</i>	1593	1580	1567	1566
	1090	1110	1157	1161
<i>CC</i>	1560	1559	1570	1576
	1060	1067	1071	1064

Pyrazine				
stretching mode	$\nu^{QMD-FF}$ (cm <sup>-1</sup> )		$\nu^{exp}$ (cm <sup>-1</sup> )	
	<i>in vacuo</i>	liquid	Ar matrix	crystal
<i>CC</i>	1073	1077	1065	1074
<i>CN</i>	1224	1226	1177	1179
	-	-	1150	1153

Table 7: Comparison between selected experimental<sup>107</sup> and QMD-FF computed stretching frequencies for the investigated diazines. The whole IR spectra computed from MD simulations are reported in the Supporting Information, Figures D1-D3.

peaks: this feature is reasonably reproduced by the QMD-FF, suggesting that the intra- and inter-molecular terms are well balanced in the current FFs.

A final test was performed by comparing the *CH* stretching frequencies of two similar five-membered heterocycles, namely furan and thiophene. The capability of the intramolecular QMD-FF term to discern small details as the shift in the *CH* stretching frequencies was already commented (see Figure 3). A further validation appears by looking at Table 8, where the experimental *CH* stretching frequencies, detected in liquid phase, are compared with those extracted from the IR spectrum (see Figure D4 in the Supporting Information) computed from the MD trajectories according to equation (21).

In fact it is evident that the QMD-FFs are not only able to deliver frequencies close to the

$\nu_{MD}^{QMD-FF}$ (cm <sup>-1</sup> )		$\nu_{liq}^{exp}$ (cm <sup>-1</sup> ) <sup>108</sup>	
Furan	Thiophene	Furan	Thiophene
3154	3104	3159	3116
3140	3097	3136	-
3118	3075	3129	3083
3111	3068	3120	3075

Table 8: Comparison between selected experimental<sup>108</sup> and QMD-FF computed *CH* stretching frequencies for furan and thiophene. The whole IR spectrum, computed from MD simulations, is displayed in the Supporting Information, Figure D4.

experimental values, but also can distinguish the *CH* vibrations of the two heterocycles, quantitatively ( $\langle \delta\nu_{CH} \rangle^{QMD-FF} \simeq 45$  cm<sup>-1</sup>) reproducing the average experimental red shift ( $\langle \delta\nu_{CH} \rangle^{exp} \simeq 45$  cm<sup>-1</sup>) in going from furan to thiophene.

## 5 Conclusions

This work was aimed to validate the JOYCE/PICKY approach in producing QMD-FFs for aromatic molecules bearing heteroatoms, focusing in particular on the possibility of integrating the recently proposed MP2<sup>mod</sup> procedure into an automated protocol. With these goals, system specific QMD-FFs were parameterized for a representative set of nine heterocyclic aromatic species, and used in the MD simulation of the corresponding liquid phases. The validation of the whole procedure was achieved by comparing thermodynamic, structural and dynamic properties to the available experimental data of the condensed phases.

A general good agreement with the experiment was found, and, the specificity of the QMD-FFs allowed for distinguishing the different properties of the investigated species, even among similar molecules. In many cases, it was found that QMD-FFs delivered more accurate predictions than common transferable FFS, especially for those compounds not included in the training set considered for their parameterization. An emblematic case concerns with the vaporization enthalpy of pyrazine, oxazole or isoxazole, whose error was lowered from  $\sim 20\%$  to  $\sim 5\%$ , in going from standard FFS to the present QMD-FFs. Besides the results achieved in the thermodynamic and structural description of the investigated

liquid phases, another point of strength of the here proposed parameterization route stands in the capability of the resulting QMD-FFS to yield, through MD simulations, reliable collections of accurate molecular structures, as confirmed in the discussion of the vibrational frequencies.

Given the importance of heterocycles in biomolecular processes and drug design, the possibility of parameterizing accurate FFs without resorting to any experimental data is appealing, in particular if novel species have to be investigated. In this context the present results are rather promising, and we plan to extend the whole procedure to larger and more complex molecules. Yet, some deviations from the average good performances call for further investigations. For instance, the results recorded for pyrrole can be traced back to some inaccuracies of the  $\text{MP2}^{mod}$  energies. Instead, the  $\sim 5\%$  average error on bulk density, also found in DFT-D derived QMD-FF for other classes of compounds, seems a systematic trend connected with the parameterization procedure, rather than with the parent QM method.<sup>8,33</sup> For these reasons work is in progress in our laboratory in order to further improve QMD-FF flexibility and their adherence with the QM-IPES. Finally, one further hypothesis to be explored concerns with the lack of three-body interactions, whose effect on bulk density was recently estimated by McDaniel and Schmidt<sup>109</sup> to be on average about 5%, on a wide set of benchmark molecules.

The second goal of this paper concerned with the integration of the  $\text{MP2}^{mod}$  method in the automated PICKY procedure. The present QMD-FF parameterization exploits two main qualities of  $\text{MP2}^{mod}$ : its adherence with the CCSD(T) intermolecular energies and its low computational cost, which allows considering large portion of IPES at a reasonable CPU time. The quality achieved for MD results seems to testify in favor of the reliability of  $\text{MP2}^{mod}$  results in describing large portions of the heterocycles IPESs. At the same time, the convenience of  $\text{MP2}^{mod}$ , which is determined by the low cost of perturbative methods and by the small basis set employed, is crucial in the present context, where the energy of about 500 dimer arrangements enter the PICKY parameterization protocol.

From a more general point of view, the present investigation should be considered a further step of a long-lasting project aimed at developing a sound and accurate QMD-FF parameterization procedures for classical computer simulations of advanced materials

and complex systems. In fact, the JOYCE and PICKY protocols have been recently tested and validated in combination with different methods (DFT<sup>8,33</sup> or wave function based techniques, as in this work), different simulation procedures (MC<sup>8</sup> or MD<sup>33</sup>), and different species (aromatic, halogenated hydrocarbons, etc.). If on the one hand the JOYCE protocol is straightforwardly extendible to almost any type of molecule, on the other hand, in its present formulation, the PICKY procedure is only applicable to homodimers, hence to the simulation of pure condensed phases and crystals. The next steps, currently in progress in our group, will consist in the extension of the intermolecular parameterization to large molecules first, and hetero-dimers in the next future. When tested and validated, such approaches will allow to extend the QMD-FF application range to mixtures, solutions, and, eventually, inhomogeneous systems.

### **Supporting Information**

Additional data and several details about the reported calculations not included in this paper are available: specific atom-type labeling, parameterization details, complete list of intra- and inter-molecular QMD-FF parameters, pyridine's  $N \cdots H$  pair correlation functions and computed IR spectra. Specific input and output files of the several codes employed throughout the present work are available from the authors upon request.

## 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] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.

- [9] 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.
- [10] Kistenmacher, H.; Popkie, H.; Clementi, E.; Watts, R. Study of structure of molecular complexes .7. Effect of correlation energy corrections to Hartree-Fock water-water potential on Monte-Carlo simulations of liquid water *J. Chem. Phys.* **1974**, *60*, 4455–4465.
- [11] Lie, G. C.; Clementi, E. Study of structure of molecular complexes .12. Structure of liquid water obtained by Monte-Carlo simulation with Hartree-Fock potential corrected by inclusion of dispersion forces. *J. Chem. Phys.* **1975**, *62*, 2195–2199.
- [12] Bolis, G.; Clementi, E. Analytical potentials from ab initio computations for interaction between biomolecules .3. Reliability and transferability of pair potentials. *J. Am. Chem. Soc.* **1977**, *99*, 5550–5557.
- [13] Clementi, E.; Corongiu, G.; Ranguino, G. Analytical potentials from ab initio computations for the interaction between biomolecules .7. Polar amino-acids and conclusions *J. Chem. Phys.* **1981**, *74*, 578–588.
- [14] Karlstrom, G.; Linse, P.; Wallqvist, A.; Jonssonf, B. Intermolecular Potentials for the H<sub>2</sub>O-C<sub>6</sub>H<sub>6</sub> and the C<sub>6</sub>H<sub>6</sub>-C<sub>6</sub>H<sub>6</sub> Systems Calculated in an ab Initio SCF CI Approximation *J. Am. Chem. Soc.* **1983**, *6*, 3777–3782.
- [15] Linse, P.; Engström, S.; Jönsson, B. Molecular dynamics simulation of liquid and solid benzene *Chem. Phys. Lett.* **1985**, *115*, 95–100.
- [16] Vaiana, A.; Schulz, A.; Wolfrum, J.; Saure, M.; Smith, J. Molecular Mechanics Force Field Parameterization of the Fluorescent Probe Rhodamine 6G Using Automated Frequency Matching *J. Comp. Chem.* **2003**, *24*, 632–639.
- [17] Cournia, Z.; Vaiana, A.; Matthias Ulmann, G.; Smith, J. Derivation of a Molecular Mechanics Force Field for Cholesterol *Pure Appl. Chem.* **2004**, *76*, 189–196.

- [18] 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.
- [19] Prampolini, G. Parametrization and Validation of Coarse Grained Force-Fields Derived from ab Initio Calculations *J. Chem. Theory Comput.* **2006**, *2*, 556–567.
- [20] Cacelli, I.; Prampolini, G. Parametrization and Validation of Intramolecular Force Fields Derived from DFT Calculations *J. Chem. Theory Comput.* **2007**, *3*, 1803–1817.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] Barone, V.; Cacelli, I.; De Mitri, N.; Licari, D.; Monti, S.; Prampolini, G. Joyce and Ulysses: Integrated and User-Friendly Tools for the Parameterization of Intramolecular Force Fields from Quantum Mechanical Data. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3736–51.



- [26] 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.
- [27] Yakovenko, O. Y.; Li, Y. Y.; Oliferenko, A. a.; Vashchenko, G. M.; Bdzholo, V. G.; Jones, S. J. M. Ab Initio Parameterization of YFF1, a Universal Force Field for Drug-Design Applications. *J. Mol. Model.* **2012**, *18*, 663–73.
- [28] Grimme, S. A General Quantum Mechanically Derived Force Field (QMDF) for Molecules and Condensed Phase Simulations *J. Chem. Theory Comput.* **2014**, 4497–4514.
- [29] 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.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] Prampolini, G.; Campetella, M.; De Mitri, N.; Livotto, P. R.; Cacelli, I. Systematic and Automated Development of Quantum Mechanically Derived Force Fields: The Challenging Case of Halogenated Hydrocarbons *J. Chem. Theory Comput.* **2016**, *12*, 5525–5540.

- [34] 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.
- [35] 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.
- [36] 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 N4 (Td) explosives *Theor. Chem. Accounts* **2016**, *135*, 1–9.
- [37] Khan, H. M.; Grauffel, C.; Broer, R.; MacKerell, A. D.; Havenith, R. W. A.; Reuter, N. Improving the Force Field Description of TyrosineCholine Cation Interactions: QM Investigation of PhenolN(Me)<sub>4</sub><sup>+</sup> Interactions *J. Chem. Theory Comput.* **2016**, *12*, 5585–5595.
- [38] Zahariev, F.; De Silva, N.; Gordon, M. S.; Windus, T. L.; Dick-Perez, M. ParFit: A Python-Based Object-Oriented Program for Fitting Molecular Mechanics Parameters to ab Initio Data *J. Chem. Inf. Mod.* **2017**, *57*, 391–396.
- [39] Vandenbrande, S.; Waroquier, M.; Van Speybroeck, V.; Verstraelen, T. The monomer electron density force field (MEDFF): A Physically Inspired Model for Noncovalent Interactions *J. Chem. Theory Comput.* **2017**, *13*, 161–179.
- [40] Piquemal, J.-P.; Jordan, K. D. From Quantum Mechanics to Force Fields: New Methodologies for the Classical Simulation of Complex Systems *Theor. Chem. Accounts* **2012**, *131*, 1207.
- [41] 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**, *49*, 1844–1851.

- [42] Piquemal, J.-p.; Jordan, K. D. Preface: Special Topic: From Quantum Mechanics to Force Fields *J. Chem. Phys.* **2017**, *147*, 161401.
- [43] 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.
- [44] 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.
- [45] 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.
- [46] 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.
- [47] 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.
- [48] Macchiagodena, M.; Mancini, G.; Pagliai, M.; Barone, V.; Rode, B. M.; Meng, E. C.; Ferrin, T. E.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. Accurate prediction of bulk properties in hydrogen bonded liquids: amides as case studies *Phys. Chem. Chem. Phys.* **2016**, *18*, 25342–25354.
- [49] Cacelli, I.; Ferretti, A.; Prampolini, G. Predicting light absorption properties of anthocyanidins in solution: a multi-level computational approach *Theor. Chem. Accounts* **2016**, *135*, 156.

- [50] 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.
- [51] Kraner, S.; Prampolini, G.; Cuniberti, G. Exciton Binding Energy in Molecular Triads *J. Phys. Chem. C* **2017**, *121*, 17088–17095.
- [52] Andreussi, O.; Prandi, I. G.; Campetella, M.; Prampolini, G.; Mennucci, B. Classical Force Fields Tailored for QM Applications: Is It Really a Feasible Strategy? *J. Chem. Theory Comput.* **2017**, *13*, 4636–4648.
- [53] 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.
- [54] Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory *J. Comp. Chem.* **2011**, *32*, 1456–65.
- [55] Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-Corrected Mean-Field Electronic Structure Methods *Chem. Rev.* **2016**, *116*, 5105–5154.
- [56] Jacobs, M.; Greff da Silveira, L.; Prampolini, G.; Livotto, P. R.; Cacelli, I. Interaction energy landscapes of aromatic heterocycles through a reliable yet affordable computational approach. *J. Chem. Theory Comput.* **2018**, *14*, 543–556.
- [57] Hohenstein, E. G.; Sherrill, C. D. Effects of Heteroatoms on Aromatic  $\pi$ - $\pi$  Interactions: Benzene-Pyridine and Pyridine Dimer *J. Phys. Chem. A* **2009**, *113*, 878–886.
- [58] Mishra, B. K.; Arey, J. S.; Sathyamurthy, N. Stacking and Spreading Interaction in N-Heteroaromatic Systems *J. Phys. Chem. A* **2010**, *114*, 9606–9616.
- [59] Wheeler, S. E.; Bloom, J. W. G. Toward a More Complete Understanding of Non-covalent Interactions Involving Aromatic Rings *J. Phys. Chem. A* **2014**, *118*, 6133–6147.

- [60] Hwang, J.; Li, P.; Carroll, W. R.; Smith, M. D.; Pellechia, P. J.; Shimizu, K. D. Additivity of Substituent Effects in Aromatic Stacking Interactions *J. Am. Chem. Soc.* **2014**, *136*, 14060–14067.
- [61] Huber, R. G.; Margreiter, M. A.; Fuchs, J. E.; von Grafenstein, S.; Tautermann, C. S.; Liedl, K. R.; Fox, T. Heteroaromatic  $\pi$ -Stacking Energy Landscapes *J. Chem. Inf. Mod.* **2014**, *54*, 1371–1379.
- [62] Sierański, T. The Intricacies of the Stacking Interaction in a Pyrrole-Pyrrole System *Struct. Chem.* **2016**, *27*, 1107–1120.
- [63] Mackie, I. D.; McClure, S. A.; DiLabio, G. A. Binding in Thiophene and Benzothiophene Dimers Investigated by Density Functional Theory with Dispersion-Correcting Potentials. *J. Phys. Chem. A* **2009**, *113*, 5476–84.
- [64] An, Y.; Doney, A. C.; Andrade, R. B.; Wheeler, S. E. Stacking Interactions between 9-Methyladenine and Heterocycles Commonly Found in Pharmaceuticals *J. Chem. Inf. Mod.* **2016**, *56*, 906–914.
- [65] Baker, C. M.; Grant, G. H. The Structure of Liquid Benzene *J. Chem. Theory Comput.* **2006**, *2*, 947–955.
- [66] Baker, C. M.; Grant, G. H. Modeling Aromatic Liquids: Toluene, Phenol, and Pyridine *J. Chem. Theory Comput.* **2007**, *3*, 530–548.
- [67] Lopes, P. E. M.; Lamoureux, G.; Mackerell, A. D. Polarizable empirical force field for nitrogen-containing heteroaromatic compounds based on the classical Drude oscillator *J. Comp. Chem.* **2009**, *30*, 1821–1838.
- [68] Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins *J. Chem. Theory Comput.* **2016**, *12*, 281–296.

- [69] Macchiagodena, M.; Mancini, G.; Pagliai, M.; Del Frate, G.; Barone, V. Fine-tuning of atomic point charges: Classical simulations of pyridine in different environments *Chem. Phys. Lett.* **2017**, *677*, 120–126.
- [70] 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.
- [71] Rezáč, J.; Hobza, P. Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. *Chem. Rev.* **2016**, *116*, 5038–5071.
- [72] 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.
- [73] 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.
- [74] 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.

- [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] Lin, S. T.; Blanco, M.; Goddard III, W. A. The Two-Phase Model for Calculating Thermodynamic Properties of Liquids from Molecular Dynamics: Validation for the Phase Diagram of Lennard-Jones Fluids *J. Chem. Phys.* **2003**, *119*, 11792–11805.
- [79] Lin, S.-T.; Maiti, P. K.; Goddard III, W. A. Two-Phase Thermodynamic Model for Efficient and Accurate Absolute Entropy of Water from Molecular Dynamics Simulations *J. Phys. Chem. B* **2010**, *114*, 8191–8198.
- [80] Pascal, T. A.; Lin, S.-T.; Goddard III, W. A. Thermodynamics of Liquids: Standard Molar Entropies and Heat Capacities of Common Solvents from 2PT Molecular Dynamics *Phys. Chem. Chem. Phys.* **2011**, *13*, 169–181.
- [81] Huang, S.-N.; Pascal, T. A.; Goddard III, W. A.; Maiti, P.; Lin, S.-T. Absolute Entropy and Energy of Carbon Dioxide Using the Two-Phase Thermodynamic Model *J. Chem. Theory Comput.* **2011**, *7*, 1893–1901.
- [82] Gereben, O.; Pusztai, L. On the accurate calculation of the dielectric constant from molecular dynamics simulations: The case of SPC/E and SWM4-DP water *Chem. Phys. Lett.* **2011**, *507*, 80–83.
- [83] Neumann, M.; Steinhauser, O. Computer simulation and the dielectric constant of polarizable polar systems *Chem. Phys. Lett.* **1984**, *106*, 563–569.
- [84] Lamoureux, G.; MacKerell, A. D.; Roux, B. A simple polarizable model of water based on classical Drude oscillators *J. Chem. Phys.* **2003**, *119*, 5185–5197.

- [85] Iftimie, R.; Tuckerman, M. E. Decomposing total IR spectra of aqueous systems into solute and solvent contributions: A computational approach using maximally localized Wannier orbitals *J. Chem. Phys.* **2005**, *122*, 214508.
- [86] Egorov, S. A.; Skinner, J. L. Semiclassical approximations to quantum time correlation functions *Chem. Phys. Lett.* **1998**, *293*, 469–476.
- [87] Vener, M. V.; Sauer, J. Environmental effects on vibrational proton dynamics in  $\text{H}_5\text{O}_2^+$ : DFT study on crystalline  $\text{H}_5\text{O}_2^+ \text{ClO}_4^-$  *Phys. Chem. Chem. Phys.* **2005**, *7*, 258–263.
- [88] Gaigeot, M.-P.; Sprik, M. Ab Initio Molecular Dynamics Computation of the Infrared Spectrum of Aqueous Uracil *J. Phys. Chem. B* **2003**, *107*, 10344–10358.
- [89] Cerezo, J.; Prampolini, G.; Cacelli, I. Developing accurate intramolecular force fields for conjugated systems through explicit coupling terms *Theor. Chem. Accounts* **2018**, *137*, 80.
- [90] Bougeard, D.; LeCalve, N.; Novak, A.; Bachanh, N. Phase Transitions of Pyrazine *Mol. Cry. Liq. Cry.* **1978**, *44*, 113.
- [91] Yaws, C. Yaws' Handbook of Thermodynamic Properties for Hydrocarbons and Chemicals; Knovel, 2009.
- [92] Lipkind, D.; Chickos, J. S. An examination of the vaporization enthalpies and vapor pressures of pyrazine, pyrimidine, pyridazine, and 1,3,5-triazine *Struct Chem* **2009**, *20*, 49–58.
- [93] Lide, D. R. CRC Handbook of Chemistry and Physics, 84th Edition, 2003-2004; CRC press: Boca Raton, FL, 2003; Vol. 53.
- [94] Tjebbes, J.; Mollberg, R.; Lindberg, B.; McKay, J.; Theander, O.; Flood, H. The Heats of Combustion and Formation of the Three Diazines and their Resonance Energies. *Acta Chem. Scand.* **1962**, *16*, 916–921.
- [95] Marcus, Y. The properties of solvents; Wiley, 1998.



- [96] McCormick, D.; Hamilton, W. The enthalpies of combustion and formation of oxazole and isoxazole *J. Chem. Thermodynam.* **1978**, *10*, 275–278.
- [97] Steele, W. V.; Chirico, R. D.; Knipmeyer, S. E.; Nguyen, A.; Smith, N. K.; Tasker, I. R. Thermodynamic Properties and Ideal-Gas Enthalpies of Formation for Cyclohexene, Phthalan (2,5-Dihydrobenzo-3,4-furan), Isoxazole, Octylamine, Dioctylamine, Trioctylamine, Phenyl Isocyanate, and 1,4,5,6-Tetrahydropyrimidine *J. Chem. Eng. Data* **1996**, *41*, 1269–1284.
- [98] Chirico, R.; Steele, W.; Nguyen, A.; Klots, T.; Knipmeyer, S. Thermodynamic properties of pyridineI. Vapor pressures, high-temperature heat capacities, densities, critical properties, derived thermodynamic functions, vibrational assignment, and derivation of recommended values *J. Chem. Thermodynam.* **1996**, *28*, 797–818.
- [99] Guthrie, G. B.; Scott, D. W.; Hubbard, W. N.; Katz, C.; McCullough, J. P.; Gross, M. E.; Williamson, K. D.; Waddington, G. Thermodynamic Properties of Furan *J. Am. Chem. Soc.* **1952**, *74*, 4662–4669.
- [100] Figuiere, P.; Szwarc, H.; Oguni, M.; Suga, H. Calorimetric study of thiophene from 13 to 300 K. Emergence of two glassy crystalline states *J. Chem. Thermodynam.* **1985**, *17*, 949–966.
- [101] McCune, J. A.; Turner, A. H.; Coleman, F.; White, C. M.; Callear, S. K.; Youngs, T. G. A.; Swadźba-Kwaśny, M.; Holbrey, J. D. Association and liquid structure of pyridineacetic acid mixtures determined from neutron scattering using a free proton’ EPSR simulation model *Phys. Chem. Chem. Phys.* **2015**, *17*, 6767–6777.
- [102] Gontrani, L.; Ramondo, F.; Caminiti, R. Furan and thiophene in liquid phase: An X-ray and molecular dynamics study *Chem. Phys. Lett.* **2006**, *422*, 256–261.
- [103] Wohlfarth, C. Static Dielectric Constants of Pure Liquids and Binary Liquid Mixtures, Landolt-Brnstein, Numerical Data and Functional Relationships in Science and Technology, New Series, Group IV, Macroscopic and Technical Properties of Matter; Springer-Verlag: Berlin, 1991; Vol. 6.

- [104] McClellan, A. Tables of Experimental Dipole Moments, Vol. 2; Rahara Enterprise: El Cerrito, 1973.
- [105] Alecu, I. M.; Zheng, J.; Zhao, Y.; Truhlar, D. G. Computational Thermochemistry: Scale Factor Databases and Scale Factors for Vibrational Frequencies Obtained from Electronic Model Chemistries *J. Chem. Theory Comput.* **2010**, *6*, 2872–2887.
- [106] Bao, J. L.; Zheng, J.; Alecu, I.; Lynch, B.; Zhao, Y.; Truhlar, D.; Minnesota Database of Frequency Scale Factors for Electronic Model Chemistries Version 3 Beta 2; <https://comp.chem.umn.edu/freqscale/version3b2.htm>; accessed on 9 february 2018.
- [107] Breda, S.; Reva, I. D.; Lapinski, L.; Nowak, M. J.; Fausto, R. Infrared spectra of pyrazine, pyrimidine and pyridazine in solid argon *J. Mol. Struct.* **2006**, *786*, 193–206.
- [108] Cataliotti, R.; Paliani, G. Infrared study of the CH stretching region of five-membered heterocyclic compounds *Can. J. Chem.* **1976**, *54*, 2451–2457.
- [109] 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*, 8042–53.