Accuracy of Quantum Mechanically Derived Force-Fields Parameterized from Dispersion-corrected DFT data: the Benzene Dimer as Prototype for Aromatic Interactions

Giacomo Prampolini,[a,*] Paolo Roberto Livotto,[b] and Ivo Cacelli[a,c]

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

[b] *Instituto de Química, Universidade Federal do Rio Grande do Sul*, Avenida Bento Gonçalves 9500, CEP 91501-970 Porto Alegre, Brazil

[c] *Dipartimento di Chimica e Chimica Industriale, Universitá di Pisa*, Via G. Moruzzi 13, I-56124 Pisa, Italy

September 17, 2015

*Corresponding author

Abstract

A multi-level approach is presented to assess the ability of several popular dispersion corrected density functionals (M06-2X, CAM-B3LYP-D3, BLYP-D3 and B3LYP-D3) to reliably describe two-body interaction potential energy surfaces (IPESs). To this end, the automated PICKY procedure (Cacelli et al., *J. Comp. Chem.* 2012, **33**, 1055) was exploited, which consists in parameterizing specific intermolecular force fields through an iterative approach, based on the comparison with quantum mechanical data. For each of the tested functionals, the resulting force field was employed in classical Monte Carlo and Molecular Dynamics simulations, performed on systems of up to one thousand molecules in ambient conditions, to calculate a number of condensed phase properties. The comparison of the resulting structural and dynamic properties with experimental data, allows us to assess the quality of each IPES, and consequently, even the quality of the DFT functionals. The methodology is tested on the benzene dimer, commonly used as benchmark molecule, prototype of aromatic interactions. The best results were obtained with the CAM-B3LYP-D3 functional. Beside assessing the reliability of DFT functionals in describing aromatic IPESs, this work provides a further step toward a robust protocol for the derivation of sound force field parameters from quantum mechanical data. This method can be relevant in all those cases where standard force fields fail in giving accurate predictions.

1 Introduction

Classical simulation methods, as Monte Carlo (MC) or Molecular Dynamics (MD) have become one of the most popular computational techniques to investigate the properties of advanced materials and biologically relevant systems. Undoubtedly,[1,2] the key ingredient of all simulation methods is the force field (FF), *i.e.* a collection of analytical functions aimed to describe the energy of the simulated system as a function of the positions of its nuclei. Consequently, the reliability of the description achieved by simulations strongly depends on the FF quality, which is in turn determined by the chosen functional form and by the set of parameters included in the selected analytical functions.[3–8] As a matter of fact, it is in the FF parameters that the chemical identity of the system under study is encoded.

The majority of popular FFs are parameterized[3,9–15] towards experimental and/or quantum mechanical (QM) data for a well determined molecular type, thus describing classes of compounds in an averaged way. On the one hand, the main advantage of such procedure is the possibility of transferring the FF parameters to similar molecules that were not initially included in the training set. On the other hand, accuracy problems may arise when the system of interest is characterized by chemical and/or physical features different from those of the original training set employed in parameterization.

One possible route to overcome these lacks is to specifically re-parameterize only some terms of the adopted FF, such as, for instance, torsional terms or point charges. A more drastic possibility is to abandon the idea of transferability in favor of FFs specific for the system under study. Following this idea, in the past decade several automated protocols have been proposed[16–25] to obtain FFs purposely tailored for molecule under investigation. In particular, as recently pointed out by Grimme,[26] there is a growing attention to novel parameterization strategies, based solely on QM data, capable of yielding very accurate FFs.[4,17,19,23–36] Despite the rather large number of different strategies, there are only few protocols, up to our knowledge, capable to parameterize specific FFs (both intramolecular and intermolecular) to be used in condensed phase simulations. Indeed, most of the proposed procedures[17,21,23,24] focus on the parameterization of the intramolecular part of the FF, deriving the parameters for the bonded interactions (*e.g.* stretching and

bending terms) from QM data computed on the isolated molecule. Only very recently, a more complete method was proposed by Grimme,[26] whose work aims to parameterize the whole FF (*i.e.* intramolecular and intermolecular parameters) from QM data with an automated procedure. Notwithstanding the protocol appears to be robust and promising, its performances in condensed phase simulations have still to be tested.

A slightly different route was developed some years ago by our group. Protocols for FF parameterization based solely on QM data were proposed separately for the intramolecular[17,18] and the intermolecular[4,27] parts and thereafter used in the calculation of thermodynamic and transport properties of a benchmark liquid crystalline compound.[30] More specifically, this quantum mechanically derived force field (QMD-FF[17]) succeeded in accurately predicting the liquid crystal phases and transition temperatures, a result usually out of the reach of conventional FFs.[37] More recently,[19] we have implemented a more robust protocol, capable to yield accurate FF parameters from QM data, for both the intramolecular and intermolecular contributions. In fact, following the suggestions of Akin-Ojo and coworkers,[25] the intermolecular FF parameters were obtained by an automated iterative approach,[19] aimed to minimize the difference between the interaction potential energy surfaces (IPESs) obtained by QM calculations and employing the QMD-FF. Such a protocol was implemented in the PICKY[19] code, and thereafter validated for pyridine, where the comparison of simulated thermodynamic, structural and transport properties with their experimental counterparts, yielded and overall good agreement.

The validation of reliable and specific parameterization routes, solely based on QM information, shifts the problem to the definition of the most appropriate QM method, whose primary requirement is a well balanced combination of accuracy and computational cost. Indeed, Hessian matrix calculations (needed for the intramolecular part) and complex IPES scans (for the intermolecular term) are often required for a given parameterization route, and the computational cost rapidly increase along with the molecular dimensions. Furthermore, when dispersion interactions are involved, as those leading to the stacking patterns common in aromatic pairs, the situation is even more complicated, as cheap QM methods might not ensure the required accuracy. A solution may come from density functional theory (DFT), where the dispersion corrected functionals[38,39]

have become very popular tools to investigate non covalent interactions. Unfortunately, the quality of the description severely depends on the chosen combination of functional, basis set and/or correction scheme.[38–44] Many of such combinations have been tested vs. high quality post-SCF methods, as for instance Coupled Cluster of singles, doubles and perturbative triples, extrapolated to the complete basis set limit (CCSD(T)/CBS), often termed the gold standard of quantum chemistry.[45] However, up to our knowledge, most of these validations were limited to few geometrical arrangements, or performed on benchmark sets of molecules. From these studies it appears that the performances of a given functional may significantly depend on the target dimer and even on its spatial arrangement.[43,46–48] This last observation suggests that it would be advisable to evaluate the performances of a given combination of functional/basis set onto larger portions of the dimer IPES, rather than benchmarking it over few, *a priori* selected conformers. This point seems particularly crucial for non-covalent complexes, where the magnitude and nature of the interaction forces may lead to relevant populations of a large number of different arrangements in condensed phase, even at room temperature.

The aforementioned PICKY program[19] allowed us to obtain a significant sample of geometrical arrangements of the dimers, that was then used to parameterize an intermolecular FF built on LJ and Coulomb functions. In previous works[4,19,30] the reference QM IPES for aromatic pairs was computed through wave function theory (WFT) calculations. In the present work we couple the PICKY parameterization procedure with dispersion corrected DFT-D descriptions of the IPES, with the aim of assessing the reliability of the considered DFT functionals in yielding accurate bulk properties. To be more precise, the accuracy of the bulk properties is used as a *criterion* for the quality of the FF. However, as the FF has been fitted onto the QM sampled energies, this criterion can be extended to the QM method employed, once a tight correspondence between QM and QMD-FF energies is assessed and the sampled geometrical arrangements can be considered adequate. In any case, two important remarks should be made at this point. First, it is clear that the simple functional forms employed in the FF (LJ + Coulomb), may undermine a precise correspondence between the FF and the QM IPES. Moreover, as the PICKY protocol is essentially based on dimer IPES sampling, the FF energy is purely

two body, neglecting three body effects on the bulk properties. However, in the recent work of McDaniel and Schmidt[33] it has been shown that the three-body contributions to the interaction energy is $\sim$ 2-5% and that its effect on the bulk density is a lowering of about 5%. On the same trend, a repulsive three-body contribution was computed for the benzene crystal lattice energy,[49] amounting to $\sim$ 7%.

The consequence of these two drawbacks (simple form of FF and pure two-body potential) is that the above statement about the correspondence between QM data and bulk properties, has to be moved to the following weaker assertion. The accuracy of the computed bulk properties is surely linked to the quality of the QM IPES, although a precise correspondence is not possible, even if, as we shall see below, it is evident how some features of the QM IPES are connected with the theoretical value of the bulk density.

Considering the large amount of required calculations, the smallest prototype of aromatic interaction was selected as a test case, *i.e.* the benzene molecule. Moreover, beside the many experimental[50–59] and computational[4,5,60–74,74–78] data available for comparison, the benzene molecule was already considered by us for QM based FF parameterization.[4] This allows us for a more thorough validation of the revised PICKY procedure here presented.

2 Parameterization routes

The QMD-FF parameterization protocols employed in this work are illustrated in Figure 1.

The basic idea substanding QMD-FF parameterization is to represent both the QM PES of the isolated target molecule and the IPES of its homo-dimer through purposely chosen analytical functions, suitable for classical MD simulations. To ease this task, the standard partition of the total FF energy ($E^{FF_{tot}}$) for a system composed of N_{mol} molecules can be invoked, *i.e.*

$$E^{FF_{tot}} = \sum_{M=1}^{N_{mol}} E_M^{FF_{intra}}(\bar{b}, \bar{\theta}, \bar{\phi}, \bar{r}^{intra}) + \sum_{A=1}^{N_{mol}} \sum_{B>A}^{N_{mol}} E_{AB}^{FF_{inter}}(\bar{r}_{AB}) \quad (1)$$

where $E_M^{FF_{intra}}$ is the intra-molecular FF contribution driving the flexibility of molecule M^{th} and depends on a collection of its internal coordinates as bonds ($\bar{b}$), angles ($\bar{\theta}$),

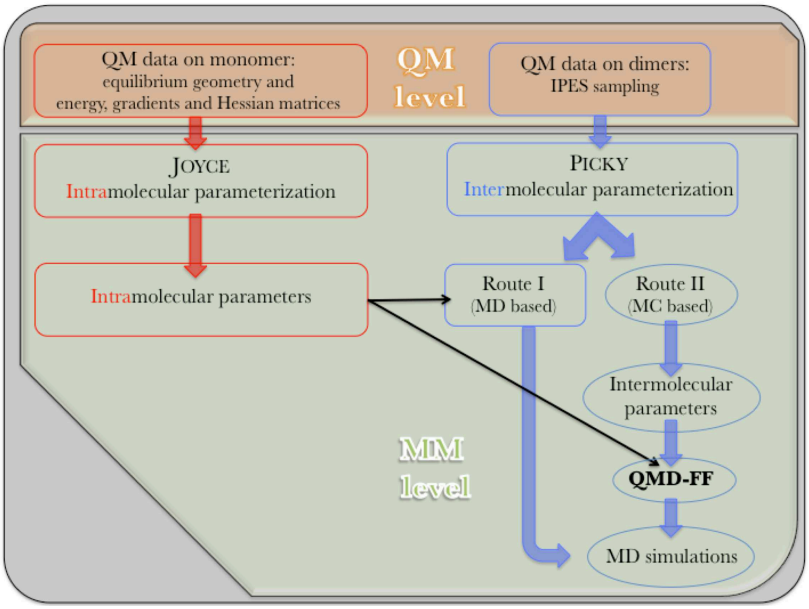


Figure 1: Flowchart of the parameterization protocols adopted for QMD-FFs. Top (brown) panel: QM calculations required for QMD-FF parameterization. The red (left) and blue (right) colors indicate if the QM data concern either intramolecular or intermolecular features, respectively. Bottom (green) panel: different parameterization routes that eventually lead to MD simulations: intramolecular parameterization (left) performed with the JOYCE method[17,18] and intermolecular parameterization obtained through the PICKY[19] method (route I or II, see text).

dihedrals ($\bar{\phi}$) or intramolecular distances ($\bar{r}^{intra}$). $E_{AB}^{FF_{inter}}$ is instead the (two-body) intermolecular part, which accounts for the interaction energy between molecules A and B, and it is a function of the set of distances $\bar{r}_{AB}$ between all interaction sites of molecule A and those of molecule B.

Within this energy partition, the parameterization of $E^{FF_{intra}}$ (highlightened in red in the left side of Figure 1) is achieved exploiting the JOYCE method, previously developed in our group,[17,18] whereas the intermolecular parameters defining $E^{FF_{inter}}$ (evidenced in blue, right side of Figure 1) are obtained through the PICKY approach. Least-squares linear (JOYCE) and not linear (PICKY) fitting procedures are performed, both aimed to minimize the difference between QM and QMD-FF computed data. As far as the intermolecular parameterization is concerned, two possible routes are illustrated in Figure 1 and, more in detail, in Figure 2. Route I was recently[19] implemented by us, and employed to parameterize a QMD-FF for the pyridine molecule. Notwithstanding the automated

protocol encoded in the PICKY software is described in detail in Ref. [19], for the sake of clarity a brief outline of PICKY 's route I is also given in the Supporting Information.

Despite the rather good results achieved,[19] route I is here further improved, handling the partition between $E^{FF_{intra}}$ and $E^{FF_{inter}}$ in a slightly more accurate manner, with the aim of enforcing the one-to-one correspondence between QM and FF energies. In fact, a possible source of inaccuracy arises from that fact that the sampling in route I is performed over MD trajectories. Since during MD simulations all the interaction sites are moved,

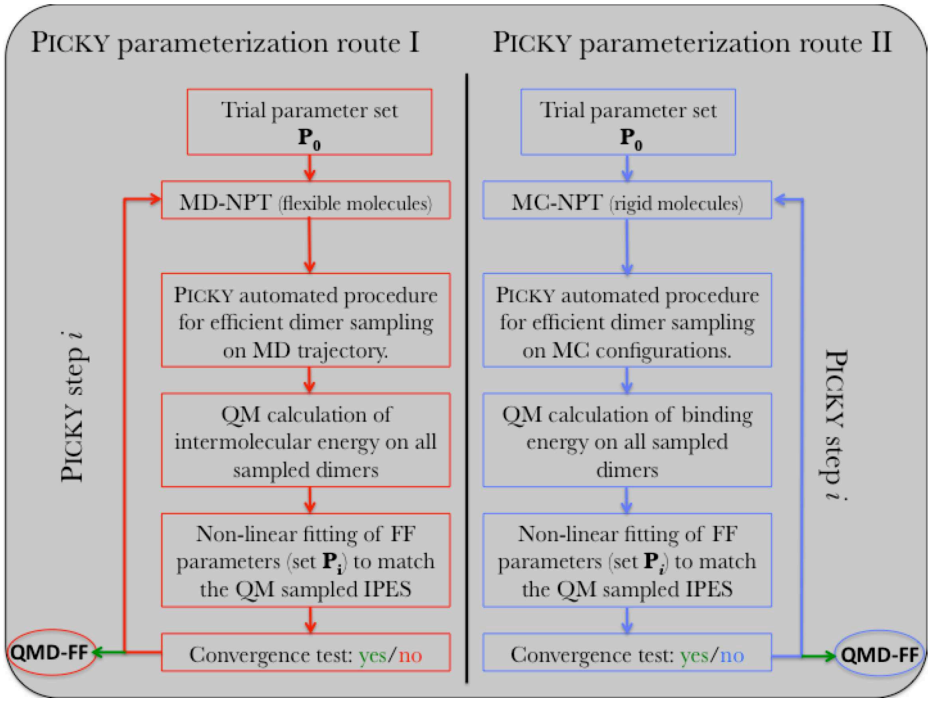


Figure 2: PICKY parameterization routes I (left) and II (right). Route I was adopted in Ref.,[19] whereas route II is implemented in the present work.

the monomers forming the sorted dimers are displaced from the equilibrium geometry obtained *in vacuo*. As a consequence, the calculation of ΔE^{inter} over such dimers should be taken with care, as the definitions of intermolecular and binding energy differ. Indeed, the binding energy ΔE^{bind} between a pair of molecules A and B is defined as

$$\Delta E^{bind} = E_{AB} - (E^0_A + E^0_B) \tag{2}$$

where E_{AB} is the dimer total energy and $E^0_{A/B}$ is the energy of the isolated A/B monomer

at its equilibrium geometry, obtained *in vacuo*. The intermolecular energy ΔE^{inter} computed during PICKY route I is instead

$$\Delta E^{inter} = E_{AB} - (E_A + E_B) \tag{3}$$

where E_{AB} is again the dimer energy, but $E_{A/B}$ is the energy of the isolated monomer A/B at the geometry adopted in the dimer. Therefore, from equations (2) and (3),

$$\Delta E^{inter} = \Delta E^{bind} - (\Delta E^{intra}_A + \Delta E^{intra}_B) \tag{4}$$

with

$$\Delta E^{intra}_M = E_M - E^0_M \quad ; \quad M = A, B \tag{5}$$

Hence, when computed over a set of MD sampled dimers, ΔE^{inter} includes a relaxation/distortion term $\Delta E^{intra}_{A/B}$ of monomer A (or B), due to the AB complex formation, which depends explicitly on the monomers internal coordinates. Given the FF partition described by equation (1), and considering that the $E^{FF_{intra}}$ contribution is derived from QM data computed on the isolated monomers (see also Supporting Information), it would be preferable to parameterize the $E^{FF_{inter}}$ over a QM IPES which does not include any source of intramolecular energy, as the $\Delta E^{intra}_{A/B}$ term.

The alternative route II proposed here satisfies this requisite, as it is based on MC simulations, performed with trial moves that only alter the molecule's relative spatial disposition, without attempting to displace their internal geometry. Since the latter is fixed to the QM equilibrium geometry of the isolated monomers (the same employed for the JOYCE intramolecular parameterization), $\Delta E^{intra}_{A/B}$ vanishes, causing ΔE^{inter} and ΔE^{bind} to coincide.

PICKY 's route II is illustrated in the right panel of Figure 2, and can be summarized as follows:

α) A reliable starting set of intermolecular parameters is assigned to the target molecule and a MC simulation is carried out in the isothermal-isobaric ensemble (NPT), in standard conditions (usually 298 K and 1 atm). Molecules may translate and rotate, but no change in the initial internal geometry of each monomer is permitted.

β) Following the PICKY automated procedure,[19] several dimers are extracted from the equilibrated MC conformation, and used for a QM calculation of ΔE^{inter}.

γ) The QMD-FF parameters are obtained by minimizing the difference between the QM computed dimer energies (ΔE^{inter}) and those obtained by means of QMD-FF intermolecular term ($E^{FF_{inter}}$). Differently from route I, both terms only depend from the distance between monomers and their relative orientation, whereas no explicit dependence of the system energy over internal coordinates is contained in the $E^{FF_{intra}}$ and $\Delta E^{intra}_{A/B}$, which are not involved in the parameterization route.

δ) The resulting QMD-FF is again employed in MC runs and the procedure iteratively repeated until some convergence *criterion* is satisfied.

Finally, as evidenced in the right panel of Figure 1, the final QMD-FF intermolecular parameters are eventually coupled to the JOYCE intramolecular term and MD simulations can be carried out for a more realistic flexible system.

3 Computational Details

3.1 Intramolecular FF

The intra-molecular FF term $E^{FF_{intra}}$ for the benzene molecule was set once and for all by the JOYCE protocol, as described in the Supporting Information. In consideration of benzene's aromatic structure, the JOYCE procedure[17,18] was applied by choosing a totally harmonic FF, based on stretching, bending and harmonic dihedrals (see equations (S1)-(S4) in the Supporting Information). The only required QM data in this case are the benzene optimized geometry *in vacuo* and the corresponding Hessian matrix. Both these quantities were computed through a suitable DFT functional, namely B3LYP coupled with the Dunning's correlation consistent basis set, cc-pVDz. Both geometry optimization and Hessian calculations were performed with the GAUSSIAN09 package.[79] The same software was employed for all the intermolecular QM calculations described in the following.

3.2 Intermolecular FF

The intermolecular energy ΔE^{inter} between a pair of benzene molecules was computed, according to equation (3), with both WFT and DFT methods.

In the first case, as already done in previous parameterizations performed by us on aromatic molecules,[4,19,30] we resort to Hobza's suggestion[65] to describe non-covalent interactions through second order Möller-Plesset perturbation theory (MP2), yet coupled with a purposely modified basis sets. Indeed, the adoption of rather small basis sets (*e.g.* 6-31G* or cc-pVDz), whose heavy atom polarization exponents have been modified with respect to their original values, was found[4,19,65,80,81] to correct the well known over-binding found for aromatic complexes when MP2 is employed with large basis sets, confirming the critical role of the basis set in such kind of calculations, recently reviewed by Hobza and coworkers .[82] For the benzene dimer, the polarization exponent for the Carbon atoms was set to 0.25,[65] and the resulting basis set therefore labeled 6-31G*(0.25). Finally, considering the reduced dimensions of the basis set employed, the Counterpoise (CP) correction[83] was applied to handle the basis set superposition error (BSSE).

As far as the DFT calculations are concerned, several combinations of functionals/basis sets were tested. In first place, the same level of theory employed in the intramolecular parameterization was considered. To take into account dispersion interactions, the D3 empirical correction term proposed by Grimme[84] was applied to the chosen functional (therefore labeled B3LYP-D3/cc-pVDz). As for MP2, the CP correction was applied in consideration of the small basis set adopted. Next, the M06-2X[85] functional was tested, coupled with the larger 6-311+G(2d,2p) basis set. Following Truhlar's suggestions for non-covalent complexes,[85] the CP correction was not applied in this case. Finally, the BLYP-D3 and CAM-B3LYP-D3 functionals were also tested, again employing the 6-311+G(2d,2p) basis set. In all cases where the D3 correction was applied, the original parameters developed by Grimme[84] and implemented in the GAUSSIAN09code[79] were employed. Since these parameters were tuned with the large def2-QZVP basis set (which could not be used herein in view of the large number of dimer calculations to be performed), a preliminary validation of the quality of the results obtained with the smaller cc-pVDz and 6-311+G(2d,2p) basis sets was carried out. Although in the original

paper[84] the dispersion correction was parameterized specifically for calculations without CP correction, in consideration of the smaller basis sets here employed all calculations were carried out with and without applying the CP scheme. The results (see Figure A in the Supporting Information) indicate that all investigated energy curves, obtained by applying the CP correction to the results obtained with the smaller basis sets are very close to those i computed with the original procedure (*i.e.* with the very large def2-QZVP basis set).

A standard formula is adopted to express the QMD-FF intermolecular term $E_{AB}^{FF_{inter}}$ reported in equation (1).

$$E_{AB}^{FF_{inter}}(\bar{r}_{AB}) = \sum_{i=1}^{N_A}\sum_{j=1}^{N_B}[E_{ij}^{LJ}(r_{ij}) + E_{ij}^{Coul}(r_{ij})] \tag{6}$$

where N_A (N_B) is the number of interaction sites of A (B) molecules and i (j) is the i^{th} (j^{th}) site of the molecule A (B). Finally, E_{ij}^{LJ} and E_{ij}^{Coul} are respectively a 12-6 Lennard-Jones (LJ) potential and the charge-charge interaction, *i.e.*

$$E_{ij}^{LJ}(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad ; \quad E_{ij}^{Coul}(r_{ij}) = \frac{q_i q_j}{r_{ij}} \tag{7}$$

The values of the parameters (LJ and charges) defining equation (7) are obtained according to the PICKY original protocol,[19] *i.e.* by minimizing the functional

$$I^{inter} = \frac{\sum_{k=1}^{N_{geom}}[(\Delta E_k^{inter} - [E_{AB}^{FF_{inter}}]_k)^2]e^{-\alpha\Delta E_k^{inter}}}{\sum_{k=1}^{N_{geom}} e^{-\alpha\Delta E_k^{inter}}} \tag{8}$$

where the index k runs over the N_{geom} considered dimer considered, while α a Boltzmann-like weight. To systematically compare the different benchmarked functionals, the PICKY protocol is adopted in all parameterizations as described in the following:

Step 0: A MC run is carried out in the NPT ensemble on a system of 216 benzene molecules, at 1 atm and 298 K. The OPLS force-field parameters[3,63] are chosen as starting set.

Step 1: From the MC equilibrated final conformation 50 different dimers are initially extracted, according to the original PICKY sampling algorithm,[19] among those whose interaction energy $E^{FF_{inter}}$ is less than 5 kJ/mol. The QM interaction energy ΔE^{inter}

is computed for all selected dimers as described above. Thereafter, the non-linear fitting, performed by means of equation (8), is carried out allowing each parameter to vary only by 10% of its initial value. The convergence *criterion* is evaluated by estimating ΔP_1 (see equation (S7)) over a grid of 10^6 points. The new set of intermolecular parameters is eventually employed to carry out a new MC-NPT simulation run, in the same thermodynamic conditions as in Step 0.

Step 2: Two different filters are applied in the dimer extraction from the final configuration of the previous MC run. The first 50 pairs are selected (attractive sampling) applying the usual PICKY algorithm on dimers whose interaction energy is less than -1.0 kJ/mol and whose distance is less than 10 Å. Additionally (repulsive sampling) other 50 dimers are extracted by imposing their interaction energy to be larger than -1.0 kJ/mol and their distance less then 8 Å. ΔE^{inter} is computed with the chosen functional for all 100 dimers, and the resulting values added to the QM database created in the previous step. The latter (*i.e.* 150 dimers) is employed in the QMD-FF parameterization, again performed constraining the maximum allowed variation to 10%. The convergence is again checked through ΔP_2, on a 10^6 point grid. The MC run with the new parameters is finally carried out.

Step 3: Same as step 2, but the maximum allowed variation during the parameters fitting is 20%.

Step 4: Same as step 2, but the maximum allowed variation during the parameters fitting is 100% and the final run is performed at 1000 atm and 400 K, to enhance the population of rather repulsive conformers.

Step 5: Same as step 2, but the maximum allowed variation during the parameters fitting is 100%.

Step n: Step 5 is iteratively repeated until convergence on ΔP_n is reached.

All details concerning MC runs, as well MD simulations are reported in the Supporting Information.

4 Results and Discussion

4.1 Preliminary tests

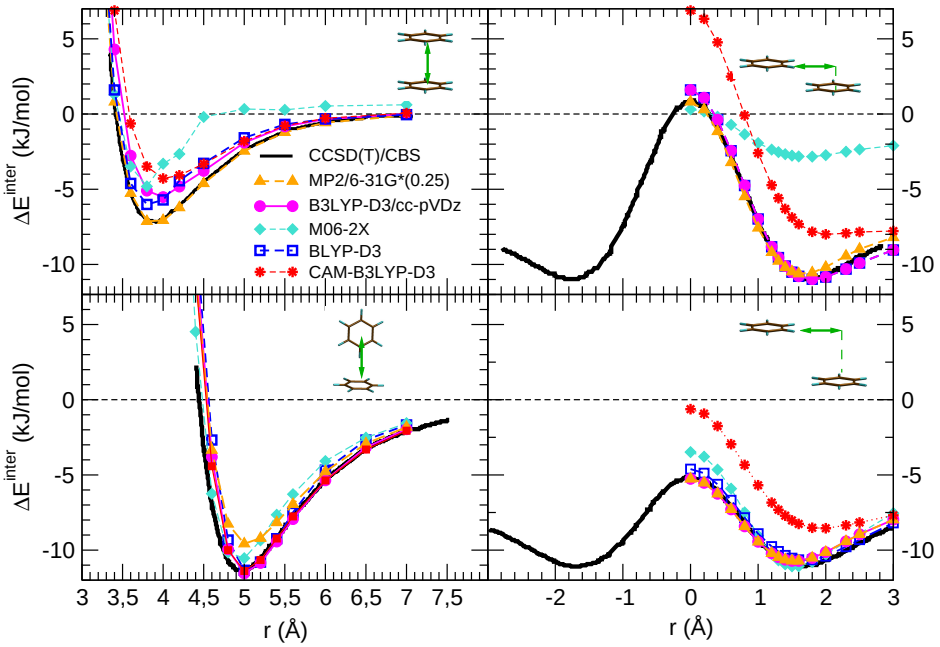


Figure 3: Selected IPES's cross sections of the considered WFT and DFT methods for different arrangements of the benzene dimer: FtF (top-left), TS (bottom-left) and PD (in two different inter-plane separations, 3.4 Å and 3.6 Å, top and bottom right panels. respectively). Unless otherwise reported in the legend, next to the label of the method, the 6-311+G(2d,2p) basis set was employed in DFT calculations. CCSD(T)/CBS reference values were taken from ref [78].

For a preliminary assessment of the performances of the considered WFT and DFT methods, ΔE^{inter} was computed for different benzene arrangements and compared with the high quality CCSD(T)/CBS reference values reported in Ref [78]. All IPES cross sections are reported in Figure 3 for some specific arrangements as displayed in the insets. In particular the face-to-face (FtF), T-shaped (TS) and parallel displaced (PD) dimer geometries are considered. Despite most of the considered methods were carried out with

and without applying the CP corrections, for the sake of clarity only the curves closer to the reference values are displayed in Figure 3. More specifically, the CP correction is applied to all methods except the M06-2X functional.

In all reported cases, the most stable conformers correspond to TS and PD geometries whose energy is about twice that of the FtF arrangements. From the comparison with the reference CCSD(T)/CBS curves, it appears how the accuracy of all methods is generally better on geometries than on interaction energies. Indeed the energy minima are found at similar distances for all methods, whereas the energy curves show different vertical shifts with respect to the reference. Major discrepancies are evident for the M06-2X functional for one PD arrangement and even for FtF. Also CAM-B3LYP-D3 seems to be rather inaccurate for PD arrangements and to lesser extent in the FtF arrangements. The remaining curves are rather close to the reference one and appear to be about of the same quality. By considering its very low computational cost, it is noteworthy noticing the accuracy of the MP2/6-31G*(0.25) curves.

The second qualitative test was inspired by two recent experimental[59] and theoretical[5]

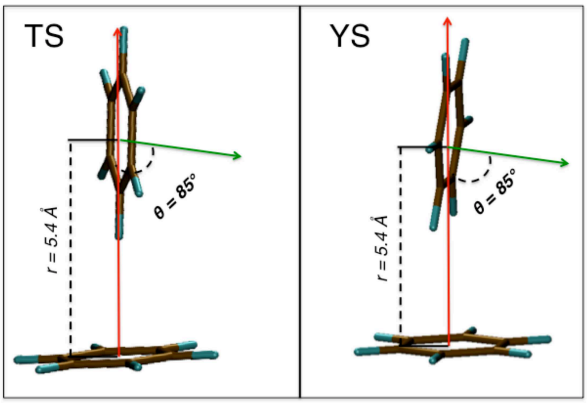


Figure 4: T-shaped and Y-shaped geometries employed in a preliminary test.

investigations which allowed for a deep insight onto the structure of liquid benzene. Both studies agree in indicating the T-shaped (TS) and Y-shaped (YS) arrangements as the most probable conformations in liquid phase for $r \sim 5.4$ Å. These statistically relevant geometries (see Figure 4) were thus used to compare the intermolecular energy computed at different level of theory. The results are reported in Table 1.

method	basis set	ΔE^{inter} (kJ/mol)	
		TS	YS
CCSD(T)	CBS	-9.30	-9.15
MP2	6-31G*(0.25)	-8.17	-8.37
B3LYP-D3	cc-pVDz	-9.50	-9.58
M06-2X	6-311+G(2d,2p)	-7.62	-6.65
BLYP-D3	6-311+G(2d,2p)	-9.25	-9.21
CAM-B3LYP-D3	6-311+G(2d,2p)	-9.21	-8.79

Table 1: Intermolecular energies ΔE^{inter}, computed at different levels of theory, for the TS and YS arrangements displayed in Figure 4.

According to CCSD(T)/CBS calculations (see Supporting Info for technical details), the interaction energies of the two conformers are similar, being the TS conformer slightly more stable. The B3LYP-D3 and BLYP-D3 functionals slightly overestimate (with average errors of 0.3 kJ/mol and 0.05 kJ/mol, respectively) the reference binding energy, whereas CAM-B3LYP-D3 shows a very small (0.2 kJ/mol in average) underestimation. Overall, all D3 corrected methods result in a good agreement with the CCSD(T)/CBS values, whereas larger underestimates (up to $\sim$2.5 kJ/mol) are observed for M06-2X. The MP2 results are still acceptable, especially considering the reduced dimensions of the employed basis set. However, these preliminary tests can give no more than a qualitative picture of the several methods. More robust assessments about their capability to give accurate two-body energies should be based, rather than on few dimer geometries, on benchmarks performed over more representative and larger regions of the IPES, as detailed in the following.

4.2 Route II validation

Route II parameterization protocol was validated adopting initially the MP2/6-31G*(0.25) method for the two-body energies, which was already considered by our group for parameterization of benzene's FF.[4]

To appreciate the differences between the new route II (using MC) and the old route I (using MD) both strategies were here followed (see Ref. [19] and Supporting Information for details). The convergence of both iterative routes I and II was monitored through

step	dimers	Route I				Route II			
		χ (kJ/mol)	ΔP (kJ/mol)	ρ (kg/m^3)	ΔH^{vap} (kJ/mol)	χ (kJ/mol)	ΔP (kJ/mol)	ρ (kg/m^3)	ΔH^{vap} (kJ/mol)
1	50	0.34	3.6	893	35.0	0.42	1.2	886	35.8
2	150	0.64	5.2	928	38.0	0.61	1.9	908	36.2
3	250	0.69	0.2	929	37.9	0.53	0.1	908	36.1
4	350	0.75	1.8	883*	-	0.53	0.9	916*	-
5	450	0.88	0.4	918	36.0	0.66	1.5	900	34.3
6	550	-	-	-	-	0.84	1.2	917	35.1
7	650	-	-	-	-	0.87	0.2	918	35.3

* simulation performed at 1000 atm and 400 K

Table 2: Monitored properties obtained during PICKY parameterization steps following route I and route II. Fitting standard deviation (χ), convergence *criterion* (ΔP), mass density (ρ) and vaporization enthalpy (ΔH^{vap}). In both cases the intermolecular energy ΔE^{inter} was computed at MP2/6-31G*(0.25) level.

the ΔP index (see equation (S7) in the Supporting Information), which accounts for the global differences between the QMD-FF IPESs of two consecutive parameterization steps. ΔP and other significant quantities arising from the MD/MC simulations, are reported in Table 2 for each PICKY step. The final intermolecular parameters, achieved with both routes, are reported in Table B in the Supporting Information.

With a convergence threshold for ΔP of 0.5 kJ/mol, both routes reach convergence in few steps, namely 5 and 7 for route I and II, respectively. Therefore, in the final step the total number of sampled dimers which have required a QM calculation is 450 for route I and 650 for route II. Owing to these large numbers and in view of future application to more complex molecules, it is clear that the QM method should be affordable. The standard deviation of the fitting, χ, increases from an initial value of ~ 0.4 kJ/mol to ~ 0.9 kJ/mol for both routes, indicating that the sample variety increases at each PICKY step. The most important result is that the two routes lead to similar results, and end up with very similar parameters (Table B). On the one hand this confirms the robustness of PICKY parameterization, on the other hand it suggests that, due to benzene scarce flexibility, the $\Delta E^{intra}_{A/B}$ distortion/relaxation term entering equation (4) plays in this case a minor role.

In Table 3 the final average mass density (ρ) and vaporization enthalpy (ΔH^{vap}) are

FF	type	n. of LJ par.	ρ (kg/m^3)	ΔH^{vap} (kJ/mol)
PICKY QMD-FF (route I)	QM derived	2	918 ± 4	35.9 ± 0.2
PICKY QMD-FF (route II)	QM derived	2	919 ± 4	35.7 ± 0.2
previous[4]	QM derived	3	911 ± 5	36.9 ± 0.4
OPT-FF[5]	QM derived	2	1045 ± 6	47.3 ± 0.1
AMBER03[5]	empirical	2	836 ± 5	30.3 ± 0.1
GAFF[5]	empirical	2	852 ± 8	31.6 ± 0.1
OPLS-AA[5]	empirical	2	867 ± 4	33.6 ± 0.1
OPLS-CS[5]	empirical	2	947 ± 3	62.2 ± 0.1
CHARMM27[5]	empirical	2	870 ± 6	34.2 ± 0.1
GROMOS 53A5[5]	empirical	2	887 ± 2	35.3 ± 0.1
GROMOS 53A6[5]	empirical	2	882 ± 3	34.5 ± 0.1
exp.	-	-	874[50]	33.9[52]

Table 3: Selected properties of benzene condensed phase, obtained with MD simulations at 1 atm and 298 K performed with different FFs. The first three rows refer to MP2/6-31G*(0.25) based parameterizations. The first and second rows are from the present work, whereas the third row, labeled "previous", is taken from Ref.[4], where a 3-parameter LJ potential and different sampling strategy were employed. Experimental values[50,52] are reported for comparison, together with literature data obtained with different FFs in Ref. [5].

compared with their experimental counterparts as well as with other simulation results, obtained by us[4] and by Fu and Tian with a variety of FFs.[5] It may be worth pointing out that the quantities referred to route II and reported in Table 3 were obtained by MD simulations performed with intermolecular FF parameters obtained through the new PICKY protocol here discussed and intramolecular ones derived through JOYCE (see Supporting Information). This includes molecular flexibility which was not considered along the PICKY steps of route II. As a consequence, both density and vaporization enthalpy slightly differ from the corresponding final values reported in Table 2. Both PICKY results are in good agreement with the experiment, being the error on ρ and $\Delta H^{vap} \sim 5\%$ and $\sim 6\%$, respectively. The third row of Table 3 reports data from a previous work of us[4] where the FF was parameterized solely from *ab initio* data computed at MP2/6-31G*(0.25) level, as in the present work. That work differs from the present one because it uses a modified 3-parameter LJ function[4] and a set of QM sampled geometries ($\sim$ 200 dimers) *a priori* chosen based on chemical intuition, *i.e.* without the systematic search for the statistically significant arrangements, that characterize the PICKY procedure. The

MD simulations of Ref. [4] reproduced with good accuracy many thermodynamic, structural and dynamic properties of benzene's crystal and liquid phases, being the largest errors an overestimation of ρ and ΔH^{vap} of $\sim 4\%$ and $\sim 9\%$, respectively.

In a recent work, Fu and Tian[5] have benchmarked the performances of several popular force-fields on liquid benzene. They have also performed MD-NPT simulations with a FF (named OPT-FF) obtained by directly fitting the few interaction curves of benzene dimer at CCSD(T)/CBS level, reported in literature[78,86] and in Figure 3. The resulting OPT-FF data reported in Table 3 show a remarkable error, much larger than we found in Ref. [4]. By considering that the MP2/6-31G*(0.25) curves displayed in Figure 3 are close to the reference CCSD(T)/CBS ones, this different accuracy might seem somehow surprising. This can be seen as a further indication that FF parameterizations and, more in general, benchmark evaluations of the capability of different methods to estimate interaction energies, should be performed on a rather wide variety of dimer arrangements (~ 200 in Ref. [4] and ~ 500 in this work), whereas the failure reported for the OPT-FF is probably connected to a small sample of the QM IPES, which was represented only by the three investigated arrangements (FtF, TS and PD). The density and enthalpy values obtained by Picky are well within the range of values reported in Ref. [5] for the considered FFs, which yielded a maximum error of $\sim 8\%$ (OPLS-CS) and a minimum one of ~ 1 % (OPLS-AA and CHARMM27). In this context it is also worth recalling that most of the FFs investigated by Fu and Tian are empirically parameterized with intermolecular parameters purposely tuned to reproduce the experimental values of ρ and ΔH^{vap}.

The overestimation of the vaporization enthalpy and of the mass density showed by both Picky routes could be connected to the pure two-body nature of the QMD-FF. Indeed, McDaniel and Schmidt[33] have recently shown that the inclusion of three-body interactions in the FF, diminishes the mass density of $\sim$ 5–6%. In that work, the authors report that, in all investigated cases, this inclusion leads to densities and vaporization enthalpies that differ from the experiment by 2% to 8%, decreasing the errors found when only the two-body contribution to the FF was accounted for, which were about 14%.[33] Thus it seems that, at least for all the systems investigated in Ref. [33], the three body

terms have a global repulsive effect on the bulk system. In the hypothesis that something similar could occur for benzene, the decrease of the mass density expected by adding the three-body terms could improve the QMD-FF results. However, the inclusion of such effects goes beyond the aims of the present work, and will be the object of a future investigation.

4.3 Benchmarking DFT functionals

The QMD-FF protocol here validated was implemented with the aim of developing specific and accurate FFs for large, flexible molecules, whose properties are not included in the training sets employed in most popular (transferable) FFs. Example of this kind of targets may stand in liquid crystalline compounds, organic dyes, functionalized polymers, organo-metal compounds, molecules in electronically excited states, *etc.* In most of these cases, the computational bottleneck consists indeed in the systematic QM computation of ΔE^{inter}. This scenario rules out very accurate approaches as for instance CCSD(T)/CBS[86] or SAPT,[33,34] whereas the computational feasibility of the DFT based methods is certainly appealing. For this reason several popular DFT approaches will be investigated in the following in their ability to accurately represent aromatic IPES's through the Picky protocol. Despite routes I and II lead to similar results at least for the case considered, route II will be used, as it is more suitable for flexible molecules, and therefore more general than route I.

Picky 's route II was applied to the benzene bulk system, using QM data computed with some popular DFT functionals, corrected in a more or less empirical way, with dispersion energy. The results are reported in Table 4 for each Picky step and for the final simulation performed with the flexible model. As stated below the density and the enthalpy along the Picky steps were obtained by MC calculations with frozen internal geometry, whereas the same properties, reported in the last row of Table 4, are calculated by a final MD simulation with no geometrical constraint, performed by coupling the Picky parameters with the Joyce FF. Both types of simulation were performed at 298 K and 1 atm, for systems of 216 and 512 molecules, respectively. The final intermolecular parameters are reported in detail in the Supporting Information, Tables C and D. As

step	B3LYP-D3				M06-2X			
	χ (kJ/mol)	ΔP (kJ/mol)	ρ (kg/m^3)	ΔH^{vap} (kJ/mol)	χ (kJ/mol)	ΔP (kJ/mol)	ρ (kg/m^3)	ΔH^{vap} (kJ/mol)
1	0.3	5.1	933	40.2	0.72	11.2	796	24.5
2	0.5	0.9	948	40.2	1.10	3.83	804	25.0
3	0.5	2.1	953	41.1	1.02	0.78	802	28.6
4	0.6	0.1	916[b)]	-	0.87	0.88	805[b)]	-
5	0.7	2.3	931	39.0	0.90	0.06	780	23.8
6	0.7	0.1	930	39.0	-	-	-	-
final[a)]	-	-	939 ± 5	39.9 ± 0.2	-	-	795 ± 8	24.3 ± 0.3

step	BLYP-D3				CAM-B3LYP-D3			
	χ (kJ/mol)	ΔP (kJ/mol)	ρ (kg/m^3)	ΔH^{vap} (kJ/mol)	χ (kJ/mol)	ΔP (kJ/mol)	ρ (kg/m^3)	ΔH^{vap} (kJ/mol)
1	0.31	0.72	934	39.8	0.22	4.0	914	37.1
2	0.67	1.56	975	43.3	0.53	0.9	911	36.5
3	0.66	0.94	968	43.0	0.55	1.2	919	36.4
4	0.65	0.12	920[b)]	-	0.60	0.5	886[b)]	-
5	0.68	0.20	960	41.3	0.72	0.4	923	36.2
final[a)]	-	-	960 ± 6	41.7 ± 0.3	-	-	926 ± 5	36.8 ± 0.2

[a)] final run performed accounting for flexibility through JOYCE intramolecular FF

[b)] simulations performed at 1000 atm and 400 K

Table 4: Monitored properties obtained during PICKY parameterization route II coupled with the investigated functionals. Fitting standard deviation (χ), convergence *criterion* (ΔP), mass density (ρ) and and vaporization enthalpy (ΔH^{vap}).

apparent from Table 4, the four functionals lead to rather different results, both along the PICKY steps and for the final value obtained by MD simulations. In particular three functionals lead to an overestimation of the density, whereas the M06-2X gives rise to a rather low density. The errors on the vaporization enthalpy are consistent with those of the density, confirming that the overestimation of both quantities is a signature of a too attractive intermolecular potential energy.

These results somehow are in qualitative accord with the data reported in Table 1, that show the intermolecular energy for the TS and YS conformers. Assuming again that the CCSD(T) results are close to the exact values, it appears that, as far as the DFT energies are concerned, there is a correspondence between the overestimation of the interaction energy and the same for both the density and enthalpy. The opposite occurs for M06-2X

which underestimate both intermolecular energy and thermodynamic properties. However a one-to-one correspondence between the accuracy of the intermolecular energy for the TS/YS conformers and that of the final thermodynamic properties cannot be established. Indeed, the best agreement among the tested D3 corrected functionals, with respect to reference CC values, is found for the two conformers with BLYP-D3, whereas the most accurate estimates of the benzene thermodynamics are obtained with the CAM-B3LYP-D3 based FF. This observation suggest that the quality of the intermolecular energies in a limited number of dimer arrangements, even if statistically significant, can not provide a robust *criterion* to predict the quality of the thermodynamic properties.

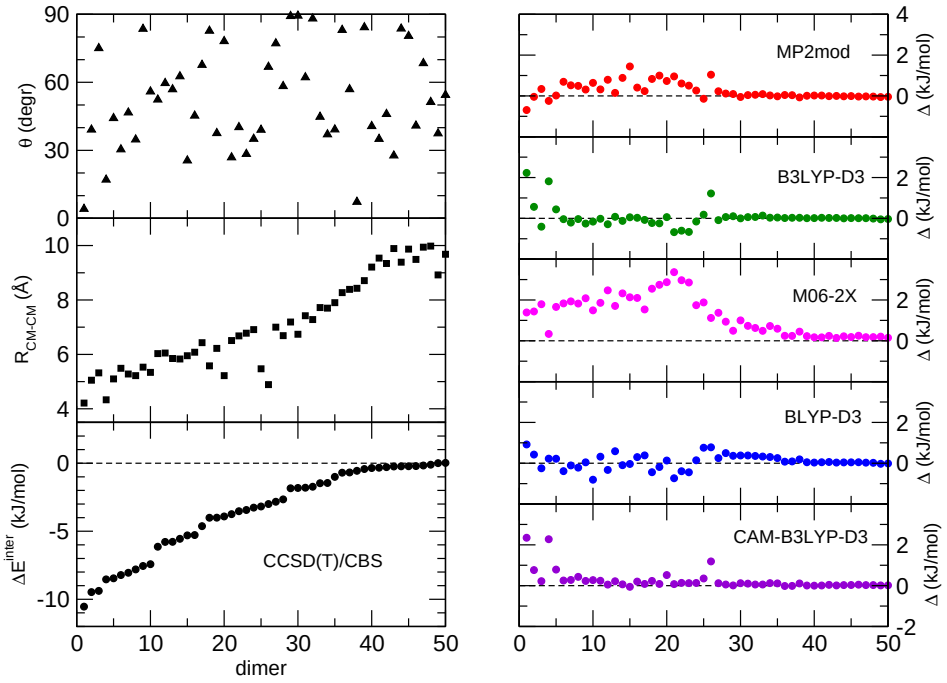


Figure 5: Left panel: interaction energies (ΔE^{inter}, computed at the CCSD(T)/CBS level), distance between the centers of mass of the two monomers (R_{CM-CM}) and angle between the vectors normal to the ring planes (θ) computed for the 50 dimers sampled in PICKY Step 1. Right panel: differences (Δ) between the interaction energies computed at CCSD(T)/CBS level and with the chosen lower level method (MP2 or DFT).

Nonetheless, it would be desirable to establish an *a priori* procedure to evaluate the

accuracy of the DFT functionals without performing the whole PICKY procedure and the consequent simulations. To this end, one may exploit the fact that in all parameterizations the first 50 dimers selected in Step 1, come from the same initial MC run and provide a statistically relevant collection of different dimer arrangements. To evaluate DFT performances in a more significant way, ΔE^{inter} was compared with that purposely computed at CCSD(T)/CBS level for all these 50 dimers. Results are reported in Figure 5: in the left panel, from bottom to top ΔE^{inter} (reference CCSD(T) values), the distance R_{CM-CM} between the centers of mass of the two monomers and the angle θ (see also Figure 4) are displayed for each conformer. The right panel displays the DFT and MP2 energies minus the reference ones, so that positive numbers correspond to energies higher (less attractive) than CCSD(T)/CBS ones. According to the left-bottom panel of Figure 5, the dimers are arranged for increasing energy which roughly correspond to an increasing distance between centers of mass. It appears that the distributions of both R_{CM-CM} and θ confirm the ability of the PICKY methodology in sampling a wide variety of arrangements.

The first general comment concerns with the distributions of the energy deviations versus the intermolecular distance. Supposing again that the CCSD(T)/CBS data are very close to the exact values, is it evident that the errors are higher at low distances and involve more or less all DFT functionals as well as MP2. Within the D3 corrected methods, it is apparent as the CAM-B3LYP-D3 functional yields the best results. The largest errors are found for dimers 1 and 4 and 26, which though correspond to PD-like arrangements at $R_{CM-CM} \simeq 4$, whose statistical weight is expected[5,59] to be moderate. There are some resemblances between the B3LYP-D3 and the CAM-B3LYP-D3 data, but the former shows a general best adherence to CCSD(T) data. The BLYP-D3 deviations are rather scattered and it is in general difficult to establish whether it is too attractive or too repulsive. The M06-2X energies are clearly too small in the attractive region probably from and underestimation of the dispersion energy, and are consistent with the low values of the density and vaporization enthalpy reported in Table 4 for this functional. Finally MP2 data seem to be less attractive than CCSD(T) data, even if this does not occur at very low intermolecular distances where the deviations are negative.

In conclusion, the data reported in Figure 5 provide a look over a squeeze of relevant dimer arrangements and allow to draw some conclusions about the quality of the DFT and MP2 intermolecular energies. Although this comparison is clearly more significant than the previous ones, based on *a priori* selected dimer arrangements, the conclusions one can draw about the final thermodynamic properties are still uncertain for energy errors scattered as in the considered cases.

4.4 Benzene condensed phase properties with QMD-FF

Based on the results of the previous section, the best DFT derived FF is the one parameterized using the CAM-B3LYP-D3 functional. The parameters of this FF are reported in Table 5 (for a list of all final parameters and fitting standard deviations see Table D in the Supporting Information). It might be worth recalling that, as the QMD-FF parameters were derived by minimizing the functional (8) with no constraint, the ϵ, σ and q parameters entering equation (6) should be considered merely as fitting parameters. Therefore, any attempt to attribute some physical meaning to a single contribution of the terms listed in (7), as for instance assigning the dispersion energy to the $\frac{1}{R^6}$ term alone, should be taken with care.

Atom	QMD-FF			QM charges	
	ϵ (kJ/mol)	σ (Å)	q (a.u.)	q_{vac} (a.u.)	q_{PCM} (a.u.)
C	0.390	3.52	-0.143	-0.128	-0.136
H	0.027	2.39	0.143	0.128	0.136

Table 5: Final intermolecular parameters obtained with PICKY parameterization route II based on CAM-B3LYP-D3//6-311+G(2d,2p) . In the last two columns RESP charges, obtained at the same level of theory, are reported for the benzene monomer *in vacuo* (*vac*) and in the neat liquid (*PCM*).

Nonetheless, for a first estimate of the physical relevance of the QMD-FF, it can be useful to compare the atomic charges obtained with the present parameterization with the ones obtained by fitting the QM computed electrostatic potential of a single monomer. To this end, the atomic charges were computed through the restrained electrostatic potential (RESP) procedure,[87] using the charge density determined at CAM-B3LYP-D3//6-311+G(2d,2p) level. Two separate calculations were conceived, in which the benzene

monomer is considered either *in vacuo* or embedded in a benzene solution, employing the polarizable continuum model (PCM).[88] The resulting charges are reported in the last two columns of Table 5. It appears that the QMD-FF charges are closer to the ones computed in PCM than to those obtained *in vacuo*. Since the former set takes into account the polarization effects due to the surrounding medium, the similar values obtained in QMD-FF can be seen as a first indication of the FF quality.

For a more robust assessment of the CAM-B3LYP based QMD-FF performances, the resulting benzene liquid phase was further characterized by computing structure, thermodynamic and dynamic properties. To better assess the reliability of the employed QMD-FF and, consequently, of its reference DFT functional, all the results are compared both to avaliable experimental data and to the outcomes of several popular FFs, reported in Ref. [5].

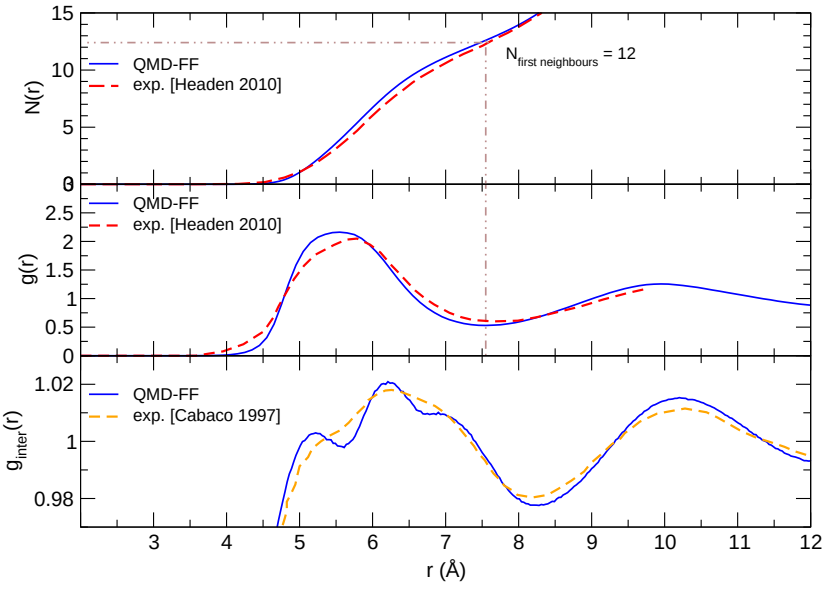


Figure 6: Pair correlation functions ($g_{inter}(r)$, bottom and g(r), middle panel) and number of neighboring molecules (N(r), top panel), computed from simulations performed with the QMD-FF (parameterized over CAM-B3LYP), are compared to their experimental counterparts, taken from Refs. [59] and Ref. [56]. Dotted brown lines evidence the number of neighbors within the first shell.

The structure of liquid benzene was investigated by radial, axial and axial-radial distri-

bution functions. In Figure 6, the $g_{inter}(r)$ and $g(r)$ radial correlation functions, computed between pairs of benzene molecules, are compared to their experimental counterparts. While the latter is the standard pair correlation function between molecular centers of mass,[1,2] the former can be derived averaging[4,56] the atom-atom pair correlations as

$$g_{inter}(r) = \frac{g_{CC}(r)}{4} + \frac{g_{HH}(r)}{4} + \frac{g_{CH}(r)}{2} \tag{9}$$

The agreement between experimental[56] and computed $g_{inter}(r)$ (bottom panel) is excellent: both functions display two rather broad bands, centered at $\sim$ 6 and 10 Å, roughly corresponding to the first two solvation shells. Interestingly, computed results also enhance a small shoulder, around 5 Å, also present, but less marked, in the experimental line. Similarly, the computed and experimental[59] $g(r)$ reported in the middle panel of Figure 6 are very close, and agree with $g_{inter}(r)$ in the description of the first and second neighbor shells. Finally, the integration of the computed $g(r)$ functions up to 7.5 Å, *i.e.* approximatively the radius of the first solvation shell, yields a number of first neighbors (12) in quantitative agreement with that reported in Ref. [59]. The structure of the neat liquid

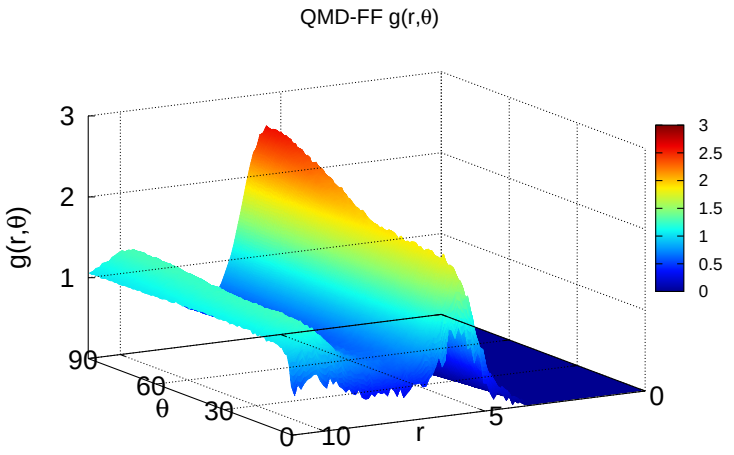


Figure 7: Angular-radial distribution function, $g(r, \theta)$, computed over the MD-NPT trajectory, at 1 atm and 298 K, performed with the QMD-FF parameterized through the CAM-B3LYP-D3 functional. Distances are reported in Å, whereas the θ angle is displayed in degrees.

can be further unraveled by considering the distribution on the angle θ (see Figure 4 and

also Refs. [59] and [5]). The scenario emerging from the analysis of the angular-radial distribution function, $g(r, \theta)$, displayed in Figure 7, is also in good agreement with the one recently reported (see also Figure 9 of Ref. [59]) by Headen and coworkers, based on high-resolution neutron diffraction measurements.[59] In fact, the employed QMD-FF not only succeeds in reliably accounting for the most probable θ ($\sim 90^\circ$, red peak in Figure 7), but also well describes the increase of stacked conformers (θ between 0° and 20°) at shorter distances (cyan region below 5 Å). These orientational features become

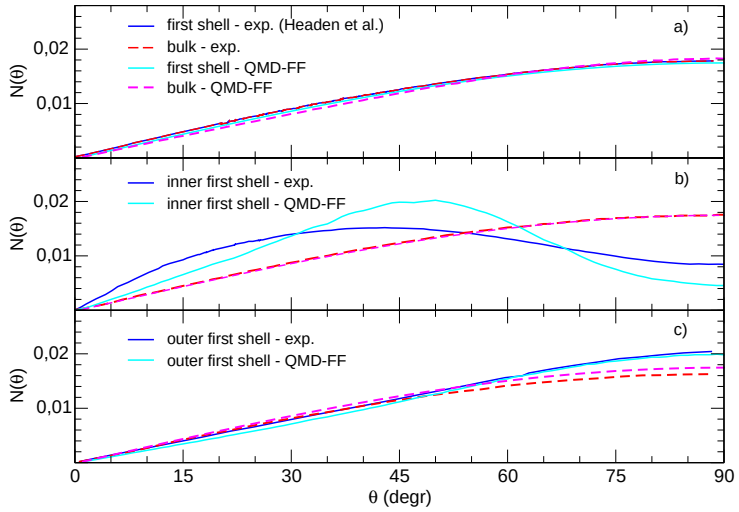


Figure 8: Number of benzene molecules, within selected regions, as a function of the angle θ between the benzene molecular planes. a) benzene molecules within the whole first coordination shell ($0 < r <$ 7.5 Å, solid blue and cyan lines, for experimental and QMD-FF computed results, respectively) *vs* the distribution achieved in the isotropic bulk (dashed red and magenta lines for experimental and QMD-FF computed results, respectively). b) benzene molecules within the inner part of first coordination shell ($0 < r < r^*$, with $r^* = 5.0$ and 4.75 Å, for experimental and QMD-FF computed results). c) benzene molecules within the outer part of first coordination shell ($r^* < r <$ 7.5Å).

even clearer by considering the axial distribution alone, which is plotted in Figure 8 and compared to the curves derived from the experiment.[59] Following Headen suggestions, the number of benzene molecules as a function of θ is computed in three different regions of the bulk (the whole first neighbor shell and its inner and outer parts) and in the bulk itself. A picture again very similar to the one drawn by the experiment appears, even if the most probable angle at short distances is somewhat larger than that suggested by the

experiment. Given the similarity ($\chi \sim 0.7$ kJ/mol) between the adopted QMD-FF and its reference DFT IPES, the above agreement with the experimental structural features *a posteriori* confirms the quality of the description of aromatic interactions achieved with the CAM-B3LYP-D3 functional.

However, as correctly pointed out by Fu and Tian,[5] the reliability of a model should be assessed not only based on structure, but also on the capability to reproduce thermodynamic behavior. To this end, bulk density, vaporization enthalpy and specific heat at constant pressure (c_P) have been computed at 1 atm and 298 K (details are given in the Supporting Info, equations (S8)-(S10)), and compared in Table 6. The errors on

FF	ρ (kg/m^3)	ΔH^{vap} (kJ/mol)	c_P (J mol^{-1} K^{-1})
GROMOS53A6	881	34.5	155.6
CHARMM27	870	34.2	157.3
AMBER03	834	30.3	153.1
GAFF	852	31.6	163.6
OPLS	867	33.6	139.3
OPT-FF	1045	47.3	147.3
QMD-FF	926	36.9	132.8
exp.	874	33.9	135.7

Table 6: Average thermodynamic properties computed at 298 K and 1 atm in this work with the QMD-FF parameterized over the CAM-B3LYP-D3 functional and through different popular FFs, as reported in Ref. [5] Experimental values are also reported in the last raw: density, vaporization enthalpy and specific heat, were taken from Refs. [50], [52], and [64], respectively.

density and vaporization enthalpy are comparable to those exhibited by the cited popular FFs. This is a good results, considering that many of the latter were tuned to reproduced these two properties specifically. Indeed, the experimental specific heat is also nicely reproduced, with an accuracy similar to the best of the empirical FFs (*i.e.* OPLS).

An even more stringent validation on the quality of a FF can be achieved by considering also dynamic properties. The translational diffusion coefficient (D), the re-orientational times of the molecular axes parallel ($\tau_{\hat{C}_2}$) or perpendicular ($\tau_{\hat{C}_6}$) to the aromatic plane and the shear viscosity (η) have been computed from purposely produced NVE trajectories. To not overwhelm the reader with unnecessary information, all computational details concerning the calculation of these quantities are reported in the Supporting Information,

equations (S11)-(S16).

By looking at Table 7, it appears that also the benzene dynamics is described with good accuracy, except the translational diffusion process. Indeed, the re-orientational times $\tau_{\hat{C}_2}$,

FF	D (10^{-9} m^2/s)	$\tau_{\hat{C}_2}$ (ps)	$\tau_{\hat{C}_6}$ (ps)	η_S (10^{-9} m^{-1} s^{-1})
GROMOS53A6	1.78	1.19	1.79	0.660
CHARMM27	1.97	1.10	1.41	0.567
AMBER03	2.79	1.09	1.401	0.364
GAFF	2.33	1.24	1.71	0.431
OPLS	1.97	1.19	1.59	0.547
OPT-FF	0.35	2.53	6.16	3.280
QMD-FF	1.20	1.40	1.90	0.620
exp.	2.20	[0.9-1.3]	-	0.601

Table 7: Dynamic properties computed at 298 K and 1 atm in this work and through different FFs, as reported in Ref. [5] Experimental values are given the last row. From left to right: translational diffusion coefficient D (Ref. [89]), spinning $\tau_{\hat{C}_2}$ and tumbling $\tau_{\hat{C}_6}$ relaxation times (Ref. [69]) and shear viscosity (η, Ref. [90].

connected to the spinning motion, is almost within the experimental range, whereas the slower relaxation of the tumbling motion ($\tau_{\hat{C}_6}$) with respect to spinning is well reproduced. This can be seen as a further confirmation of the well balanced description of the benzene IPES achieved with the QMD-FF. Conversely, the remarkable overestimation of $\tau_{\hat{C}_6}$ found in Ref. [5] for the OPT-FF, is possibly connected to the biased sampling adopted, and to the consequent overestimation of the statistical weight of the stacked interactions, which might be the cause of the enormous slow down ($\tau_{\hat{C}_6} > 6$ ps) of the tumbling motion. The shear viscosity is also in excellent agreement with the experimental value, testifying that the collective motions are well reproduced. Turning to the translational diffusion, the rather large underestimation of the D coefficient is probably connected to the overestimated bulk density, which, though only $\sim 5\%$ higher than the experimental value, seems the major lack of the QMD-FF. Considering the similarity (see Figure 5) of the CAM-B3LYP IPES with the reference CCSD(T) one, the former inaccuracy (as that registered on ΔH^{vap}) could be in principle ascribed to the missing three-body terms, as previously discussed.

Finally, there is another point of strength of QMD-FF parameterization that deserves

being addressed. Since a pure *ab initio* potential does not depend on the applied thermodynamic conditions, the same level of accuracy should be expected from simulations performed at different temperatures and pressures. This does not necessarily hold for empirical "effective" potentials, which are tuned in well determined thermodynamic conditions, outside of which accurate performances cannot be ensured. Since the specific heat at constant pressure is connected to the enthalpy derivative with respect to the temperature, the good agreement showed for c_P in Table 6 can be seen as a first confirmation of the aforementioned QMD-FF feature. To further investigate this issue, additional properties were computed from the variation of some relevant quantities as a function of temperature. Supplementary MD runs, at 280 K and 320 K, were performed on system of 512 benzene molecules at 1 atm, again adopting the QMD-FF. First, the thermal expansion coefficient α was computed from the variation of the volume of the simulation box in the three NPT runs. Next, since diffusion process are expected to follow an Arrhenius behavior, the activation energy of the translational and re-orientational processes can be computed from the dependence on temperature of D and τ's, respectively, as:

$$C(T) = C_0 e^{-\frac{E_k^a}{RT}} \quad ; \quad C = D, \tau_{\hat{C}_2}, \tau_{\hat{C}_6} \tag{10}$$

The resulting values of these properties are reported in Table 8 together with their experimental counterparts and those obtained through OPLS FF in Ref. [4]. From inspection of

FF	α (10^{-3} K^{-1})	E^a_{tr} (kJ/mol)	$E^a_{rot,\hat{C}_2}$ (kJ/mol)	$E^a_{rot,\hat{C}_6}$ (kJ/mol)
OPLS	1.56	16.0	8.1	11.3
QMD-FF	1.19	12.5	8.1	12.9
exp.	1.198	13.0	8.2	13.3

Table 8: Additional properties computed from the simulations at 280 K, 298 K and 320 K. The values with the OPLS FF were previously obtained in Ref. [4]. Experimental values are given the last raw. From left to right: thermal expansion coefficient,[91] activation energy for the translational diffusion Ref.[53] and activation energies for the spinning and tumbling motions.[69]

the latter Table, it is evident that the QMD-FF results nicely agree with the experiment, hence enforcing the reliability of the PICKY protocol.

5 Conclusions

In this work the automated Joyce and Picky procedures, previously developed in our group, are exploited to set up a multi-level protocol with the aim of assessing the ability of a chosen QM method to describe two-body IPESs. Several dispersion corrected DFT functionals, as well as other WFT based methods, were benchmarked for the benzene molecule, chosen as prototype of aromatic π-π interactions.

First, the quality of each tested functional was assessed with respect to reference CCSD(T)/CBS data, computed for selected geometrical arrangements. The benchmarked functionals showed remarkably different performances, especially in estimating the strength of the interaction, which was found to sensibly depend on the considered dimer arrangement. After Picky parameterization, these deviations reflected in different QMD-FF parameters and, therefore, in different properties, predicted by simulations.

The best results were obtained by employing the QMD-FF parameterized over the IPES sampled with the Grimme dispersion correction coupled with the CAM-B3LYP functional. Indeed, thermodynamic, structure and dynamic properties of the benzene liquid phase, computed through MD simulations performed with the CAM-B3LYP-D3 derived FF, were in good agreement with both experimental measures and the recently reviewed outcomes of several popular FFs. This success was traced back in the ability of the CAM-B3LYP-D3 functional to yields accurate estimates (with respect to the reference CCSD(T) data) of the dimer interaction energies for the most populated geometrical arrangement. Conversely, despite their ability in accurately mimic the CCSD(T)/CBS values for some conformations, other functionals were found to fail in giving a well balanced representation of the benzene IPES.

It is important noticing that the differences among CAM-B3LYP-D3, B3LYP-D3 and BLYP-D3 become apparent when a large portion of the IPES is considered, whereas benchmarks performed over few selected geometries could erroneously lead to the conclusion that they all provide a similar accuracy. The capability to extract a statistically relevant sample of conformers, given by the Picky procedure, is therefore crucial in the attempt to obtain classical FFs from QM information only. In fact, the remarkable improvement in the prediction of condensed phase properties found for all considered QMD-FFs, with

respect to the literature OPT-FF, is probably connected with the inadequate number of conformers employed in the parameterization of the OPT-FF.

On the contrary, the good results achieved here for benzene and, previously, for pyridine, seem to enforce the robustness of the PICKY protocol in sampling dimer IPES. Nevertheless, few points appear to call for further investigation and improvements. First, the PICKY parameterization protocol relies on solely two-body IPES. As mentioned throughout the text, neglecting the three-body terms leads to an overestimation of the interaction energies, which, in turn, causes an overestimation of the bulk densities. Second, the simplicity of the potential functions (standard LJ and Coulomb terms) and the lack of polarizable charges in the adopted QMD-FFs lead to an imperfect fit between the QM and MM IPESs, undermining the connection between the DFT functional and the final simulation outcomes. Finally, in its present formulation, the PICKY protocol is only applicable to pure liquids and crystals. However, work is in progress to extend the parameterization to mixtures, solutions and, eventually, inhomogeneous systems.

The fact that some of the macroscopic properties computed with empirical FFs are in better agreement with their experimental counterparts should not be surprising. As previously stated, most of the empirical parameters characterizing the latter FFs were tuned specifically to reproduce benzenes density and vaporization enthalpy, thus these quantities should not be considered when comparing the performances of empirical against QMD force-fields. Moreover, it should be pointed out that rather than seeking general and transferable parameters, the present protocol is aimed to set up specific force-fields, purposely tailored for the target molecule under study. For this reason QMD-FFs can be exploited in all those cases where standard FFs fail and/or an empirical parameterization is impossible due to the lack of experimental data. Examples of this applications are liquid crystalline phases, probe molecules in their excited states, inorganic-organic hybrid materials, condensed phases in non-standard thermodynamic conditions, unusual substituents, *etc.* . Some of these cases are currently under investigation in our group.

6 Aknowledgements

The research leading to these results has received funding from the Italian Ministry of University and Research (MIUR, PRIN 2010FM738P, and 2010PFLRJR).

7 Supporting Information Available

Details on JOYCE and PICKY parameterization routes, the list of all QMD-FF parameters, validation tests on CCSD(T) reference calculations and dispersion corrected DFT calculations, as well details and additional information on MD and MC simulations were reported in the Supporting Information. This information is available free of charge via the Internet at http://pubs.acs.org The JOYCE code and binaries are freely available at http://www.pi.iccom.cnr.it/joyce. The PICKY code and binaries will be soon freely available at http://www.pi.iccom.cnr.it/picky.

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] 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.

[5] Fu, C.-F.; Tian, S. X. A Comparative Study for Molecular Dynamics Simulations of Liquid Benzene *J. Chem. Theory Comput.* **2011**, *7*, 2240–2252.

[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] 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.

[8] 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.

[9] Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.

[10] Jorgensen, W. L.; Maxwell, D. S.; Tirado-rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids *J. Am. Chem. Soc.* **1996**, *7863*, 11225–11236.

[11] Sun, H. COMPASS: An ab Initio Force-Field Optimized for Condensed-Phase ApplicationsOverview with Details on Alkane and Benzene Compounds *J. Phys. Chem. B* **1998**, *102*, 7338–7364.

[12] MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–616.

[13] 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.

[14] Mackerell, A. D. Empirical force fields for biological macromolecules: Overview and issues *J. Comp. Chem.* **2004**, *25*, 1584–1604.

[15] Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. The GROMOS software for biomolecular simulation: GROMOS05. *J. Comp. Chem.* **2005**, *26*, 1719–51.

[16] Palmo, K.; Mannfors, B.; Mirkin, N.; Krimm, S. Potential Energy Functions: From Consistent Force Fields to Spectroscopically Determined Polarizable Force Fields *Biopolymers* **2003**, *68*, 383–394.

[17] Cacelli, I.; Prampolini, G. Parametrization and Validation of Intramolecular Force Fields Derived from DFT Calculations *J. Chem. Theory Comput.* **2007**, *3*, 1803–1817.

[18] 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.

[19] 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.

[20] Burger, S.; Lacasse, M.; Verstraelen, T.; Drewry, J.; Gunning, P.; Ayers, P. Automated Parametrization of AMBER Force Field Terms from Vibrational Analysis with a Focus on Functionalizing Dinuclear Zinc(II) Scaffolds *J. Chem. Theory Comput.* **2012**, *8*, 554–562.

[21] 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.

[22] 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.

[23] 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.

[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] 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.

[26] Grimme, S. A General Quantum Mechanically Derived Force Field (QMDFF) for Molecules and Condensed Phase Simulations *J. Chem. Theory Comput.* **2014**, *10*, 4497–4514.

[27] Amovilli, C.; Cacelli, I.; Cinacchi, G.; Gaetani, L.; Prampolini, G.; Tani, A. Structure and Dynamics of Mesogens Using Intermolecular Potentials Derived from Ab Initio Calculations *Theor. Chem. Accounts* **2007**, *117*, 885–901.

[28] Stone, A. J.; Misquitta, A. J. AtomAtom Potentials from Ab Initio Calculations *Int. Rev. Phys. Chem.* **2007**, *26*, 193–222.

[29] Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-Up Strategy. *J. Chem. Theory Comput.* **2007**, *3*, 1960–1986.

[30] Cacelli, I.; Lami, C. F.; Prampolini, G. Force-field Modeling through Quantum Mechanical Calculations: Molecular Dynamics Simulations of a Nematogenic Molecule in its Condensed Phases. *J. Comp. Chem.* **2009**, *30*, 366–378.

[31] McDaniel, J. G.; Schmidt, J. R. Physically-Motivated Force Fields from Symmetry-Adapted Perturbation Theory. *J. Phys. Chem. A* **2013**, *117*, 2053–66.

[32] Semrouni, D.; Cramer, C. J.; Gagliardi, L. AMOEBA Force Field Parameterization of the Azabenzenes *Theor. Chem. Accounts* **2014**, *134*, 1590.

[33] McDaniel, J. G.; Schmidt, J. R. First-Principles Many-Body Force Fields from the Gas Phase to Liquid: a "Universal" Approach. *J. Phys. Chem. B* **2014**, *118*, 8042–8053.

[34] 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

[35] Chattoraj, J.; Risthaus, T.; Rubner, O.; Heuer, A.; Grimme, S. A multi-scale approach to characterize pure CH4, CF4, and CH4/CF4 mixtures. *J. Chem. Phys.* **2015**, *142*, 164508.

[36] 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. Comp. Chem.* **2015**, *36*, 1015–27.

[37] Tiberio, G.; Muccioli, L.; Berardi, R.; Zannoni, C. Toward in Silico Liquid Crystals. Realistic Transition Temperatures and Physical Properties for n-Cyano-biphenyls via Molecular Dynamics Simulations *ChemPhysChem.* **2009**, *10*, 125.

[38] Grimme, S. Density functional theory with London dispersion corrections *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, 1, 211–228.

[39] Burns, L. a.; Vázquez-Mayagoitia, A.; Sumpter, B. G.; Sherrill, C. D. Density-Functional Approaches to Noncovalent Interactions: a Comparison of Dispersion Corrections (DFT-D), Exchange-Hole Dipole Moment (XDM) Theory, and Specialized Functionals. *J. Chem. Phys.* **2011**, *134*, 084107.

[40] Hohenstein, E. G.; Chill, S. T.; Sherrill, C. D. Assessment of the Performance of the M052X and M062X Exchange-Correlation Functionals for Noncovalent Interactions in Biomolecules *J. Chem. Theory Comput.* **2008**, *4*, 1996–2000.

[41] Vazquez-Mayagoitia, A.; Sherrill, C. D.; Apra, E.; Sumpter, B. G.; Apra, E. An Assessment of Density Functional Methods for Potential Energy Curves of Non-bonded Interactions: The XYG3 and B97-D Approximations *J. Chem. Theory Comput.* **2010**, *6*, 727–734.

[42] Riley, K. E.; Pitoák, M.; Cerný, J.; Hobza, P. On the structure and geometry of biomolecular binding motifs (hydrogen-bonding, stacking, X-Hπ): WFT and DFT calculations *J. Chem. Theory Comput.* **2010**, *6*, 66–80.

[43] Forni, A.; Pieraccini, S.; Rendine, S.; Sironi, M. Halogen Bonds with Benzene: an Assessment of DFT Functionals. *J. Comp. Chem.* **2014**, *35*, 386–94.

[44] Ehrlich, S.; Moellmann, J.; Grimme, S. Dispersion-Corrected Density Functional Theory for Aromatic Interactions in Complex Systems. *Acc. Chem. Res.* **2013**, *46*, 916–26.

[45] 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.

[46] Marianski, M.; Oliva, A.; Dannenberg, J. J. A Reinvestigation of the Dimer of para-Benzoquinone and Pyrimidine with MP2, CCSD(T), and DFT Using Functionals Including Those Designed to Describe Dispersion. *J. Phys. Chem. A* **2012**, *116*, 8100–5.

[47] Antony, J.; Sure, R.; Grimme, S. Using Dispersion-Corrected Density Functional Theory to Understand Supramolecular Binding Thermodynamics. *Chem. Comm.* **2014**, *51*, 1764–1774.

[48] Paytakov, G.; Dinadayalane, T.; Leszczynski, J. Toward Selection of Efficient Density Functionals for van der Waals Molecular Complexes: Comparative Study of C-Hπ and N-Hπ Interactions. *J. Phys. Chem. A* **2015**, *119*, 1190–1200

[49] Kennedy, M. R.; McDonald, A. R.; DePrince III,A. E.; Marshall, M. S.; Podeszwa, R.; Sherrill, C. D. Resolving the three-body contribution to the lattice energy of crystalline benzene: Benchmark results from coupled-cluster theory *J. Chem. Phys.* **2014**, *140*, 121104

[50] *Handbook of Chemistry and Physicis*; CRC press: Boca Raton, 1997.

[51] Timmermans, J. *Physical Chemical Constants of Pure Organic Compounds*; Elsevier, 1950.

[52] Majer, V.; Svoboda, V. *Enthalpies of Vaporization of Organic Compounds: A Critical Review and Data Compilation*; Blackwell Scientific Publications: Oxford, 1985.

[53] Falcone, D. R.; Douglass, D. C.; McCall, D. W. Self-diffusion in benzene *J. Phys. Chem.* **1967**, *71*, 2754–2755.

[54] Tanabe, K. Determination of Rotational Diffusion Constants of Liquid Benzene from Measurements of Infrared and Raman Line Widths *Chem. Phys. Lett.* **1979**, *63*, 43–46.

[55] Battaglia, M.; Buckingham, A.; Williams, J. The electric quadrupole moments of benzene and hexafluorobenzene *Chem. Phys. Lett.* **1981**, *78*, 421–423.

[56] Cabaço, M. I.; Danten, Y.; Besnard, M.; Guissani, Y.; Guillot, B. Neutron Diffraction and Molecular Dynamics Study of Liquid Benzene and Its Fluorinated Derivatives as a Function of Temperature *J. Phys. Chem. B* **1997**, *101*, 6977.

[57] Chelli, R.; Cardini, G.; Procacci, P.; Righini, R.; Califano, S.; Albrecht, A. Simulated Structure, Dynamics, and Vibrational Spectra of Liquid Benzene *J. Chem. Phys.* **2000**, *113*, 6851–6863.

[58] Bhide, S. Y.; Kumar, A. V. A.; Yashonath, S. Diffusion of Hydrocarbons in Confined Media: Translational and Rotational Motion *J. Chem. Sci.* **2001**, *113*, 559–577.

[59] Headen, T. F.; Howard, C. A.; Skipper, N. T.; Wilkinson, M. A.; Bowron, D. T.; Soper, A. K. Structure of π-π Interactions in Aromatic Liquids. *J. Am. Chem. Soc.* **2010**, *132*, 5735–42.

[60] Karlstrom, G.; Linse, P.; Wallqvist, A.; Jonssonf, B. Intermolecular Potentials for the H20-C6H6 and the C6H6-C6H6 Systems Calculated in an ab Initio SCF CI Approximation *J. Am. Chem. Soc.* **1983**, *6*, 3777–3782.

[61] Linse, P.; Engström, S.; Jönsson, B. Molecular dynamics simulation of liquid and solid benzene *Chem. Phys. Lett.* **1985**, *115*, 95–100.

[62] Bartell, L. S.; Sharkey, L. R.; Shi, X. Electron Diffraction and Monte Carlo Studies of Liquids. 3. Supercooled Benzene *J. Am. Chem. Soc.* **1988**, *110*, 7006–7013.

[63] Jorgensen, W. L.; Severance, D. L. Aromatic-Aromatic Interactions: Free Energy Profiles for the Benzene Dimer in Water, Chloroform, and Liquid Benzene *J. Am. Chem. Soc.* **1990**, *112*, 4768–4774.

[64] Grolier, J.-P.; Roux-Desgranges, G.; Berkane, M.; Jiménez, E.; Wilhelm, E. Heat capacities and densities of mixtures of very polar substances 2. Mixtures containing N, N-dimethylformamide *J. Chem. Thermodynam.* **1993**, *25*, 41–50.

[65] Hobza, P.; Selzle, H. L.; Schlag, E. W. Potential Energy Surface for the Benzene Dimer. Results of ab Initio CCSD(T) Calculations Show Two Nearly Isoenergetic Structures: T-Shaped and Parallel-Displaced *J. Chem. Phys.* **1996**, *100*, 18790–18794.

[66] Nakagawa, T.; Umemura, J.; Hayashi, S.; Oobatake, M.; Miwa, Y.; Machida, K. Molecular Dynamics Study of the Spectroscopic Properties of Liquid Benzene *Mol. Phys.* **1996**, *88*, 1635–1643.

[67] Kim, J. H.; Lee, S. H. Molecular Dynamics Simulation Studies of Benzene, Toluene, and p-Xylene in a Canonical Ensemble *Bull. Korean Chem. Soc.* **2002**, *23*, 5–10.

[68] Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. Origin of Attraction and Directionality of the π/π Interaction: Model Chemistry Calculations of Benzene Dimer Interaction *J. Am. Chem. Soc.* **2002**, *124*, 104–112.

[69] Witt, R.; Sturz, L.; Dölle, A.; Müller-Plathe, F. Molecular Dynamics of Benzene in Neat Liquid and a Solution Containing Polystyrene. 13 C Nuclear Magnetic Relaxation and Molecular Dynamics Simulation Results *J. Phys. Chem. A* **2000**, *104*, 5716–5725.

[70] Milano, G.; Müller-Plathe, F. Cyclohexane-Benzene Mixtures: Thermodynamics and Structure from Atomistic Simulations *J. Phys. Chem. B* **2004**, *108*, 7415–7423.

[71] Grimme, S. Improved second-order MøllerPlesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies *J. Chem. Phys.* **2003**, *118*, 9095.

[72] Sinnokrot, M. O.; Sherrill, C. D. Highly Accurate Coupled Cluster Potential Energy Curves for the Benzene Dimer: Sandwich, T-Shaped, and Parallel-Displaced Configurations *J. Phys. Chem. A* **2004**, *108*, 10200–10207.

[73] Podeszwa, R.; Bukowski, R.; Szalewicz, K. Potential Energy Surface for the Benzene Dimer and Perturbational Analysis of π-π Interactions *J. Phys. Chem. A* **2006**, *110*, 10345–10354.

[74] Janowski, T.; Pulay, P. High accuracy benchmark calculations on the benzene dimer potential energy surface *Chem. Phys. Lett.* **2007**, *447*, 27–32.

[75] Lee, E. C.; Kim, D.; Jurečka, P.; Tarakeshwar, P.; Hobza, P.; Kim, K. S. Understanding of Assembly Phenomena by Aromatic-Aromatic Interactions: Benzene Dimer and the Substituted Systems *J. Phys. Chem. A* **2007**, *111*, 3446–3457.

[76] Pavone, M.; Rega, N.; Barone, V. Implementation and Validation of DFT-D for Molecular Vibrations and Dynamics: The Benzene Dimer as a Case Study *Chem. Phys. Lett.* **2008**, *452*, 333–339.

[77] Bludský, O.; Rubeš, M.; Soldán, P.; Nachtigall, P. Investigation of the Benzene-Dimer Potential Energy Surface: DFT/CCSD(T) Correction Scheme *J. Phys. Chem.* **2008**, *128*, 1–8.

[78] David Sherrill, C.; Takatani, T.; Hohenstein, E. G. An Assessment of Theoretical Methods for Nonbonded Interactions: Comparison to Complete Basis Set Limit Coupled-Cluster Potential Energy Curves for the Benzene Dimer, the Methane Dimer, Benzene-Methane, and Benzene-H_2S *J. Phys. Chem. A* **2009**, *113*, 10146–10159.

[79] 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 D.01; Gaussian, Inc.; Wallingford CT; 2009.

[80] Barone, V.; Cacelli, I.; Crescenzi, O.; D'Ischia, M.; Ferretti, A.; Prampolini, G.; Villani, G. Unraveling the interplay of different contributions to the stability of the quinhydrone dimer *RSC Adv.* **2014**, *4*, 876.

[81] Prampolini, G.; Cacelli, I.; Ferretti, A. Intermolecular Interactions in Eumelanins: a Computational Bottom-Up Approach. I: Small Building Blocks. *RSC Adv.* **2015**, *5*, 38513-38526.

[82] Riley, K. E.; Platts, J. A.; Rezáč, J.; Hobza, P.; Hill, J. G. Assessment of the Performance of MP2 and MP2 Variants for the Treatment of Noncovalent Interactions *J. Phys. Chem. A* **2012**, *116*, 4159

[83] Boys, S.; Bernardi, F. The calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors *Mol. Phys.* **1970**, *19*, 553–566.

[84] 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.

[85] Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Function *Theor. Chem. Accounts* **2008**, *120*, 215–241.

[86] Sherrill, C. D.; Sumpter, B. G.; Sinnokrot, M. O.; Marshall, M. S.; Hohenstein, E. G.; Walker, R. C.; Gould, I. A. N. R. Assessment of Standard Force Field Models Against High-Quality Ab Initio Potential Curves for Prototypes of π-π, CH/π, and SH/π Interactions. *J. Comp. Chem.* **2009**, *30*, 2187–93.

[87] Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269

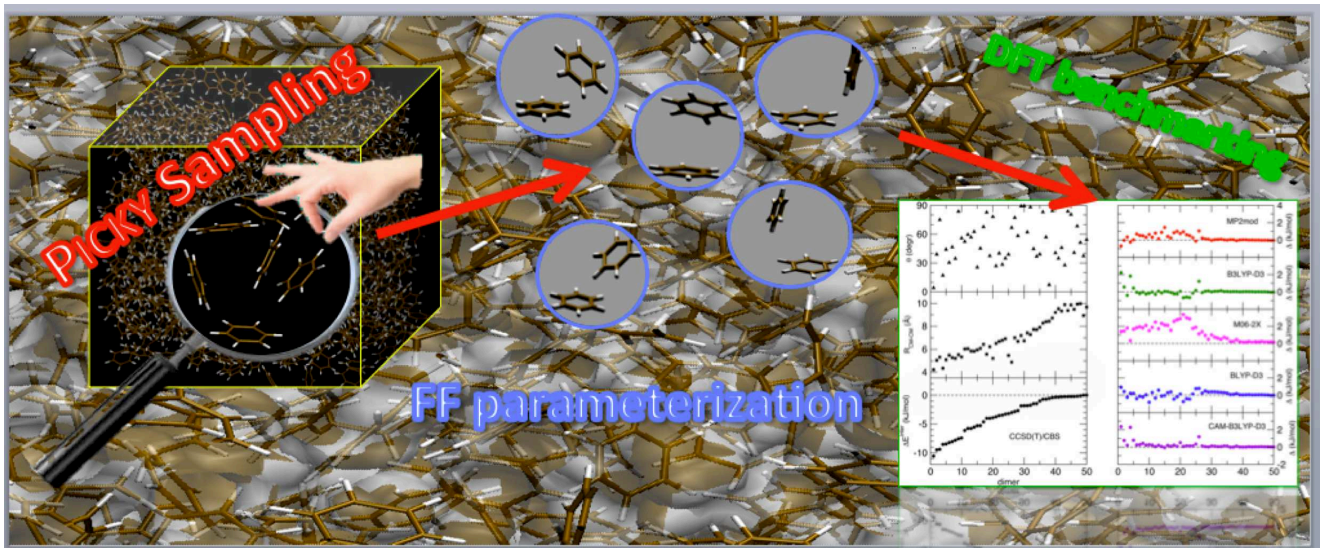
[88] Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999

[89] McCool, M. A.; Collings, A. F.; Woolf, L. A. Pressure and Temperature Dependence of the Self-Diffusion of Benzene. *J. Chem. Soc. Faraday Trans.* **1972**, *68*, 1498.

[90] Dymond, J. H.; Robertson, J.; Isdale, J. Transport Properties of Nonelectrolyte Liquid Mixtures - IV. Viscosity Coefficients for Benzene, Perdeuterobenzene, Hex-

afluorobenzene, and an Equimolar Mixture of Benzene + Hexafluorobenzene from 25 to 100^o Pressures up to the Freezing Pressure. *Int. J. Thermophys.* **1981**, *2*, 223.

[91] Klüner, R. *Rotatorische Dynamik von Kohlenwasserstoffen in flüssiger Phase*; Shaker-Verlag: Aachen, 1995.



TOC graphic:

Accuracy of Quantum Mechanically Derived Force-Fields Parameterized from Dispersion-corrected DFT data:
the Benzene Dimer as Prototype for Aromatic Interactions