An Automated Approach for the Parameterization of Accurate Intermolecular Force-Fields: Pyridine as a Case Study

Ivo Cacelli,[a] Antonella Cimoli,[a] Paolo Roberto Livotto,[b] Giacomo Prampolini*[c,d]

An automated protocol is proposed and validated, which integrates accurate quantum mechanical calculations with classical numerical simulations. Intermolecular force fields, (FF) suitable for molecular dynamics (MD) and Monte Carlo simulations, are parameterized through a novel iterative approach, fully based on quantum mechanical data, which has been automated and coded into the PICKY software, here presented. The whole procedure is tested and validated for pyridine, whose bulk phase, described through MD simulations performed with the specifically parameterized FF, is characterized by computing several of its thermodynamic, structural, and transport properties, comparing them with their experimental counterparts. © 2012 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.22937

Introduction

Thanks to the massive increase of computational resources in the past two decades, computer simulation techniques,[1,2] as molecular dynamics (MD) or Monte Carlo, have nowadays become powerful tools in many different fields of supramolecular chemistry,[3] as the study of soft matter structural and dynamic properties[4–7] or the investigation of biomolecular functions in solution.[8–10] Indeed, it is now possible to perform MD atomistic simulations on bulk phases up to thousands of medium sized molecules for several tenths of nanoseconds,[6,7,11–13] whereas large biomolecules can be studied in solution up to the microsecond regime.[10,14] Because of the high computational cost, an adequate exploration of the phase space of such large systems can only be achieved with simulation methods based on classical dynamics. However, an accurate statistical treatment turns useless if not complemented by "realistic" force fields (FF), that is, capable of retaining most of the structural details which specify the chemical identity of the simulated systems. The latter is a crucial point which may become critical when a delicate balance of non covalent forces determines the specificity of the intermolecular interactions and, hence, the final supramolecular structure (e.g., host–guest chemistry, molecular recognition, liquid–crystalline transitions, etc.).

The majority of numerical simulations are performed with empirical potentials (e.g., OPLS,[15,16] GROMOS,[17] AMBER,[18,19] and CHARMM[20–22]), which are parameterized toward experimental and quantum chemical values for a selected set of similar molecules. For instance, OPLS FF parameters are tuned to reproduce experimental density and vaporization enthalpy of target ensembles made of similar molecules of relatively small dimensions. In the case of larger molecules, for example, polymers, liquid crystals or biomacromolecules, a corresponding tuning of the FF can be very time consuming both for the long equilibration times required by their dynamics and for the large number of parameters needed for their description, making this route rather hard to be pursued. Furthermore, the quality of the description is highly correlated to the simulation time used in the parameterization, and extentions of the total length of the MD runs may reveal the appropriateness of the FF.[23] For the description of larger molecules, parameters are therefore transferred to the same atoms for which they where devised, under the condition that they are embedded in a similar chemical framework as that found in their smaller homologues. However, in some of these cases, the straightforward adoption of the literature FF's, based on the transferability hypothesis, may result in very approximate results, as for instance reported by Zannoni and coworkers,[13] in an MD study on the transition temperatures of a liquid crystalline homologue series.

Broadly speaking, an *a priori* validation of FF parameters, in particular those regarding intermolecular interactions, is a rather difficult task, also due to the scarceness of reliable benchmark data. Among these, the S22 and JCSH-2005 databases[24] contain several interaction energies, for different stacked, H-bonded and mixed small to medium-sized molecular dimers, computed at high computational level [Coupled Cluster with Singles and Doubles excitations and perturbative Triples, CCSD(T), at the complete basis set (CBS) limit]. Both databases have been

[a] *I. Cacelli, A. Cimoli*
Dipartimento di Chimica e Chimica Industriale, Università degli Studi di Pisa, via Risorgimento 35, I-56126 Pisa, Italy

[b] *P. R. Livotto*
Instituto do Quìmica, Universidade Federal do Rio Grande do Sul, Avenida Bento Gonalves 9500, CEP 91501-970 Porto Alegre, Brazil

[c] *G. Prampolini*
Instituto dei Processi Chimico-Fisici (IPCF), Consiglio Nazionale delle Ricerche (CNR), Area della Ricerca, via Moruzzi 1, 56124 Pisa, Italy
Scuola Normale Superiore, piazza dei Cavalieri 7, I-56126 Pisa, Italy
E-mail: giacomo.prampolini@pi.ipcf.cnr.it

recently used to validate the performance of widely used FF,[25] resulting in a relatively good performance for stacked and mixed complexes, but not for H-bonded complexes. In 2010, Hobza and coworkers extended[26] such study by considering also dimer geometries different from the minimum energy conformation, finding a less close agreement between Quantum Mechanical (QM) and FF values. Despite a surprisingly good overall standard deviation, large deviations were reported for single cases, as for instance an overestimation of the stacking stabilization energy of aromatic systems of about 25%, which can become even worst for large distances. It is worth noticing that notwithstanding that the latter dimer energies are smaller than those at equilibrium geometry, their number grows proportionally to the square of the intermolecular distance, so that their sum contributes non-negligibly to the total interaction energy. These findings were confirmed by Ref. [27], where a comparison between Lennard-Jones (L J) and electrostatic (charge–charge) FF energy terms with the contributions arising from a SAPT-DFT[28,29] approach, showed that in even those cases where a good agreement with the QM data is achieved, this comes from a large compensation of errors arising from the different contributions, being the error on the separate contributions approximately one order of magnitude larger than the one reported for the total energy. In conclusion, it appears that in many cases the straightforward adoption of literature transferable FF's may not lead to accurate predictions, and alternative routes to model potential construction should be devised to complement existing databases.

A possibility is to adopt a parameterization route based on QM calculations. As the intramolecular part of the FF is regarded, QM-based parameterizations have been proposed by several groups and used in widely used FF.[19,30–34] Among these, the procedure recently proposed by our group and implemented in the JOYCE software[34] derives equilibrium internal coordinates and force constants of the target molecule by fitting QM optimized energies, gradients and Hessian matrix. Conversely, the QM route to intermolecular FF is much more involved, because it requires in principle the creation of large energy versus geometry QM databases. In a previous work on benzene,[35] we have shown that an accurate, QM derived, two-body potential is capable to account for several bulk properties, both in crystalline and liquid phases, obtaining an overall performance comparable to that of the well-known OPLS effective potential,[36] purposely tuned on some benzene's experimental properties. Thereafter, the QM route to FF parameterization, combined with the fragmentation reconstruction method,[37–39] was extended to larger molecules[12,37–45] in the field of liquid crystals, and was shown to successfully reproduce experimental transition temperatures, rotational viscosity coefficients, shear viscosity and orientational ordering. The major problem found in some of these parameterizations is the reproduction of the bulk density, which, in the worst cases,[38] was overestimated by almost $\sim$6%,[38] causing dramatic effects on the translational dynamics.[44]

The protocol presented here is an improved version of the first intermolecular parameterization route proposed by our group.[37,39] In its original version the QM-based parameterization was accomplished in three steps: (i) a large set of dimer geometries was created by translating and rotating one monomer with respect to the other; (ii) the intermolecular energy was QM-computed for all generated dimer geometries; (iii) a model energy function was fitted against the database of QM energies and the intermolecular FF parameters derived. One of the main limitations of such protocol can be found in the sampling procedure adopted to select significant dimer geometries. These arrangements were all chosen *a priori*, following the "chemical intuition" of the operator and therefore their representativeness was not granted. In other words, the overall good reproduction of the QM energies of the selected dimers was by no means ensured for many different configurations, which could be populated in the bulk phase at room temperature but were not included in the *a priori* selected database. A second limit of the previous protocol is that the internal geometry of all monomers was always kept frozen. At short intermolecular distances, this constraint could undermine the representativity of the sampling, as important molecular distortions may take place to avoid short interatomic contact distances. This biased sampling may in turn reflect in significant inaccuracies on the simulated bulk density.

In this article, we present a new strategy to overcome these lacks. Resorting to the suggestions of Akin-Ojo et al.,[46] several dimer configurations will now be sampled directly from MD trajectories according to a well-established protocol, their interaction energies *ab initio* computed and used to obtain the model potential energy functions of the FF. A trial set of FF parameters is chosen to start up the first simulation, and the procedure iteratively repeated until some convergence criteria are satisfied.

As case study, we have chosen the pyridine bulk phase for the following reasons. First of all, $\pi-\pi$ interactions are of fundamental importance in bio-systems, as they appear in protein–ligand interactions,[47] protein folding,[48] and DNA structure.[49] Despite benzene has often been used as model system to study aromatic $\pi-\pi$ complexes, nitrogen hetero-atoms are present in nucleic acid bases, influencing the stacking energies between bases, and hence the structure of DNA. More in general, many biologically important molecules as amino acids often contain aromatic heterocycles, and their interactions have been shown[50–53] to differ consistently from those found in the benzene dimer. Secondarily, pyridine has been recently[54] tested as a benchmark for the implementation of a new FF, the OPLS-CS, whose results were shown to improve the standard OPLS[55] description. Unfortunately, due to the more complex functional forms, the OPLS-CS FF resulted in a three times increased computational costs. In this perspective, it may be interesting to verify if similar improvement with respect to the standard OPLS description could be gained by the present protocol, without the adoption of adjunctive model functions. Finally, very recent high level QM calculations on pyridine trimers[51,53] have revealed that three-body effects contribute negligibly, enforcing the reliability of our protocol, which is based on the parameterization of QM computed two-body interactions.

The article is organized as follows: in the next section a description of the presented novel protocol is given, while computational details of the other used techniques are given in

"Computational Detail" section; results are discussed in "Results" section and main conclusions are drawn in the last section.

Picky: An Intermolecular Parameterization Protocol

The main steps of the present multilevel approach, aimed to evaluate an accurate intermolecular FF, are illustrated in Figure 1.

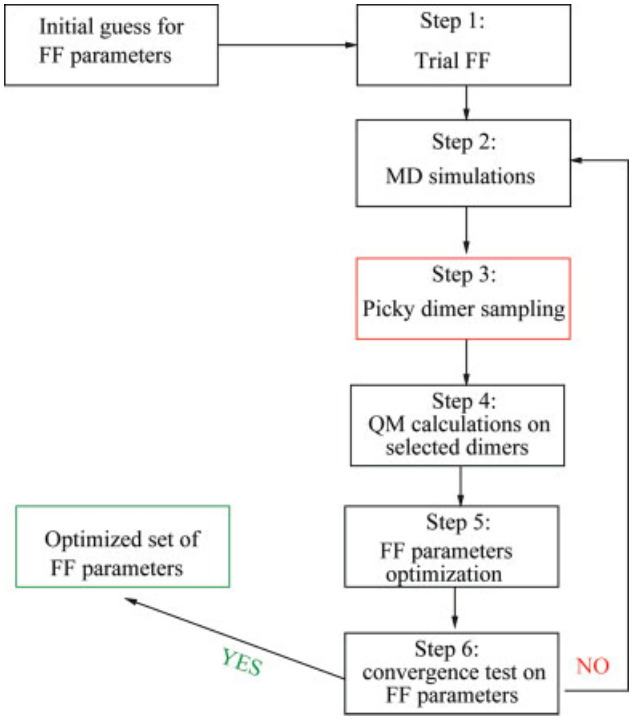

Figure 1. FF parameterization integrated approach based on QM data and MD simulations. Main steps of the method are shown. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

The procedure goes on iteratively until a convergence threshold on the value of the FF parameters is reached. As shown in Figure 1, a single iterative cycle can be summarized into the following six steps.

1. Set up a trial FF from a chosen set of parameters (initial guess).
2. Run an MD simulation.
3. Sample selected molecular dimers from the resulting MD trajectory.
4. Perform QM calculations of the intermolecular energy for the selected dimers and add these to the previous QM database.
5. Obtain, through a least square fitting procedure, the best set of FF parameters, with respect of the computed QM energy data.
6. Test the convergence of the resulting parameters relative to the ones obtained in the previous cycle. If convergence is reached, perform final MD simulations, otherwise perform a further cycle (from step 2.).

In the following, each step will be separately considered and described.

Step 1: Choice of model potential functions

A model potential is chosen for the target molecule.

In this application, a full atomistic model, shown in Figure 2, has been chosen for pyridine. The intermolecular part of the adopted FF is expressed by:

$$E^{\text{inter}} = \sum_{i=1}^{N_{\text{sites}}} \sum_{j=1}^{N_{\text{sites}}} \left[E_{ij}^{\text{LJ}} + E_{ij}^{\text{Coul}} \right] \tag{1}$$

where E_{ij}^{LJ} and E_{ij}^{Coul} are the interactions (12–6 LJ potential and charge-charge, respectively) between a pair of sites i, j of two different molecules, with

$$E_{ij}^{\text{LJ}} = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]; \quad E_{ij}^{\text{Coul}} = \frac{q_i q_j}{r_{ij}} \tag{2}$$

LJ parameters ϵ_{ij} and σ_{ij} are obtained from site parameters through the Lorentz–Berthelot mixing rules.[1]

It should be pointed out that the proposed parameterization approach is completely independent from the chosen model, and it can be applied to intermolecular functions other then LJ, as for instance Buckingham or Morse potentials.

Step 2: MD simulations

Numerical simulations are performed with the current FF in several conditions at constant temperature and pressure. When the system is equilibrated some dimer configurations are extracted from selected MD snapshots, with criteria that will be discussed below.

Step 3: Sampling algorithm

We have coded a Fortran program, named PICKY, which has the task of selecting from a snapshot of a MD run a given number of dimers to be considered for a QM calculation, with the aim of exploring as far as possible the full configurational dimer space. Thus, the selection should be able to discard the dimers "similar" to any of those already included in the QM database and to accept those dimers which are rather different from the stored ones. This task has to take into account also the energy of the dimer under scrutiny, to avoid those dimers having high energy and representing a portion of the dimer potential energy surface (PES) with a small probability of being populated at room temperature. In this way, the representativity of the QM database can be gradually improved and the FF parameters are expected to be more and more adequate to the whole PES, within the limit of the functional form reported in Eq. (1). The comparison between the inquired dimer and those present in the QM database is based upon the creation of an index which measures the difference between two dimers.

As a simple example, a pyridine dimer can be considered (Fig. 2). Four atoms are chosen for each monomer to define a long (N—C_3) and a short (C_1—C_1) molecular axis, whose versors ($\hat{L}$ and $\hat{S}$, respectively) uniquely define the orientation of the molecule. The relative position, irrespective of the distance, of the A and B monomers forming a dimer, can be specified with the aid of the

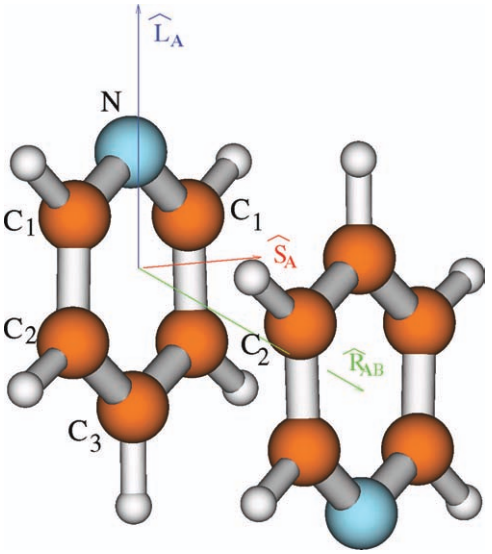

Figure 2. A pyridine dimer: a generic configuration is shown. Site labels are displayed for monomer A. hydrogens are labeled accordingly to the carbon site they are bonded to (i.e., H_1 is an hydrogen bonded to a C_1 site). The long ($\hat{L}_A$) and short ($\hat{S}_A$) molecular axis are respectively shown in blue and red. The intermolecular barycenter axis ($\hat{R}_{AB}$) is also displayed in green. The same labels and axis ($\hat{L}_B$, $\hat{S}_B$) are used for monomer B.

intermolecular versor $\hat{R}_{AB}$ joining the two barycenters. Besides the intermolecular A–B distance, a possible quantity that can be used to quantify the difference between dimers $(AB)_1$ and $(AB)_2$ is

$$O = \frac{1}{8}[|(\hat{L}_A \cdot \hat{L}_B)_1 - (\hat{L}_A \cdot \hat{L}_B)_2| + |(\hat{L}_A \cdot \hat{R}_{AB})_1 - (\hat{L}_A \cdot \hat{R}_{AB})_2| + |(\hat{L}_B \cdot \hat{R}_{AB})_1 - (\hat{L}_B \cdot \hat{R}_{AB})_2| + |(\hat{S}_A \cdot \hat{S}_B)_1 - (\hat{S}_A \cdot \hat{S}_B)_2|] \quad (3)$$

where the normalizing factor 1/8 assures that O spans the [0–1] range. Furthermore, as the difference in the A–B distance between two dimers is concerned, it can be conveniently taken into account by comparing the trace of the inertia tensors T_1 and T_2 which are invariant under rotation

$$T = \frac{|T_1 - T_2|}{T_1 + T_2} \quad (4)$$

Finally, the internal conformations are taken into account by the term

$$F = f\left(\frac{1}{N_A}\sum_i^{N_A}\left(\lambda_i^{A_1} - \lambda_i^{A_2}\right)^2 + \frac{1}{N_B}\sum_i^{N_B}\left(\lambda_i^{B_1} - \lambda_i^{B_2}\right)^2\right) \quad (5)$$

The summation indexes λ run over the molecular internal coordinates of monomers A_k and B_k, where $k = 1, 2$ refers to the dimer [i.e., $(AB)_1$ or $(AB)_2$] formed by the A and B units. f is a constant that scales the F term, to make it comparable with the O and T terms, which are bounded to assume values lower than 1. All these terms concur in the formation of a difference index (DI) which should measure the difference between the dimers $(AB)_1$ and $(AB)_2$ in the high dimensional coordinate space

$$DI_{1,2} = O + T + F \quad (6)$$

Such an algorithm can be used to select the most different dimers to achieve the largest representativity of the multi dimensional PES of the dimers. In practice, once a snapshot is extracted from the MD trajectory, all resulting pairs of molecules are considered, and arranged according to their minimum DI with respect to all dimers already included in the current QM database, set up in previous iterations. To improve the flexibility of the sorting, some constraints may be used to select the dimers from the snapshot. Those used in the present article are the energy (computed with the current FF) and the distance R_{AB}. For instance those dimers having energy higher than a given threshold or distance shorter than a fixed value, may be discarded. A given number of dimers, those showing the maximum DI and satisfying the constraints, is then considered for QM calculation and added to the QM database. From this extended QM database a new set of FF parameters is determined and used for a further MD run in an iterative procedure as above explained.

Step 4: QM calculations on dimers selected by Picky

Obviously, interaction energies of the selected dimer configurations have to be *ab initio* computed at a reliable level of theory. High-level QM accurate calculations should be used, but the number of selected dimer grows rapidly during the iterative

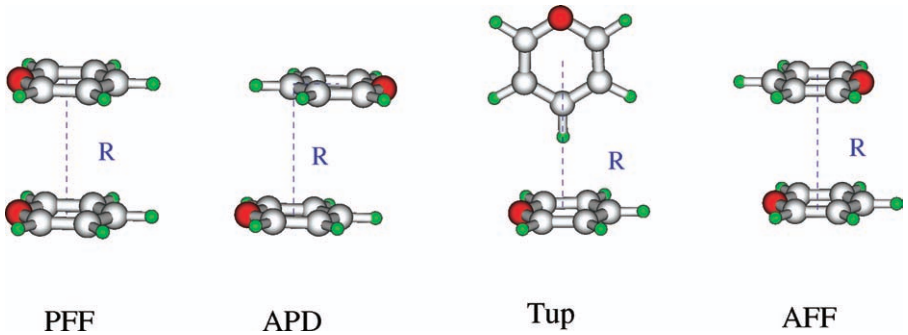

Figure 3. Pyridine dimer geometries taken into account in reference CCSD(T)/CBS calculations. From left to right: PFF, APD, Tup, and AFF. Last arrangement has not been used in the fitting of the polarization exponent.

procedure, and a good compromise between computational cost and accuracy is requisite. However, the chosen lower level method should be first validated against the results obtained, on few selected conformations, with a reliable high-level method.

In this application, this can be a particularly hard task, as most of the attractive interactions in pyridine dimers derive from dispersion forces.[51–53] As high-level reference method, we have chosen the CCSD(T)/CBS protocol, which was separately proposed in slightly different implementations by the groups of Tsuzuki,[56] Hobza,[57,58] and Sherrill,[59,60] and often referred to as the "gold standard" of quantum chemistry, in particular for aromatic interactions. Turning to the lower level method, to be used in the extensive sampling, one could resort to either wavefunction-based methods or Density funtcional theory (DFT)-based techniques. Among the formers, computational convenience would suggest the use of Møller–Plesset second order perturbation (MP2) theory, but the remarkable over-binding reported for aromatic interactions,[59,61–64] rules out the choice of this method with standard basis sets. As a matter of fact, the standard MP2 method could still be adopted if coupled with a small basis set, where the polarization exponent is modified to a different value. Indeed, the 6-31G*(0.25) and cc-pVDZ(0.25) basis sets (which consist in the standard 6-31G* and cc-pVDZ basis sets with the polarization exponent on Carbon modified from 0.80 to 0.25) were first suggested by Hobza and coworkers,[65] and recently showed to well perform[35,58,66,67] with respect to more expensive computational methods. Conversely, within the DFT framework, Truhlar and co-workers[68–70] recently succeeded in a reparameterization aimed to take dispersion into account, achieving a reasonable agreement[70,71] with reference data for benchmark systems, but very recent results[58] evidenced some defects in the overall representation of the intermolecular energy curves. In the same article, somewhat better performance was reported for the dispersion corrected DFT-D approach[59,60,72,73].

It should be pointed out that the choice of the high and low-level QM methods is by no means unique, and different techniques can be coupled with the PICKY procedure. In this work, the MP2/cc-pVDZ(mod) will be used for the systematic computation of the dimer interaction energy, while work is in progress [Cacelli et al. (work in progress)] to test the performances of the proposed multi level parameterization approach coupled with DFT-D computations for a different molecule.

Step 5: FF parameters optimization

The intermolecular parameters are obtained from a least square fitting procedure, by minimizing the difference between the QM-computed dimer energy (U) and the parameterized one (E^{inter}) for all sampled geometries through the functional

$$I^{\text{inter}} = \frac{\sum_{k=1}^{N\text{geom}} \left[\left(U_k - E_k^{\text{inter}}\right)^2\right] e^{-\alpha U_k}}{\sum_{k=1}^{N\text{geom}} e^{-\alpha U_k}} \tag{7}$$

where N_{geom} is the number of geometries considered for the target dimer, α a Boltzmann-like weight, U_k is the QM energy of the kth dimer arrangement and E_k^{inter} the value of the fitting model function computed, according to Eq. (1). The minimization of functional (7) is performed by a slightly modified version of the STEPIT[74] routine, encoded in the PICKY software. All the PICKY package is freely available upon request to the authors.

Computational Details

Intramolecular calculations

All QM single molecule calculations, used in the intramolecular PES parameterization, were performed with the GAUSSIAN 03 package,[75] with the well-tested density functional B3LYP method[76] with a standard correlation consistent basis set cc-pVDZ. The absolute energy minimum was obtained by a complete geometry optimization. Vibrational frequencies, gradients and Hessian matrix were computed from this optimized conformation.

The intramolecular part of the pyridine FF was described as

$$E^{\text{intra}} = E^{\text{stretch}} + E^{\text{bend}} + E^{\text{Rtors}} \tag{8}$$

All terms have an harmonic expression

$$E^{\text{stretch}} = \frac{1}{2}\sum_{\mu}^{N_s} k_\mu^s \left(r_\mu - r_\mu^0\right)^2;\; E^{\text{bend}} = \frac{1}{2}\sum_{\mu}^{N_b} k_\mu^b \left(\theta_\mu - \theta_\mu^0\right)^2;$$
$$E^{\text{Rtors}} = \frac{1}{2}\sum_{\mu}^{N_{Rt}} k_\mu^t \left(\phi_\mu - \phi_\mu^0\right)^2 \tag{9}$$

where k_μ^s, k_μ^b, k_μ^t and r_μ^0, θ_μ^0, ϕ_μ^0 are the force constants and equilibrium values for stretching, bending and rigid torsional internal coordinates, respectively. All intramolecular parameters were taken from Ref. [34], where they were derived from the single molecule QM data through the JOYCE program. Details of the fitting procedure can be found in Ref. [34].

Intermolecular calculations

As stated in the previous section, QM calculations on the sampled dimers are here performed at the MP2/cc-pVDZ(mod) level. In particular, the d polarization exponent on C atoms is 0.25, as suggested by Hobza and coworkers.[58] However, no indications were found in literature about the exponents to be used for the nitrogen atom. To obtain the most appropriate values of the p and d exponents for the pyridine H and N atoms, respectively, high level accurate calculations are first performed for a limited number of dimer arrangements. Reference data are thus obtained at the CCSD(T) level of theory, estimated at the CBS limit, which is in turn obtained from the MP2/CBS value ($\Delta E_{\text{CBS}}^{\text{MP2}}$) by adding the ΔCCSD(T) correction term

$$\Delta E_{\text{CBS}}^{\text{CCSD(T)}} = \Delta E_{\text{CBS}}^{\text{MP2}} + \Delta\text{CCSD(T)} \tag{10}$$

where the correction term ΔCCSD(T) is obtained from the difference

$$\Delta\text{CCSD(T)} = \left[\Delta E^{\text{CCSD(T)}} - \Delta E^{\text{MP2}}\right]_{\text{small basis set}} \tag{11}$$

Despite this procedure has been extensively used by several groups,[52, 56–58, 77] up to our knowledge, the best route to a correct estimate of both ΔE^{MP2}_{CBS} and ΔCCSD(T) values has not yet been assessed. In this work, computational convenience has prompted us to follow the suggestions of Ref. [57] and the MP2 value in the CBS limit is estimated through the procedure proposed by Truhlar[78] (using the cc-pVDZ and cc-pVTZ basis sets), whereas the ΔCCSD(T) is computed with the cc-pVDZ(0.25,0.15) modified basis set. The results obtained in this way for few selected dimer geometries are then used to obtain the best polarization exponents, for H and N atoms, in the cc-pVDZ(mod) basis set, to be used in the MP2 extensive calculations. This has been done by a fully automated procedure, coded in an "in house" software named Exopt, and consists in minimizing the energy differences between CCSD(T)/CBS reference energies and those obtained with the MP2/cc-pVDZ(α_C, α_N, and α_H) method, being α_C, α_N, and α_H, the polarization function exponents of carbon (set to 0.25), nitrogen, hydrogen atoms, respectively. QM reference and low level extensive calculations of dimer energies were all performed with the Gaussian 03 package.[75] Finally, the intermolecular FF's reported in Eq. (1) was parameterized through the Picky code, by minimizing the functional (7), imposing for all geometries a Boltzmann-like weight with $\alpha = 0.05$ mol/kJ.

MD simulations

The QM-parameterized FF has been used in MD runs perform with a parallel version of the Moscito 3.9[79] package, on systems of 512 pyridine molecule. In all runs, bond lengths were constrained at their equilibrium value using the Shake algorithm,[80] with a relative tolerance of 0.0001 Å. Ignoring high frequency bond vibrations allows for a time-step of 1 fs to be used. Charge–charge long range forces were treated with the particle mesh Ewald method,[81,82] using a convergence parameter α of $5.36/2R_c$ and a 4th order spline interpolation, whereas the short range interactions were truncated at $R_c = 10$ Å using standard corrections for energy and virial.[1] All simulations in the isothermic-isobaric (NPT) ensemble were performed for 2 ns keeping temperature and pressure constant with the use of the weak coupling scheme of Berendsen et al.[83]

The equilibrated systems were characterized in the microcanonical (NVE) ensemble also by measuring dynamic single molecule and collective properties: the translational and rotational diffusion coefficients (D and D^{rot}) and the shear viscosity η_S. The isotropic translational diffusion coefficient is computed as

$$D = \lim_{t\to\infty} D(t) = \lim_{t\to\infty} \frac{1}{6t}\langle[\mathbf{r}(t) - \mathbf{r}(0)]^2\rangle \tag{12}$$

where $\langle \ldots \rangle$ indicates a double average over all configurations and molecules. Molecular spinning and tumbling have been studied by computing parallel ($D^R_{\parallel}$) and perpendicular ($D^R_{\perp,1}$ and $D^R_{\perp,2}$), rotational diffusion coefficients, respectively. The latter are defined as

$$D^R_k = \int_0^\infty \langle \omega_k(t) \cdot \omega_k(0)\rangle dt; \quad k = \parallel; \perp, 1; \perp, 2 \tag{13}$$

where $\omega_{\parallel}$, $\omega_{\perp,1}$, and $\omega_{\perp,2}$ are the parallel and perpendicular projections of the molecular angular velocity $\boldsymbol{\omega}$ onto the axis normal to the plane ring. Finally, shear viscosity is computed as

$$\eta_S = \lim_{t\to\infty} \frac{V}{6k_BT}\int_0^t C_\sigma(t')dt' \tag{14}$$

where V is the volume of the simulation box, k_B the Boltzmann constant and $C_\sigma(t')$ the correlation function of the off-diagonal elements of the stress tensor $\hat{\boldsymbol{\sigma}}$. The long time limit of η_S was thereafter extrapolated by fitting the resulting function with a double exponential, as done previously in other applications.[44]

Results

Reference QM calculations and basis set tuning

Four different pyridine–pyridine arrangements, shown in Figure 3, were chosen from preliminary calculations. For each arrangement, namely parallel face-to-face (PFF), antiparallel displaced (APD), T-shaped (Tup), and antiparallel face-to-face (AFF), three different dimer geometries were created, by varying the intermolecular distance R (Fig. 3). Interaction energies for the 12 resulting dimer conformations were estimated at the CCSD(T) level in the CBS limit according to Eq. (10) and all results are reported in Table 1.

Table 1. Energy contributions (kJ/mol) entering in the estimation of the CCSD(T)/CBS interaction energy according to Eq. (10) and computed for selected dimer geometries.

R (Å)	$\Delta E^{CCSD(T)}_{CBS}$	ΔCCSD(T)	$\Delta E^{MP2}_{cc-pVDZ}$	$\Delta E^{MP2}_{cc-pvTz}$	ΔE^{MP2}_{CBS}
PFF					
3.40	−2.94	9.21	1.23	−6.53	−12.15
3.80	−7.97	5.35	−4.86	−9.86	−13.32
4.50	−4.24	2.18	−2.80	−4.92	−6.42
APD					
3.00 (a)	−4.96	15.84	3.77	−10.64	−20.79
3.40 (a)	−17.54	8.67	−11.60	−20.23	−26.22
4.50 (a)	−7.93	2.07	−5.75	−8.31	−9.99
Tup					
4.50	−5.72	5.57	4.09	−5.18	−11.30
5.00	−12.61	2.52	−7.85	−12.16	−15.13
6.00	−5.80	0.71	−4.39	−5.66	−6.51
AFF					
3.20	−0.96	11.38	6.36	−4.62	−12.33
3.70	−13.49	5.88	−8.88	−15.18	−19.37
4.50	−7.43	2.19	−5.50	−7.96	−9.62

(a) displacement was set to 1.2 Å.

In agreement with the results reported in Refs. [52, 53], the antiparallel arrangements are the most stable, due to both the dispersion contribution and the favorable antiparallel disposition of the molecular dipoles. Conversely, when the dipoles are aligned (PFF), the interaction energy decreases considerably, similar to that found for the T-shaped conformations. The most stable complex is found at $R = 3.4$ Å in a APD conformation and its energy (−17.54 kJ/mol) is rather close to that reported

Table 2. Comparison between CCSD(T) and MP2/cc-pVDZ(mod) interaction energies (kJ/mol).

R (Å)	$\Delta E_{CBS}^{CCSD(T)}$	$\Delta E_{cc-pVDZ(mod)}^{MP2}$	$\Delta E_{MP2/cc-pVDZ(mod)}^{CCSD(T)/CBS}$
PFF			
3.40	−2.94	−3.92	0.98
3.80	−7.97	−8.69	0.72
4.50	−4.24	−4.52	0.28
APD			
3.00[a]	−4.96	−4.45	−0.51
3.40[a]	−17.54	−17.34	−0.20
4.50[a]	−7.93	−7.51	−0.42
Tup			
4.50	−5.72	−1.13	−4.59
5.00	−12.61	−10.92	−1.69
6.00	−5.80	−5.19	−0.61
AFF			
3.20	−0.96	−0.42	−0.54
3.70	−13.49	−13.53	0.04
4.50	−7.43	−7.38	−0.05
Other[52]			
P1a	−9.37	−11.60	−2.22
P1b	−10.63	−12.88	−2.25
P2a+	−11.63	−13.11	−1.48
P2a−	−15.48	−17.23	−1.75
P2b	−16.07	−17.71	−1.65
T1	−8.45	−9.90	−1.44
T2	−5.15	−4.47	0.68

Last rows refer to CCSD(T) data reported in Ref. [52], as well as the labels reported in the first column. [a] displacement was set to 1.2 Å.

(−16.07) by Ref. [52] for a similar (R = 3.4 Å; disp. = 1.6 Å) APD arrangement, estimated at the CCSD(T)/CBS with a slightly different procedure. The binding energy of the Tup arrangement at R = 5.0 Å) is also not far (−12.61 kJ/mol) to the value reported from Hohenstein and Sherrill[52] for a Tup conformer at R = 5.1 Å (−8.45 kJ/mol).

Conversely, it is evident by looking at the last three columns of Table 1 that MP2 with large basis sets strongly overestimates the binding energy, in all the considered conformations. This is not surprising, as agrees with the results[59,61–64] reported from several groups in the case of aromatic–aromatic interactions. Moreover, by looking at the interaction energies obtained at the MP2/cc-pVDZ level, it appears that, also in the case of the pyridine dimer, the use of small basis sets tends to compensate the overestimation error. Following Hobza's idea,[57,65] a cc-pVDZ(mod) basis set is therefore devised to be coupled with the MP2 calculations to yield reasonable energies for the pyridine compound. To this aim, basis set polarization exponents on both nitrogen and hydrogen atoms were optimized to reproduce CCSD(T)/CBS values obtained for the PFF, APD and Tup arrangements. Conversely, the AFF set was not included in the exponent optimization procedure, but it was used as a validation test for the "predictive" capabilities of the MP2/cc-pVDZ(mod) estimates. The iterative optimization procedure, performed through the EXOPT software, yielded a standard deviation of 1.7 kJ/mol, with the largest error of 4.6 kJ/mol, found for the Tup arrangement at short distances. The best exponents resulted to be α_N = 0.41 and α_H = 0.32 for N and H atoms, respectively, while

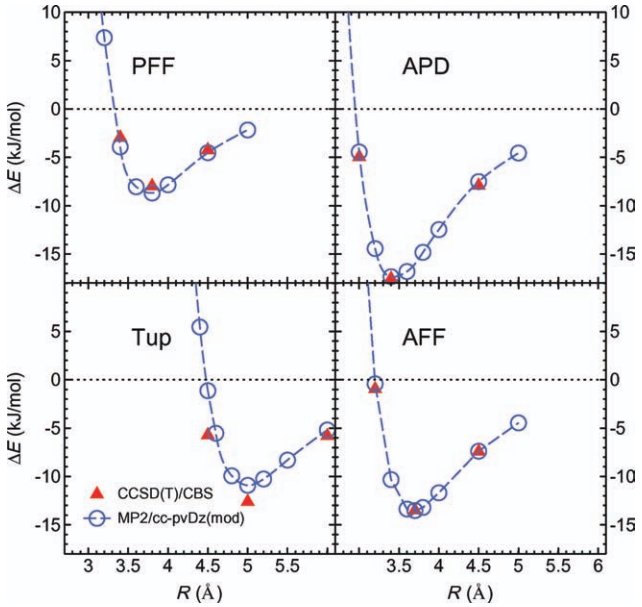

Figure 4. Comparison between CCSD(T)/CBS and MP2/cc-pVDZ(mod) pyridine interaction energies in the selected dimer arrangements. Labels refer to those reported in Figure 3.

the one for C atoms was left to 0.25 as suggested by Hobza and coworkers.[57,65] All CCSD(T)/CBS estimates are compared with the energy values obtained by the MP2/cc-pVDZ(mod) calculations in Table 2 and Figure 4. In the bottom right panel of the latter, the comparison between CCSD(T) and MP2 data is reported for the AFF arrangement, which was excluded from the database used for the exponent optimization procedure. It appears that the MP2/cc-pVDZ(mod) procedure is able to predict intermolecular energies in good agreement with those obtained with the higher level calculations. Moreover, for the Tup conformations, the rather large overestimation found at 5.0 Å seems to derive from a very slight shift, in the x-axis, between the two curves.

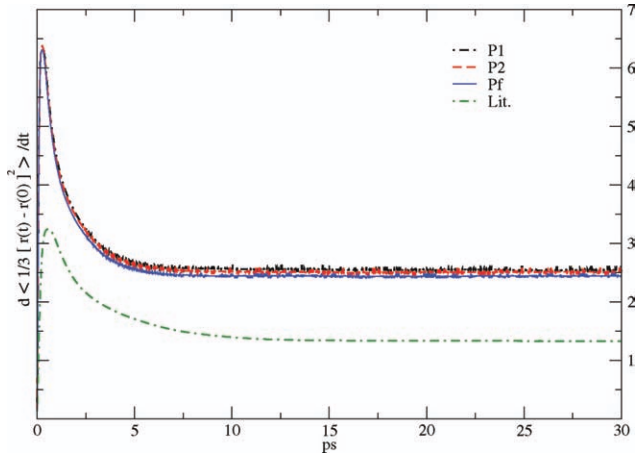

Figure 5. Mean square displacement: its derivative in time is presented for all sets. Half the values of the curves at reasonable long time limit represent the diffusion coefficients, D, in 10^{-8} m^2/s.

A further validation of the reliability of the MP2/cc-pVDZ(mod) results comes from the comparison with the estimates reported in Ref. [52], shown in the last rows of Table 2. The MP2 values differ in average by less than ~2 kJ/mol, maintaining the same relative ordering among the considered geometries. These results reinforce the suitability of this technique as a reliable tool to sample pyridine dimer PES at an affordable computational cost.

FF parameterization

The parameters of the FF intramolecular part, derived from accurate QM calculations, were already reported in Ref. [34]. The internal FF is determined once and for all and is never changed during this study.

OPLS atomistic parameters,[55] q, σ, and ϵ, were used as starting values for the iterative PICKY procedure. The set is reported in Table 4, together with the final optimized parameters. In the first iteration, a system of 512 molecules was equilibrated at 298 K and 1 atm starting from a simple cubic arrangement. To gain experience and to test the robustness of the present method, two different snapshots of this first equilibrated system were chosen. In this way two different paths (P1 and P2) both starting from the OPLS FF parameters and following different routes towards the final FF are considered. This procedure may provide useful information on the dependence of the final FF from the choice of the snapshot and other possible details of the algorithm which are rather arbitrary, in the sense that a well-established protocol is hard to be precisely defined.

The final QM database consists of 450 dimers, gradually increased in five iterations, for both paths P1 and P2, according to Table 3. In the same table, the constraints on R_{AB} and E_{inter} and the pressure and temperature of the NPT simulations are reported for each iteration. The first iteration is oriented to include the lowest energy portion of the PES in the QM database. The fitting of the resulting QM database was accomplished, through the minimization of functional (7) as described in "Computational Details" section, allowing the parameters to change up to ±10%, to avoid spurious minima in view of the limited size of this first database. Runs II and III extend the range of intermolecular energies thus improving the PES sampling; the parameters were here allowed to change to a larger extent. Run IV was performed at high T and P with the clear goal of forcing the system to populate portions of the PES corresponding to short intermolecular distances. In the last two steps, the parameters were allowed to change with no constraint.

The final parameters are reported in Table 4 for both P1 and P2. The two sets have very similar standard deviations, 1.04 and 0.98 kJ/mol, respectively. It is evident that final values of the parameters of P1 and P2 show some marked differences both in ϵ and in σ, whereas minor differences are observed for the atomic charges. The changes with respect to the starting OPLS values are even more relevant. In particular, the final ϵ values of C and N atoms are strongly increased, whereas those of the H atoms are decreased to very low values. The charges tend to diminish with respect to the starting values, preserving the sign in all cases. Consequently, a lower dipole moment is expected both for P1 and P2.

Table 3. Sampling criteria for intermolecular parameterization iterations applied in paths P1 and P2. In the last two rows are reported the standard deviations ΔP, defined in Eq. (15) between the two PESs generated with the parameters of a pair of subsequent runs, in path P1 and P2, respectively.

	Run I	Run II	Run III	Run IV	Run V
n. sorted dimers	50	100	100	100	100
QM database	50	150	250	350	450
R_{AB} (Å)	<10.0	<7.0	<7.0	<7.0	<7.0
E_{inter} (kJ/mol)	<−1.0	<1.0	<3.0	<3.0	<3.0
P (atm)	1	1	1	1000	1
T (K)	298	298	298	400	298
$\Delta P_{path\ 1}$ (kJ/mol)	0.239	0.178	0.073	0.074	0.038
$\Delta P_{path\ 2}$ (kJ/mol)	0.270	0.165	0.103	0.072	0.019

Table 4. Intermolecular parameters for pyridine model obtained with PICKY parameterization paths (P1, P2, and **Pf**).

Site	Start	P1	P2	**Pf**
ϵ (kJ/mol)				
N	0.711	1.162	0.791	0.959
C1	0.293	0.398	0.473	0.440
C2	0.293	0.380	0.348	0.369
C3	0.293	0.815	0.881	0.825
H1	0.125	0.005	0.005	0.005
H2	0.125	0.005	0.005	0.005
H3	0.125	0.005	0.005	0.005
σ (Å)				
N	3.25	3.00	3.14	3.06
C1	3.55	3.42	3.35	3.38
C2	3.55	3.49	3.52	3.51
C3	3.55	3.34	3.39	3.35
H1	2.42	2.86	2.85	2.86
H2	2.42	2.83	2.85	2.83
H3	2.42	2.88	2.83	2.86
q (e^-)				
N	−0.678	−0.532	−0.517	−0.521
C1	0.473	0.283	0.253	0.267
C2	−0.447	−0.301	−0.289	−0.295
C3	0.227	0.040	0.064	0.052
H1	0.012	0.066	0.076	0.070
H2	0.155	0.143	0.137	0.140
H3	0.065	0.110	0.099	0.105

Because of the aforementioned differences, it might seem that our final FFs are very different both from the starting OPLS one and to each other. It is in some way unpleasant that the parameters obtained through two different paths (P1 and P2) do not eventually coincide, as this could be a signature of an unwanted dependence of the final FF from some details of the protocol, like the initial choice of the snapshot in a MD run. Moreover no reliable criteria to drive this choice seems to exist. As it will be discussed in the following in some details, the reason of such an apparent discrepancy between the two final parameter sets is strongly connected with the marked redundancy of the atom–atom LJ and charge–charge functions. This redundancy can be partially reduced by increasing the QM database but can never be eliminated. A rather unfortunate consequence is

that, in the case of full redundancy, the same function can be represented with different sets of parameters. Therefore, also considering the Lorentz–Berthelot mixing rules, marked differences between the atomic parameters of several sets do not necessarily lead to marked differences in the resulting PES. In fact, as it will be showed in the following, the P1 and P2 PESs are more similar to each other as it would seem at first sight, and lead to the prediction of very similar bulk properties.

A further insight into this problem is provided by the last two rows of Table 3 which report the standard deviation between the dimer PESs generated by two consecutive sets of parameters, i.e.

$$\Delta P_k = \left(\frac{1}{N_{\text{geoms}}} \sum_{i}^{N_{\text{geoms}}} \left[\left(E_i^{\text{inter}}\right)_k - \left(E_i^{\text{inter}}\right)_{k-1} \right]^2 \right)^{\frac{1}{2}} \quad (15)$$

where $(E_i^{\text{inter}})_k$ is the intermolecular dimer energy computed with the FF parameters, according to Eqs. (1) and (2), obtained at the kth parameterization step for the ith geometry. This quantity allows for an estimate of the variation of the FF PES, on going form parameterization run $k-1$ to k. The sum is over a dense six-dimensional grid of geometrical points (X,Y,Z, α, β, γ), which identifies the relative geometry of the two molecules in the dimer. Clearly, the unphysical points corresponding to overlapping molecules are not included in the grid. Both the data reported in Table 3 show that the change of PES along the runs tends to diminish, being the last changes for paths P1 and P2 0.038 and 0.019 kJ/mol, respectively, both below a threshold value of 0.05 kJ/mol, which was considered small enough to stop the iterations. The same quantity computed between the PESs generated with the final parameters of the two paths amounts to 0.043 kJ/mol, which is higher than the change of PES obtained in the last run of P1 or P2, but still lower the chosen threshold. This seems to indicate that the convergence of both paths is rather slow and that a unique set of parameters could be obtained only for very extended QM database, even if the possibility of very similar potentials surfaces described by different set of parameters cannot be ruled out.

Before attempting such database extension, one possibility of reducing this redundancy problem could be to decrease the number of independent parameters to be optimized, by forcing similar atoms to take the same value of the FF parameters. Thus, we impose the same σ and ϵ LJ parameters to all hydrogen atoms (σ_H and ϵ_H) and to the C1, C2 sites (σ_C and ϵ_C). With this reduced number of free parameters the standard deviations of the minimization of functional (7), performed on each dimer database of the last runs of the two paths, increase from 1.041 to 1.095 kJ/mol for P1, and from 0.976 to 1.021 kJ/mol for P2. The ΔP function between these new PESs and the PESs yielded by the final parameters of Table 4, is very small for P1, whereas larger differences are found for P2. Indeed, owing to the larger difference between the parameters of C1 and C2 in the original P2 parameter set, this results is not surprising. However, the difficulty in finding a unique set of parameters is not overcome, as the parameters of P1 and P2 still show some differences. This again seems to indicate that a unique FF, within the limitation of the chosen intermolecular functional form, cannot be obtained with the current QM databases.

A computationally costless solution is to extend database dimensions by joining the the two QM sets, obtained in paths P1 and P2 respectively, into a unique database containing 900 dimer energies. This has been done and the resulting extended database was used for two additional fittings, performed by minimizing functional (7) using as starting parameters the P1 and P2 parameters sets, respectively. Both fittings yielded a final standard deviation of 1.023 kJ/mol and eventually converged to a unique set of parameters, named **Pf** and reported in Table 4. Not surprisingly, the final **Pf** parameters show values which are intermediate between those of sets P1 and P2. Nevertheless, it should be noted that the difference between the PESs subtended to the P1, P2, and **Pf** parameters sets is very small, and this is expected to reflect in negligible differences in the properties of the three bulk systems described by MD simulations performed with the different sets. In fact, in view of the results reported in the following sections, all sets turned out to be able to accurately describe the bulk phase properties of the target system, confirming that a ΔP convergence threshold of 0.05 kJ/mol is an appropriate indication to stop the iterations of the PICKY procedure.

FF validation: MD runs on pyridine liquid phase

Pyridine bulk structural and dynamic properties were computed by means of MD simulations, performed at 298 K and 1 atm with all sets FF parameters, as obtained from sampling paths (P1 and P2) and from the fitting with the extended database (**Pf**). Within fluctuations, all three runs ended up in describing two pyridine bulk liquid phases with the same macroscopic properties.

Table 5 shows dipole values, bulk density, and vaporization enthalpy obtained for all iterations of parameterization paths P1 and P2 and for the final set **Pf**.

Table 5. Intermolecular parameterization iterations and final results obtained with the **Pf** set.

	Dipole (D)	Density (g/cm^3)	ΔH_{vap} (kJ/mol)
Run 0 (OPLS)	2.35	0.963 ± 0.002	41.6 ± 0.2
Run I	2.19	1.063 ± 0.003	46.9 ± 0.4
	2.00	0.980 ± 0.004	40.7 ± 0.2
Run II	2.23	1.011 ± 0.005	43.3 ± 0.3
	2.18	0.998 ± 0.002	41.0 ± 0.3
Run III	2.20	1.005 ± 0.005	41.9 ± 0.3
	2.20	1.001 ± 0.003	40.8 ± 0.2
Run IV	2.15	1.018 ± 0.002	41.6 ± 0.2
	2.13	1.007 ± 0.004	40.2 ± 0.4
Run V (final)	2.11	1.017 ± 0.004	41.4 ± 0.3
	2.10	1.008 ± 0.005	40.6 ± 0.3
Pf	2.10	1.012 ± 0.004	40.9 ± 0.03
Exp	2.19	0.98[84]	40.3[55]

First and second row reported for each run refer to path P1 and P2, respectively. All MD runs are performed at 298 K and 1 atm (NPT ensemble).

The final densities for P1 and P2 sets are 1.017 and 1.008 g/cm^3, in rather good agreement with the experimental[84] value

of 0.98 g/cm^3. The relative error is 3.7 and 2.8% for set P1 and P2, respectively, not far from the one obtained with the OPLS set (2%), which was tuned to reproduce this property. The density yielded by the **Pf** set is 1.012 g/cm^3, an intermediate value with respect of the P1 and P2 sets and 3.1% larger than the experimental value. As vaporization enthalpies are concerned, the two parameterization paths again converged to very similar values (41.4 and 40.6 kJ/mol, respectively), in very good agreement with both experimental measures (40.3 kJ/mol, rel. err. 3 and 1.8%) and OPLS computed data (41.60 kJ/mol). The latter result is quite encouraging, as ΔH_{vap} is the second target property used to tune the OPLS parameters. Again, the result obtained using final **Pf** set of parameters resulted in a value (40.9 kJ/mol) averaged between the P1 and P2 ones and in excellent agreement with its experimental counterpart.

To further validate the present FF's, dynamic and transport properties of pyridine bulk phase were computed from MD simulations performed in the NVE ensemble, on the systems previously equilibrated at 298 K and 1 atm. Additionally, to compare all results achieved with P1, P2, and **Pf** sets with those obtained making use of literature parameters, another FF was constructed with OPLS parameters[55] for the intermolecular part, and AMBER parameters[18,85] for the description of intramolecular flexibility. A new set of MD (NPT and NVE) simulations was therefore carried with this literature FF (*Lit.*) on a system of 512 pyridine molecules. In accordance with the result of the guess set (Table 5), the *Lit.*-NPT simulations resulted in a bulk density of 0.962 $\pm$ 0.010 g/cm^3 and in a ΔH_{vap} of 41.21 $\pm$ 0.70 kJ/mol. The four sets, namely P1, P2, **Pf**, and *Lit.*, were then used in the NVE ensemble to compare their dynamic properties.

Diffusion coefficients were computed in a 10 ns run, using a correlation time window of 30 ps. D were obtained as stated in Eq. (12) and reported as a function of time in Figure 5, whereas their long time limit values are reported in Table 6. It is apparent from Table 6 the good accordance of the results of all parameterized sets with the experimental value.[86] More importantly, they show to be macroscopically equivalent also as translational diffusion is concerned. Conversely, the literature set performs much worse as the computed diffusion coefficient is largely underestimated $0.7 \cdot 10^{-9}$ m^2/s. It might be interesting to note that a similar deficiency of *Lit.* parameters was found also for the translational diffusion of the benzene molecules,[35,87] whereas a QM derived FF appeared[35] to give better results also in that case.

Table 6. Computed and experimental[86] translational diffusion coefficients [D, ($\times 10^{-9}$) m^2/s].

	P1	P2	**Pf**	*Lit.*	Exp[86]
D	1.3 $\pm$ 0.04	1.3 $\pm$ 0.04	1.3 $\pm$ 0.05	0.7 $\pm$ 0.04	1.54

As molecular rotational motions are regarded, the angular velocity autocorrelation functions were calculated from the NVE trajectories and reported in Figure 6 for all sets as a function of time. It can be observed from the upper panel that for all sets tumbling motions ($\alpha = \perp$), which correspond to rotations around one of the two axes lying in the plane of the heteroaromatic ring, are almost equal and much slower than the spinning one (rotation around the normal to the ring plane). These results confirm the presence of rotational anisotropy, as already found experimentally and in other MD studies (see Ref. [88] and references therein). All correlation functions were used to derive rotational diffusion coefficients, according to Eq. (13). Results are reported in Table 7, with their experimental counterparts. The ratio of the spinning and tumbling rotational coefficients is ~2.9, ~3.0, and ~2.7 for sets P1, P2, and **Pf**, respectively, in rather good agreement with the experimental ratio of 2.5.[89] However, the rotational diffusion coefficients are slightly underestimated, with respect to experimental values, by our FF and by the literature data, which are very similar. For the sake of comparison it should be also mentioned that the experimental values of the rotational diffusion coefficients are known to be highly dependent on the model adopted for their extrapolation.

Finally, shear viscosity was investigated correlating the off-diagonal elements of the stress tensor $\hat{\sigma}$ for 50 ps and extrapolating its value with a double exponential function. η final values are reported in Table 8 for all sets. The experimental value for the shear viscosity of pyridine at 298 K ranges from 0.855[84] to 0.877 mPa/s.[86] Both values can be considered in

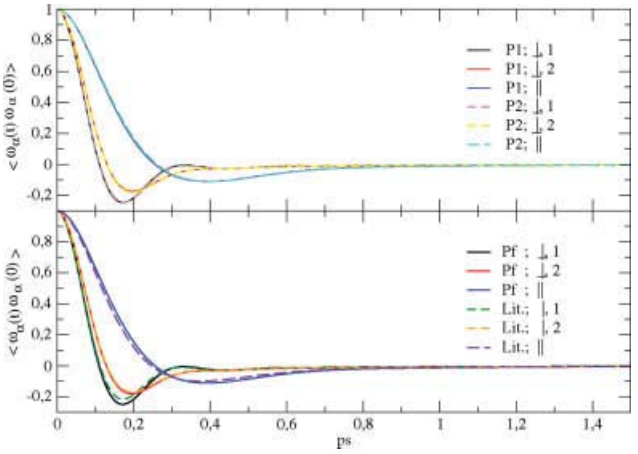

Figure 6. Angular velocity autocorrelation functions. $\alpha = \parallel$ is related to spinning motion, while $\alpha = \perp$ components are the projection of ω along the N—C_3, C_2—C_2 axes, respectively, and describe the molecular tumbling. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 7. Rotational diffusion coefficients (10^{10} rad^2/s)..... $\perp$, 1, and $\perp$, 2 refer to the N—C_3 and C_2—C_2 axes, respectively.

	P1	P2	**Pf**	*Lit.*	Exp[89]
$D^{rot}_{\perp,1}$	2.4	2.4	2.8	2.7	4.0
$D^{rot}_{\perp,2}$	3.4	3.2	3.4	3.2	4.0
$D^{rot}_{\parallel}$	8.3	8.5	8.3	7.9	10.0

$\parallel$ refers to the ring normal represents the rotation in the molecular plane.

Table 8. Shear viscosity data in mPa/s for all sets.

	P1	P2	**Pf**	*Lit.*	Exp[84,86]
η	0.86 ± 0.2	0.87 ± 0.2	0.86 ± 0.2	0.83 ± 0.2	0.85 – 0.88

very good agreement with those found with all our FF's and Lit.

The overall picture that arises from the comparison of the computed bulk properties counterparts makes us confident of the proposed parameterization approach, that has proven to be capable of reproducing with good accuracy structural and dynamic condensed phase properties of the target molecule. A further evidence of the method robustness, can be deduced from the comparison of the computed properties between P1 and P2 parameterization paths and the final **Pf** set, which yield very similar results, in contrast with the seemingly different values of their underlying FF parameters.

Conclusions

The present work situates itself in a broader framework of a research project, started some years ago,[34,37,38] and aimed to compute condensed phase properties with the aid of data obtained by QM calculations. In this context, the main goal of this article was to validate the new sampling strategy coded into the PICKY software. The pyridine molecule was chosen as a test case, and selected structural and dynamic properties were computed by performing MD-NPT simulations with purposely QM parameterized FF. It should be pointed out that, rather than exploiting the wealth of detailed information provided by MD simulations to gain a deeper insight into the physics of pyridine bulk phase itself, this article is focused on the validation of the FF through the comparison of the computed properties with their experimental counterparts.

The accurate yet affordable *ab initio* scheme [MP2 with cc-pvDz(mod) basis set], adopted for the FF parameterization, was validated by comparing the results obtained for selected dimer interaction energies with those yielded by higher level techniques, either reported in literature or purposely calculated. As the difference with more expensive techniques (CCSD(T)/CBS) is in average less than 2 kJ/mol, this scheme can be confidently used when extensive sampling of vdW PES is requested.

The iterative scheme for FF parameterization converged in few iterations, but two different sets of parameters were obtained from two different paths. Nevertheless, for all computed properties, both sets yielded almost identical results, suggesting that a certain degree of redundancy is implicit in the adopted model FF functions. Moreover, even when both set converged to the final **Pf** set, by extending the fitting to the larger QM database, the same averaged properties were obtained, within fluctuations. Indeed, the overall picture emerging from all computed bulk properties shows that both QM-derived FF are able to describe in a quantitative way a wide array of structural and dynamical properties of pyridine, ranging from vaporization enthalpy to collective dynamics.

More importantly, when compared to the results obtained with QM derived FF constructed with the previous sampling strategy,[35,37–39] the PICKY algorithm shows an overall improvement. In fact, the most noticeable weak points of the previous strategy were less accurate estimates of of the bulk density and the single particle dynamics. In this case, the density has been reproduced with an average error of ~3%, which meliorates the ~6% error reported in previous parameterizations,[38] and well compares with that obtained with density-tuned FF (~2%). As the description of pyridine dynamics is concerned, both single-molecule and collective properties are also in good agreement with the experiment. Furthermore, it is worth noticing that, as already found for the benzene molecule,[35] QM-derived FF's give a better description of the translational dynamics with respect to the literature FF adopted for comparison.

A last important point to be emphasized deals with molecular flexibility. The discussed improvement is connected not only to a more reliable sampling of the dimers selected for QM calculations, but also with the possibility of relaxing the internal geometry, which is expected to be important, above all, at short intermolecular distances. If the overall rigidity of the pyridine molecule may not adequately account for such improvement of the parameterization method, we expect this feature to be much more important when dealing with large flexible molecules, which is our ultimate aim. In fact, if a strictly theoretical approach to the FF development is to be maintained, the interplay between intermolecular and intramolecular interactions can not be neglected, and parameterizations of larger and less rigid molecules are currently in progress.

Keywords: force-field parameterization • aromatic interactions • pi-pi stacking • molecular dynamics

How to cite this article: I. Cacelli, A. Cimoli, P. R. Livotto, G. Prampolini, *J. Comput. Chem.* **2012**, *33*, 1055–1067. DOI: 10.1002/jcc.22937

[1] M. P. Allen, D. J. Tildesley, Computer Simulation of Liquids; Clarendon: Oxford, **1987**.

[2] D. Frenkel, B. Smith, Understanding Molecular Simulations; Academic Press: San Diego, **1996**.

[3] B. Kirchner, M. Reiher, In Theoretical Methods for Supramolecular Chemistry; C. Shalley, Ed.; Wiley-VCH Verlag Gmbh & Co. KGaA: Weinheim, **2007**.

[4] P. Pasini, C. Zannoni, Advances in the Computer Simulations of Liquid Crystals NATO ASI series, Kluwer: Dordrecht, **2000**.

[5] K. Kremer, *Macromol. Chem. Phys.* **2003**, *204*, 257.

[6] P. Pasini, C. Zannoni, S. Zumer, Eds. Computer Simulations of Liquid Crystals and Polymers NATO ASI series; Kluwer: Dordrecht, **2005**.

[7] C. M. Care, D. J. Cleaver, *Rep. Prog. Phys.* **2005**, *68*, 2665.

[8] M. Karplus, J. A. McCammon, *Nat. Struct. Biol.* **2002**, *9*, 646.

[9] A. D. Mackerell, *J. Comput. Chem.* **2004**, *25*, 1584–1604.

[10] M. W. van der Kamp, K. E. Shaw, C. J. Woods, A. J. Mulholland, *J. R. Soc. Interface* **2008**, *5*, S173–S190.

[11] R. Berardi, L. Muccioli, C. Zannoni, *Chem. Phys. Chem.* **2004**, *5*, 104.

[12] L. De Gaetani, G. Prampolini, *Soft Matter* **2009**, *5*, 3517.

[13] G. Tiberio, L. Muccioli, R. Berardi, C. Zannoni, *Chem. Phys. Chem.* **2009**, *10*, 125.

[14] S. Hampton, P. Agarwal, S. Alam, P. Croziers, In Proceedings of the 2010 International Conference on High Performance Computing and Simulation, HPCS 2010 **2010**, art. no. 5547149, p. 98.

[15] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225.

[16] W. Damm, A. Frontera, J. Tirado-Rives, W. L. Jorgensen, *J. Comput. Chem.* **1997**, *18*, 1955.

[17] J. Hermans, H. J. C. Berendsen, W. F. van Gusteren, J. P. V. Postma, *Biopolymers* **1984**, *23*, 1.

[18] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollmann, *J. Am. Chem. Soc.* **1995**, *117*, 5179.

[19] J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157.

[20] B. R. Brooks, R. E. Bruccoeri, B. D. Olafson, D. J. States, S. Swaminanthan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187.

[21] A. D. MacKerell, D. Bashford, M. Bellott, R. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586.

[22] N. Foloppe, A. D. MacKerell, *J. Comput. Chem.* **2000**, *21*, 86–104.

[23] A. Pérez, I. Marchán, D. Svozil, J. Šponer, T. Cheatham III, C. Laughton, M. Orozco, *Biophys. J.* **2007**, *92*, 3817.

[24] P. Jurečka, J. Černỳ, J. Šponer, P. Hobza, *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.

[25] R. Paton, J. Goodman, *J. Chem. Inf. Model.* **2009**, *49*, 944.

[26] M. Kolář, K. Berka, P. Jerečka, P. Hobza, *Chem. Phys. Chem.* **2010**, *11*, 2399.

[27] M. Zgarbová, M. Otyepka, J. Šponer, P. Hobza, P. Jurečka, *Phys. Chem. Chem. Phys.* **2010**, *12*, 10476.

[28] A. J. Misquitta, R. Podeszwa, B. Jeziorski, K. Szalewicz, *J. Chem. Phys.* **2005**, *123*, 214103.

[29] A. Hesselmann, G. Jansen, M. Schutz, *J. Chem. Phys.* **2005**, *122*, 014103.

[30] S. Dasgupta, T. Yamasaki, W. A. Goddard III, *J. Chem. Phys.* **1996**, *104*, 2898.

[31] K. Palmo, B. Mannfors, N. G. Mirkin, S. Krimm, *Biopolymers* **2003**, *68*, 383.

[32] J. R. Maple, M.-J. Hwang, T. P. Stockfish, U. Dinur, M. Waldman, C. S. Ewig, A. T. Hagler, *J. Comput. Chem.* **1994**, *15*, 162.

[33] T. A. Halgren, *J. Comput. Chem.* **1996**, *17*, 490.

[34] I. Cacelli, G. Prampolini, *J. Chem. Theory Comput.* **2007**, *3*, 1803.

[35] I. Cacelli, G. Cinacchi, G. Prampolini, A. Tani, *J. Am. Chem. Soc.* **2004**, *126*, 14278.

[36] W. L. Jorgensen, D. L. Severance, *J. Am. Chem. Soc.* **1990**, *112*, 4768.

[37] C. Amovilli, I. Cacelli, S. Campanile, G. Prampolini, *J. Chem. Phys.* **2002**, *117*, 3003.

[38] I. Cacelli, G. Prampolini, A. Tani, *J. Phys. Chem. B* **2005**, *109*, 3531.

[39] I. Cacelli, A. Cimoli, L. De Gaetani, G. Prampolini, A. Tani, *J. Chem. Theory Comput.* **2009**, *5*, 1865.

[40] A. Amovilli, I. Cacelli, G. Cinacchi, L. De Gaetani, G. Prampolini, A. Tani, *Theor. Chem. Acc.* **2007**, *117*, 885.

[41] I. Cacelli, L. De Gaetani, G. Prampolini, A. Tani, *Mol. Cryst. Liq. Cryst.* **2007**, *465*, 175.

[42] I. Cacelli, L. De Gaetani, G. Prampolini, A. Tani, *J. Phys. Chem. B* **2007**, *111*, 2130.

[43] L. De Gaetani, G. Prampolini, A. Tani, *J. Chem. Phys.* **2008**, *128*, 194501.

[44] M. Cifelli, L. De Gaetani, G. Prampolini, A. Tani, *J. Phys. Chem. B* **2008**, *112*, 9777.

[45] I. Cacelli, C. Lami, G. Prampolini, *J. Comput. Chem.* **2009**, *30*, 366.

[46] O. Akin-Ojo, Y. Song, F. Wang, *J. Chem. Phys.* **2008**, *129*, 64108.

[47] E. Meyer, R. Castellano, F. Diederich, *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 1210.

[48] W. Saenger, Principles of Nucleic Acid Structure; Springer-Verlag: New York, **1984**.

[49] S. Burley, G. Petsko, *Science* **1985**, *229*, 23.

[50] S. Tsuzuki, K. Honda, R. Azumi, *J. Am. Chem. Soc.* **2002**, *124*, 12200.

[51] M. Piacenza, S. Grimme, *Chem. Phys. Chem.* **2005**, *6*, 1554.

[52] E. Hohenstein, D. Sherrill, *J. Phys. Chem. A* **2009**, *113*, 878.

[53] B. Mishra, J. Arey, N. Sathyamurthy, *J. Phys. Chem. A* **2010**, *114*, 9606.

[54] C. Baker, G. Grant, *J. Chem. Theory Comput.* **2007**, *3*, 530.

[55] W. L. Jorgensen, N. A. McDonald, *J. Mol. Struct.* **1998**, *424*, 145.

[56] S. Tsuzuki, K. Honda, T. Uchimaru, M. Mikami, K. Tanabe, *J. Am. Chem. Soc.* **2002**, *124*, 104.

[57] P. Hobza, R. Zahradník, K. Müller-Dethlefs, *Collect. Czech. Chem. Commun.* **2006**, *71*, 443.

[58] K. Riley, M. Pitoňák, J. Černý, P. Hobza, *J. Chem. Theory Comput.* **2010**, *6*, 66.

[59] C. D. Sherrill, T. Takatani, E. Hohenstein, *J. Phys. Chem. A* **2009**, *113*, 10146.

[60] A. Vasquez-Mayagoitia, C. D. Sherrill, T. Aprà, B. G. Sumpter, *J. Chem. Theory Comput.* **2010**, *6*, 727.

[61] R. L. Jaffe, G. D. Smith, *J. Chem. Phys.* **1996**, *105*, 2780.

[62] P. Hobza, H. L. Selzle, E. W. Schlag, *J. Phys. Chem.* **1996**, *100*, 18790.

[63] S. Tsuzuki, T. Uchimaru, K. Matsamura, M. Mikami, K. Tanabe, *Chem. Phys. Lett.* **2000**, *319*, 547.

[64] M. Pitoňák, P. Neogrády, J. Černý, S. Grimme, P. Hobza, *Chem. Phys. Chem.* **2009**, *10*, 282.

[65] J. Špóner, J. Leszczynski, P. Hobza, *J. Phys. Chem.* **1996**, *100*, 5590.

[66] K. Riley, P. Hobza, *J. Phys. Chem. A* **2007**, *111*, 8257.

[67] L. Rutledge, H. Durst, S. Wetmore, *J. Chem. Theory Comput.* **2009**, *5*, 1400.

[68] Y. Zhao, D. Truhlar, *J. Chem. Theory Comput.* **2007**, *3*, 289.

[69] Y. Zhao, D. Truhlar, *J. Chem. Phys.* **2006**, *125*, 194101.

[70] Y. Zhao, D. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215.

[71] E. Honenstein, S. Chill, C. Sherrill, *J. Chem. Theory Comput.* **2008**, *4*, 1996.

[72] S. Grimme, *J. Comput. Chem.* **2004**, *25*, 1463.

[73] S. Grimme, *J. Comput. Chem.* **2006**, *27*, 1787.

[74] J. Chandler, *Behav. Sci.* **1969**, *14*, 81.

[75] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian 03, Revision C.02; Gaussian: Wallingford, CT, **2004**.

[76] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.

[77] C. D. Sherrill, B. Sumpter, M. O. Sinnokrot, M. Marshall, E. Hohenstein, R. Walker, I. Gould, *J. Comput. Chem.* **2009**, *30*, 2187.

[78] D. Truhlar, *Chem. Phys. Lett.* **1998**, *294*, 45.

[79] D. Paschen, A. Geiger, MOSCITO 3.9; Department of Physical Chemistry: University of Dortmund, **2000**.

[80] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *55*, 3336.

[81] T. A. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089.

[82] U. Essmann, L. Perera, M. L. Berkowitz, A. Darden, H. Lee, L. Pedersen, *J. Chem. Phys.* **1995**, *103*, 8577.

[83] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gusteren, A. Di Nola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684.
[84] J. N. Nayak, M. I. Aralaguppi, U. S. Toti, T. M. Aminabhavi, *J. Chem. Eng. Data* **2003**, *48*, 1483.
[85] E. Megiel, T. Kasprzycka-Guttman, A. Jagielska, L. Wrblewska, *J. Mol. Struct.* **2001**, *569*, 111.
[86] M. H. Muller, X. A. Mao, D. Seiferling, A. Sacco, *J. Chem. Phys.* **1996**, *104*, 669.
[87] R. Witt, L. Sturz, A. Dölle, F. Müller-Plathe, *J. Phys. Chem. A* **2000**, *104*, 5716.
[88] I. Bako, T. Radnai, G. Palinkas, *Z. Naturforsch. A* **1996**, *51*, 859.
[89] B. Stryczek, *J. Mol. Struct.* **1986**, *143*, 297.

Received: 31 August 2011
Revised: 23 December 2011
Accepted: 27 December 2011
Published online on 16 February 2012