

# JOYCE3.0: A General Protocol for the Specific Parameterization of Accurate Intramolecular Quantum Mechanically Derived Force-Fields

Samuele Giannini,<sup>a</sup> Pablo M. Martinez,<sup>b,c</sup> Abderrahmane Semmeq,<sup>a</sup>  
J. Pablo Galvez,<sup>d</sup> Anna Piras,<sup>a</sup> Alessandro Landi,<sup>e</sup> Daniele Padula,<sup>f</sup>  
J.G. Vilhena,<sup>b,c</sup> Javier Cerezo,<sup>g,\*</sup> and Giacomo Prampolini<sup>a,\*</sup>

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

<sup>b</sup>*Instituto de Ciencia de Materiales de Madrid (ICMM-CSIC), Madrid 28049, Spain*

<sup>c</sup>*Departamento de Física Teórica de la Materia (Universidad Autónoma de Madrid)  
E-28049 Madrid, Spain*

<sup>d</sup>*Departamento de Química Física (Universidad de Murcia),  
30100 Murcia, Spain*

<sup>e</sup>*Dipartimento di Chimica e Biologia "Adolfo Zambelli" (Università di Salerno)  
Via Giovanni Paolo II, 132, I-84084 Fisciano (SA), Italy*

<sup>f</sup>*Dipartimento di Biotecnologie, Chimica e Farmacia (Università di Siena)  
via Aldo Moro 2, 53100 Siena, Italy*

<sup>g</sup>*Departamento de Química and  
Institute for Advanced Research in Chemical Sciences (IAdChem) (Universidad Autónoma de Madrid),  
28049 Madrid, Spain*

---

\*email: giacomo.prampolini@pi.iccom.cnr.it, javier.cerezo@uam.es

## Abstract

While the intrinsically multi-scale nature of most advanced materials necessitates the use of cost-effective computational models based on classical physics, a reliable description of the structure and dynamics of their components often requires a quantum-mechanical treatment. In this work, we present JOYCE3.0, a software package for the parameterization of accurate, quantum-mechanically derived force-fields (QMD-FFs). Since its original release, the code has been extensively automated and expanded, with all novel implementations thoroughly discussed. To illustrate its general applicability, QMD-FFs are parameterized for seven benchmark cases, encompassing molecules with diverse structure and properties. These range from exotic stiff scaffolds, flexible polymeric chains, and polyenes of biological interest to transition-metal complexes. On the one hand, JOYCE3.0 FFs consistently outperform available general purpose descriptions, achieving excellent agreement with higher-level theoretical methods or available experimental validation data. On the other hand, the remarkable accuracy found in the description of the molecular structures extends to electronic excited states, enabling the integration of the JOYCE3.0 QMD-FFs into multi-level protocols aimed at reliably predicting selected properties and spectral line-shapes in advanced optoelectronic materials. The high quality of the results — spanning molecular structures, condensed-phase properties, and spectroscopic features — combined with the enhanced interface with popular quantum-mechanical codes and molecular dynamics engines, as well as its applicability to chemically diverse species, strongly suggests that JOYCE3.0 could play a pivotal role in the rational design of functionalized materials and heterogeneous systems.

# 1 Introduction

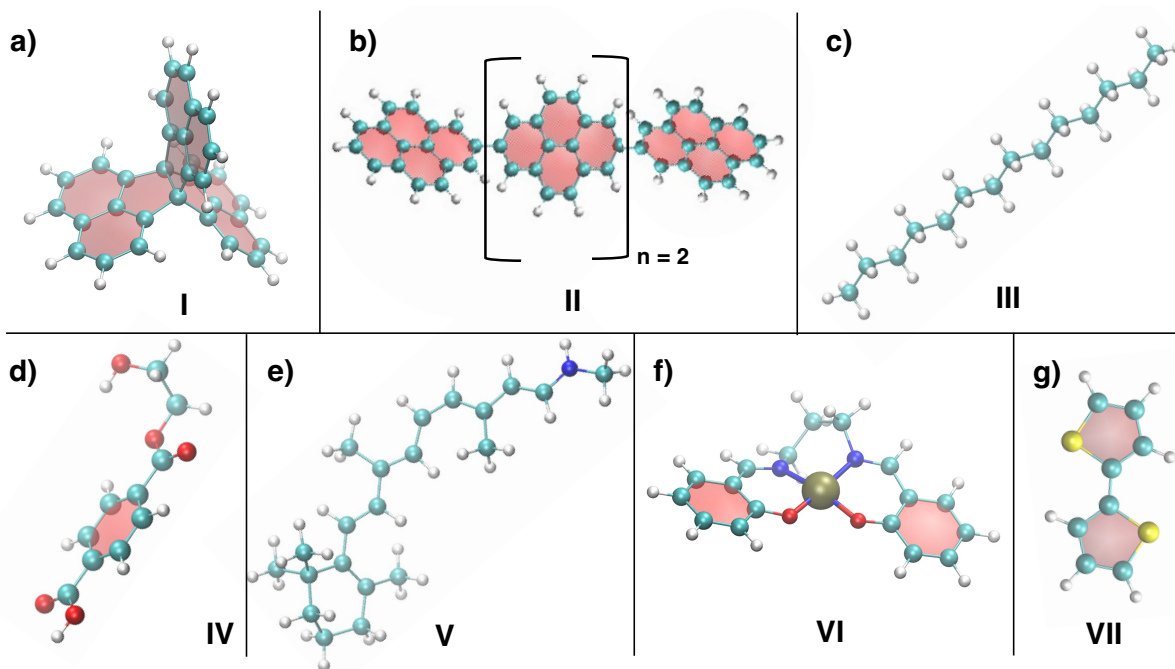
The functionalities and performances of advanced materials and novel devices rely on the ability to control processes at the molecular level. Such control, in turn, hinges upon our understanding of the detailed mechanisms involving each molecular player, the interactions among them, and their constituent atoms.<sup>1-4</sup> Indeed, a deep comprehension of molecular structures and supra-molecular interactions is decisive for the design and development of a plethora of advanced materials, as for instance dyes sensitized solar cells (DSSC) and photovoltaic devices,<sup>5,6</sup> functionalized smart polymers,<sup>7-9</sup> photonic liquid crystals<sup>10</sup> or transition metal based technological applications.<sup>11</sup> In the rational design of such materials, these requirements pose a significant challenge to computational techniques, which must achieve atomistic/molecular-level accuracy –often only attainable through quantum mechanical (QM) methods—while simultaneously accounting for a sufficiently large portion of the simulated material and providing a long enough observation time.<sup>12-14</sup> In this framework, computational techniques based on simplified descriptions of the potential created by electrons, namely Molecular Mechanics (MM), along with methods of classical (CL) physics, such as Monte Carlo (MC) and Molecular Dynamics (MD),<sup>15</sup> serve as a statistical tool for bridging such length/time scale gap. These techniques can be employed to gather reliable ensembles of (supra-)molecular configurations, which in turn can be used for successive QM calculations.<sup>14, 16, 17</sup> In such sequential CL/QM methods,<sup>16</sup> thermalized statistical ensembles of MC/MD configurations of large and complex systems are exploited in subsequent accurate QM calculations, allowing for an accurate characterization of material properties which intrinsically require a quantum treatment,<sup>18, 19</sup> as for instance spectroscopic features.<sup>16, 20, 21</sup>

The key ingredient of any CL simulation method is the force-field (FF), a collection of analytical model functions designed to represent the potential energy of the target system as a function of the nuclei's coordinates,<sup>15</sup> where the chemical identity of each system component is entailed in the parameters specifying the model functions. The FF should clearly be capable of faithfully and accurately describing such a potential as best as possible.<sup>22</sup> As a consequence, notwithstanding their computational convenience, general CL/QM multi-level approaches have

two main drawbacks: i) the so-called “structural mismatch”,<sup>23</sup> which stems from a description of the potential energy surface (PES) delivered by available general-purpose FF parameters that is significantly different from the one provided by the subsequent QM approach; and ii) the unavailability of FF parameters for specific components of the simulated material. To overcome these drawbacks, the possibility to derive from accurate QM calculations FF parameters suitable for simple CL simulations has been explored for small molecules since the late 1970’s.<sup>24</sup> Nonetheless, their application to more complex systems has attracted a renewed interest only more recently.<sup>12–14,25–29</sup> In fact, such an approach can in principle solve the two aforementioned drawbacks of general-purpose FFs based on transferability. On the one hand, tailoring the FF model on specific and accurate descriptors purposely obtained at QM level not only may allow for overcoming the structural mismatch, but also enforces the bridge connecting the CL and QM scales in subsequent analysis. On the other hand, modern electronic structure methods allow for the accurate representation of the structure and dynamics of molecules containing up to hundreds of atoms, thus paving the way for the FF parameterization of novel species or excited states.

In this context, our group has dedicated a long lasting and continuous effort in the development of reliable protocols to parameterize quantum-mechanically-derived FFs (QMD-FFs).<sup>25,30–32</sup> Despite the original procedures separately addressed the QMD-FF’s intra-<sup>31</sup> and inter-<sup>32</sup> molecular terms (which respectively govern the flexibility of each molecule and the interactions among them), a complete parameterization protocol has been also recently reported.<sup>33</sup> Therein, the inter-molecular parameterization was handled by the PICKY procedure,<sup>32</sup> whereas the QMD-FF intra-molecular parameters were obtained following the original JOYCE protocol.<sup>31</sup> Since then, the JOYCE code has been stepwise refined, and, along this development, it has been applied to different topics, ranging from the electronic properties of flexible organic molecules<sup>19,21,34</sup> and transition metal complexes (TMCs) with octahedral symmetry,<sup>35</sup> to the nano-scale friction of polymeric chains at metal surfaces.<sup>36,37</sup> Moreover, its capability of building reliable excited state FFs<sup>38,39</sup> has recently led to the integration of the JOYCE protocol in multi-level approaches for computational spectroscopy<sup>40</sup> and, more recently, to investigate the photophysics of molecular systems.<sup>41</sup>

Most of the aforementioned developments are now consistently implemented in a new version of the code, JOYCE3.0,<sup>42</sup> which constitutes the focus of this work. The whole parameterization procedure has been in fact significantly automated and updated: among the most relevant improvements, we highlight the new interfaces with electronic structure methods and popular MD codes, the automatic generation of the target molecule internal coordinates and the implementation of different routes to consistently tackle highly flexible molecular scaffolds. Moreover, all the documentation is now updated and available at a dedicated website,<sup>43</sup> where tutorials, templates and additional tools can be downloaded together with the code, hence making JOYCE3.0 a well-documented and user-friendly tool for accurate QMD-FF generation of varied materials. The aim of this paper is therefore three-fold: i) review and rationalize in a



**Figure 1:** Structure of the molecular targets considered in the present work. From a) to g): **I:** triperylene [3,3,3] propellane,<sup>44,45</sup> **II:** poly(2,7-pyrenylene),<sup>36,37</sup> **III:** *n*-tetradecane,<sup>46–48</sup> **IV:** ethylene terephthalate,<sup>49,50</sup> **V:** 11-*cis* retinal protonated Schiff base (rPSB11),<sup>51,52</sup> **VI:** *N,N'*-propylenebis(salicylaldiminato)nickel(II) ([Ni(salpn)]) complex,<sup>53</sup> **VII:** dithiophene.<sup>54,55</sup> In all cases, cyan, white, red, blue, brown and yellow spheres are employed for Carbon, Hydrogen, Oxygen, Nitrogen, Nickel and Sulfur atoms, respectively.

consistent framework the possible applications of the JOYCE-based QMD-FF parameterization protocol; ii) present the most significant features and strategies, which have been implemented

into the new version of the package; iii) discuss a novel application of QMD-FF in computational spectroscopy, and its possible impact on multiscale approaches. To this end, seven different target molecules, sketched in Figure 1, were parameterized with the new JOYCE3.0 code, and the resulting QMD-FFs were employed in MD simulations to retrieve selected structural, thermodynamic or spectroscopic properties of molecular as well as bulk morphologies. Such a selection is intended to show the general applicability of the JOYCE3.0 protocol, which can now be straightforwardly applied to remarkably different molecular targets, yet yielding consistent and highly accurate force-fields, which will be shown to outperform the performances delivered by general purpose descriptions.

## 2 Methods and Computational Details

### 2.1 JOYCE3.0 workflow

The general workflow of the JOYCE3.0 code is summarized in Figure 2. Following this scheme,

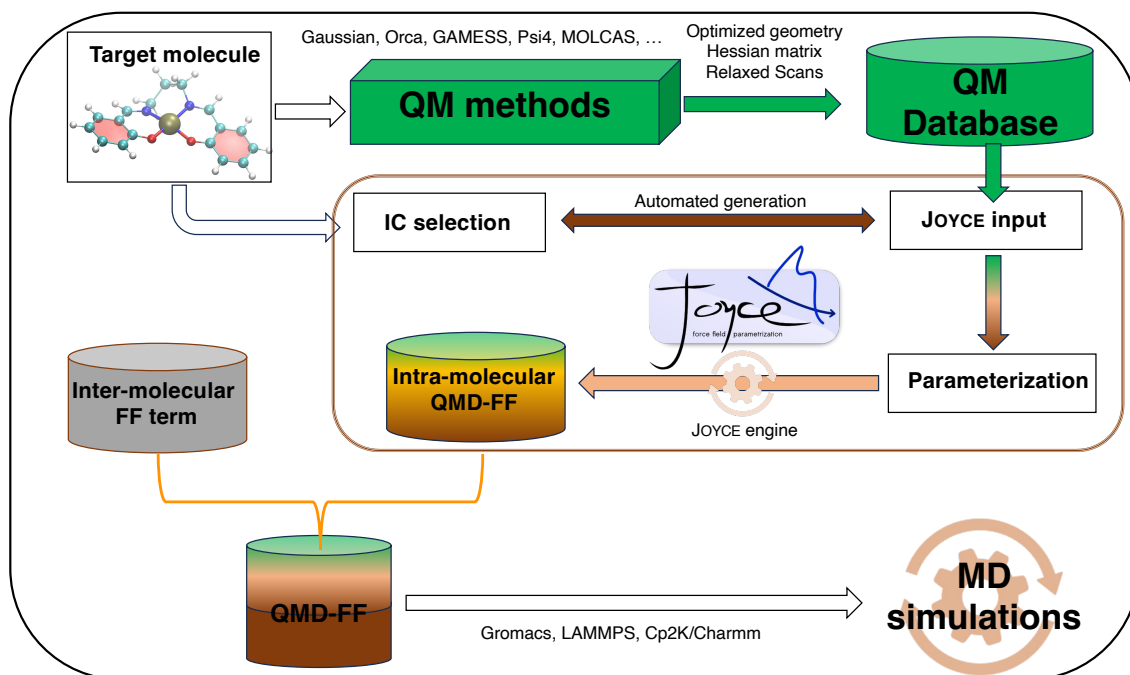


Figure 2: Workflow of the JOYCE3.0 scheme.

the parameterization protocol starts from the sole knowledge of the target molecule's chemical structure (top left corner). The latter is described in terms of internal coordinates (ICs), i.e. a collection of bond lengths, bond angles, dihedrals and selected intra-molecular distances. As discussed in the next sections, an important update with respect to the previous implementations of the code, consists in a direct generation of the target IC's, that are automatically prepared and classified by the new JOYCE3.0 package,<sup>42</sup> or by user's selection. Once all involved IC's have been automatically classified and selected, and possibly hand-refined by the user, the JOYCE3.0 protocol retrieves all parameters for the preparation of the QMD-FF, by deriving them from selected molecular descriptors purposely computed for the target at the QM level. This step is represented by the green color boxes in the upper part of Figure 2. It is important to mention that, while the previous versions of the code<sup>31</sup> were interfaced only with QM data obtained by the GAUSSIAN16 software,<sup>56</sup> in the JOYCE3.0 package the QM data can now be obtained using several popular packages suited for electronic structure calculations, such as ORCA,<sup>57</sup> GAMESS,<sup>58</sup> Psi4,<sup>59</sup> (Open)Molcas<sup>60</sup> and others. Concretely, the accepted format of the QM database has been extended from the GAUSSIAN16 *.fchk*, to the *FCclasses3* native *.fcc* format,<sup>61</sup> which includes the geometry, the energy, the gradient and the Hessian matrix. Such files can be automatically obtained from the output of the aforementioned electronic structure codes through the interconversion package *fcc\_tools*, recently distributed by some of us.<sup>62</sup> A detailed description of the *.fcc* format can be found in Section **S1** of the Supporting Information.

The final FF parameters defining the intramolecular term of the resulting QMD-FF<sup>31</sup> can then be obtained by the parameterization procedure described in Section 2.3. In this framework, a further new feature now available in JOYCE3.0, consists of the possibility of exporting the full list of parameters directly in the format expected by several MD packages of widespread use, such as GROMACS,<sup>63</sup> LAMMPS<sup>64</sup> or CHARMM (as implemented in the CP2K program package<sup>65</sup>) so that it can be added to any collection of intermolecular parameters. However, as discussed below, caution is necessary when using codes other than GROMACS in conjunction with the JOYCE QMD-FF, as not all functions are supported by other software. Notably, the JOYCE approach is based on an exact partition of the total QMD-FF energy into separate intramolecular and intermolecular terms, allowing for the straightforward merging of the

JOYCE3.0 parameters with the most suitable set of intermolecular ones, either sourced from literature databases or obtained through other QMD-FF parameterizations.<sup>33</sup>

## 2.2 QM calculations

As indicated in the top right corner of Figure 2, the QM calculations required by JOYCE3.0 pa-

Target	Class/Structure	QM database <sup>a)</sup>	Properties
<b>I</b>	polycyclic hydrocarbon stiff	PBE/6-31+G(d,p) 1 Hessian matrix	structural
<b>II</b>	polymer with PAH <sup>b)</sup> units semi-flexible	B3LYP/cc-pvDz 1 Hessian matrix 1 torsional scan	structural adhesion forces
<b>III</b>	polymer hydrocarbon flexible	cam-B3LYP/6-31+G(d,p) 1 Hessian matrix 5 torsional scans	structural bulk density viscosity heat capacity
<b>IV</b>	monomer of a polyester flexible	B3LYP/6-311+g(d,p) 1 Hessian matrix 7 torsional scans	structural
<b>V</b>	conjugated polyene flexible	MP2/6-31G(d) 1 Hessian matrix 10 torsional scans	structural spectroscopic
<b>VI</b>	transition metal complex square planar (stiff)	MP2/6-31G(d) 1 Hessian matrix	structural spectroscopic solvation
<b>VII</b>	polycyclic organic dye semi-flexible	PBE0/6-31G(d) 2 Hessian matrices <sup>c)</sup> 2 torsional scans <sup>c)</sup>	structural spectroscopic solvation

**Table 1:** Summary of investigated systems and computed properties. Systems (**I**: triperylene [3,3,3] propellane,<sup>44,45</sup> **II**: poly(2,7-pyrenylene),<sup>36,37</sup> **III**: *n*-tetradecane,<sup>46-48</sup> **IV**: ethylene terephthalate,<sup>49,50</sup> **V**: rPSB11,<sup>51,52</sup> **VI**: [Ni(salpn)] complex,<sup>53</sup> **VII**: dithiophene,<sup>54,55</sup> see Figure 1) are classified according to either their chemical class or the stiffness of their structure. <sup>a)</sup> level of theory (method/basis set) at which Hessian matrices and relaxed torsional energy scan were computed; <sup>b)</sup> Polycyclic Aromatic Hydrocarbon; <sup>c)</sup> Hessian matrix and torsional scans were computed for the ground and the first excited state.

parameterization consist in: i) a geometry optimization carried out in the target electronic state carried out specifically for each molecule; ii) a calculation of the Hessian matrix for each opti-



mized geometry; iii) a number of relaxed torsional energy scans along each rotatable dihedral (*vide infra*) of the considered species, obtained minimizing the internal energy with respect to all ICs, but the scanned one. To demonstrate JOYCE3.0's capability to parameterize QMD-FFs for diverse chemical species and across different levels of theory used to build the QM reference database, Table 1 summarizes the QM descriptors and methods employed in this work for the selected targets illustrated in Figure 1.

### 2.3 Intra-molecular QMD-FFs

As sketched in Figure 2, the computed QM database is the only input information required for the JOYCE3.0 parameterization engine. The main output consists in the list of all the FF parameters concerning the ICs ( $\mathbf{r}^{intra}$ ) of the target molecule, i.e. the intramolecular QMD-FF,  $E_{intra}^{QMD-FF}$ . The total FF potential energy ( $E_{tot}^{FF}$ ) is taken as a sum of the former with an intermolecular term,  $E_{inter}^{FF}$ , which instead depends on the distances ( $\mathbf{r}^{inter}$ ) between atoms belonging to different molecules:

$$E_{tot}^{FF} = E_{inter}^{FF}(\mathbf{r}^{inter}) + \sum_{k=1}^{N_{mol}} |E_{intra}^{QMD-FF}|_k(\mathbf{r}^{intra}) \quad (1)$$

Such an expression has been used for the compounds reported in Table 1. Notably, when performing simulations in condensed phase (i.e. for compounds **III**, **VI** and **VII**),  $E_{inter}^{FF}$  is expressed through a pairwise sum,

$$E_{inter}^{FF}(\mathbf{r}^{inter}) = \sum_{i=1}^{N_{at}} \sum_{j=1}^{N_{at}} u_{ij}^{inter}(\mathbf{r}^{inter}) \quad (2)$$

where  $N_{at}$  is the number of atoms of all species composing the system,  $i$  and  $j$  are dummy indexes running over the atoms of two different interacting molecules, and  $u_{ij}^{inter}$  takes the standard expression:

$$u_{ij}^{inter} = 4\varepsilon_{ij}^{inter} \left[ \left( \frac{\sigma_{ij}^{inter}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}^{inter}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{r_{ij}}. \quad (3)$$

Here,  $\varepsilon_{ij}^{inter}$  and  $\sigma_{ij}^{inter}$  are the 12-6 Lennard-Jones (LJ) parameters and  $q_i$  and  $q_j$  the Coulomb point charges (see Section S2 of the Supporting Information for further details). The first term

on the right side of equation (1) is null for an isolated molecule and the whole potential energy energy of the system in the gas phase coincides with the QMD intra-molecular term:

$$E_{tot}^{QMD-FF} = E_{intra}^{QMD-FF}(\mathbf{r}^{intra}) \quad (4)$$

In this work, to allow for the portability of JOYCE QMD-FF within the most popular MD engines mentioned in the previous paragraphs, the class I JOYCE expression<sup>31</sup> is adopted for  $E_{intra}^{QMD-FF}$ . That is, the molecular internal energy is further partitioned into five main contributions, namely

$$E_{QMD-FF}^{intra} = E_s + E_b + E_{st} + E_{ft} + E_{nb} \quad (5)$$

It is worth mentioning that equation (5) can be complemented with additional terms recently implemented within the JOYCE protocol,<sup>66</sup> which take into account also the couplings between the five contributions. This, in principle, increases even further the accuracy of the resulting QMD-FF.

Limiting the discussion to the intramolecular terms parameterized in the present work with the JOYCE3.0 code, the first three contributions on the right side of equation (5), refer to “stiff” IC’s ( $\mathbf{r}^{stiff}$ ), which are most likely subjected to small oscillations around their equilibrium value, and can be approximated through harmonic potentials.<sup>55</sup> Among these, the first two terms are routinely employed in most general-purpose force-fields, describing respectively bond stretching ( $s$ ) and angle bending ( $b$ ), i.e.:

$$E_s = \frac{1}{2} \sum_i^{N_s} k_i^s (r_i - r_i^0)^2 \quad (6)$$

and

$$E_b = \frac{1}{2} \sum_i^{N_b} k_i^b (\theta_i - \theta_i^0)^2 \quad (7)$$

where  $N_{s(b)}$  is the total number of stretching (bending) modes,  $k_i^{s(b)}$  the force constant and  $r_i^0$  ( $\theta_i^0$ ) the equilibrium value of the considered stiff coordinate  $r_i$  ( $\theta_i$ ). The third term,

$$E_{st} = \frac{1}{2} \sum_i^{N_{st}} k_i^{st} (\phi_i - \phi_i^0)^2 \quad (8)$$

refers to stiff torsions (*st*). Within the JOYCE protocol,<sup>31</sup> it is often employed to describe dihedrals  $\phi$  (as for instance those ruling the planarity of aromatic rings or conjugated double bonds), whose internal energy variation is better described by such an harmonic term with respect to the Fourier-like series usually employed by transferable FFs. In most large-sized molecules, however, the effectiveness of a local harmonic approximation (LHA) cannot be fully exploited. This is because these species are very often characterized by flexible “soft” coordinates, i.e. expected to experience large amplitude distortions during their dynamics. This characteristic requires model function beyond LHA. An example of these kind of ICs are those dihedrals  $\delta$  which define rotations around single C-C or C-O bonds, whose torsional potential may present multiple minima separated by rather accessible energy barriers. For the description of such flexible torsions (*ft*), JOYCE3.0 employs a Fourier-like sum:

$$E_{ft} = \sum_i^{N_{ft}} \sum_j^{N_i^{cos}} k_{ij}^{ft} [1 + \cos(n_{ij}\delta_i - \gamma_j)] \quad (9)$$

where  $N_i^{cos}$  is the user-selected number of cosine functions employed to model dihedral  $\delta_i$ , while  $k_{ij}^{ft}$ ,  $n_{ij}$  and  $\gamma_j$  are respectively the force constant, multiplicity and phase of the  $j^{th}$  term of the series. Finally, since the MD runs are expected to explore a large portion of the PES, in which the target molecule might be in different conformations, intramolecular non-bonded interactions may be introduced between selected atom pairs. This is necessary for instance to account for the steric repulsion between close atoms or the possible intramolecular hydrogen bonds (HBs). Such interactions can be introduced through specific LJ intramolecular model functions, depending on soft intramolecular distances ( $r^{soft}$ ):

$$E_{nb} = \sum_{i=1}^{N_{pairs}} \sum_{j=1}^{N_{pairs}} u_{ij}^{intra}(r_{ij}^{soft}) \quad (10)$$

where

$$u_{ij}^{intra}(r_{ij}^{soft}) = 4\epsilon_{ij}^{intra} \left[ \left( \frac{\sigma_{ij}^{intra}}{r_{ij}^{soft}} \right)^{12} - \left( \frac{\sigma_{ij}^{intra}}{r_{ij}^{soft}} \right)^6 \right]. \quad (11)$$

It should be stressed that  $\epsilon^{intra}$  and  $\sigma^{intra}$  are intramolecular parameters aiming at describing specific interactions occurring between selected atom pairs belonging to the same molecule. In

contrast to  $\epsilon^{inter}$  and  $\sigma^{inter}$ , reported in (3), which are instead designed to mimic the interaction among atoms pertaining to two different molecules. Consequently, *intra* and *inter* LJ parameters can have, in principle, different values. Further details about QMD-FF and their standard JOYCE parameterization are reported in the original paper,<sup>31</sup> while in the following section we describe the most relevant features introduced in this work within the new JOYCE3.0 procedure.

## 2.4 JOYCE3.0 parameterization and new features

### Atom types

Compared to the previous versions of the code,<sup>31,66</sup> the additional new features implemented in the JOYCE3.0 procedure are generally aimed at expanding the base of potential users. This was done by providing interfaces for JOYCE3.0 with more QM and MD codes, and making parameterization more user-friendly for nonexperts. Nonetheless, it should be emphasized that the accuracy of QMD-FF stems from its specificity, and on the possibility to tailor in detail the FF description for each considered target. For this reason, although a preliminary automatic atom type assignment can be obtained using popular software and web servers,<sup>67–69</sup> this assignment can be further refined by the user, allowing for the specification of user-customized atom types. To facilitate this process while providing sufficient flexibility and specificity, JOYCE3.0 now implements a four-subscript labeling system. Specifically, each atom composing the target molecule can take the label  $X_{abcd}$ , where  $X$  is the atom name, and  $a, b, c, d$  are alphanumeric characters, which can be used to identify a particular atom type.

Each  $X_{abcd}$  atom type should reflect symmetry considerations and chemical equivalence. The latter is defined in terms of topologically equivalent atoms, meaning they have identical environments in terms of connectivity: they are connected to the same number and type of atoms, and those atoms themselves exhibit the same connectivity in a recursive manner. This equivalence arises from the symmetry of the labeled molecular graph, a graph-theoretical representation of the molecule. Dependencies between FF parameters stem from the chemical equivalence of atoms, which determines the atom types to which they are assigned. Typically, the number of data points is smaller than the number of parameters in the FF, resulting in many strongly correlated parameters. To enhance the stability of the fitting process and avoid over-

fitting, equivalent force constants are not independently varied during parameterization. Additionally, the equilibrium values of equivalent FF terms are averaged to produce symmetrized values.

An alternative strategy to significantly reduce the number of fitting parameters involves determining equivalent atoms based on their local connectivity, defined up to a specific neighbor depth. This approach retains the specificity of QMD-FFs while leveraging the transferability of parameters between atoms expected to exhibit similar chemical behavior. Currently, we are pursuing different strategies<sup>43,62,70,71</sup> to efficiently produce a set of parameters with minimal intervention by the user, implementing them in the ecosystem of tools that enrich the JOYCE code. These strategies, generally exploit the powerful graph theoretical algorithms to inspect of the local environment (the  $k$ -depth neighborhood) of each atom. Namely, the local environment of an atom is represented by the subgraph comprising the atom itself and all atoms reachable within a depth  $k$ ; if any of the neighbors belong to a cycle, all atoms in the cycle are included. Local equivalence is established by identifying whether a local automorphism exists that maps the subgraph of one atom onto that of another, ensuring identical connectivity up to the chosen depth.

### **FF internal coordinates**

A crucial step in building any FF stands in the selection of the ICs used to represent the PES of the target molecule. In this respect, the previous versions of JOYCE code<sup>31,66</sup> provided a list of all possible ICs for the target molecule, and the initial, time-consuming step consisted in a manual classification and selection of the ones to retain. Although a 'black-box' strategy that retains all internal coordinates (ICs) might seem like the simplest option as it requires no user intervention, it can lead to overfitting issues due to the non-orthogonal and highly redundant nature of a basis set consisting of all ICs. In JOYCE3.0, once the atom types to be employed in the QMD-FF have been defined, the new code provides a preliminary list of all ICs (based on the optimized QM geometry) specifically classified in terms of bond stretching, angle bending, dihedrals and non-bonded distances and proposes options for a further refinement. Specifically, the ICs automated selection occurs by:

- i) identifying the connectivity within the molecule, determined using Pauling bond orders, and assessing the resulting covalent bond lengths based on whether the distance is no more than 30% longer than the sum of the covalent radii of the involved atoms.
- ii) exploiting the computed connectivity. That is, JOYCE3.0 builds and classifies all the ICs in terms of bonds ( $\mathbf{r}$ ), angles ( $\theta$ ), dihedrals ( $\phi$  or  $\delta$ ), and non-bonded intramolecular distances ( $\mathbf{r}^{soft}$ ).
- iii) refining the dihedrals selection and related atom quadruplets by defining each torsion according to the bond order of the inner atoms. That is, grouping the coordinates in stiff, ( $\phi$ ) or flexible ( $\delta$ ) coordinates according to equations (8) and (9) and automatically assigning them to either harmonic or periodic potential model functions, respectively.
- iv) automatically identifying the internal distances  $\mathbf{r}^{soft}$ . JOYCE3.0 provides a list of all atom pairs that are found to be more than three bonds apart, from which the user can select only the pairs necessary to represent specific intramolecular non-bonded interactions, such as those involved in intramolecular HBs, or to avoid 'bad contacts', which can occur in very flexible molecules, between atoms that are far apart.
- v) providing a starting topology file where all ICs are associated with the corresponding FF term, as expressed by equations (6) to (10), through which the QMD-FF parameterization can be started.

### Non-bonded intramolecular interactions

Another important novel implementation in JOYCE3.0 concerns the intra-molecular non-bonded term defined in equation (11) and the procedure to parameterize it. In most of the previous applications following the original JOYCE protocol,<sup>31,66</sup> the QMD-FF parameterization was carried out by minimizing the objective function

$$I^{intra} = \sum_{K \leq L}^{3N-6} \frac{2W_{KL}''}{C} \left[ H_{KL} - \left( \frac{\partial^2 E_{QMD-FF}^{intra}}{\partial Q_K \partial Q_L} \right) \right]_{g_0}^2 + \sum_g^{N_{geom}} W_g [\Delta U - E_{QMD-FF}^{intra}]_{g/g'}^2 \quad (12)$$

In the first term on the right of Eq. (12), the double sum runs over QM normal modes (where  $Q_K$  is the  $K^{th}$  mode),  $C$  is a normalization factor and  $N$  the number of atoms. The normalized

diagonal elements of the weight matrix  $\mathbf{W}''$  are usually set at twice the value of those corresponding to the off diagonal terms. The QM and QMD-FF Hessian matrix elements,  $H_{KL}^{QM}$  and  $\frac{\partial^2 E_{QMD-FF}^{intra}}{\partial Q_K \partial Q_L}$ , are both evaluated at the equilibrium geometry ( $g_0$ ). In the second term, which is introduced only in the case of molecules bearing flexible dihedrals  $\delta$ , the weights  $W_g$  are all initially set to the same value,  $\Delta U_g$  is the QM computed energy difference between the  $g$  and  $g_0$  geometries and  $N_{geom}$  is the number of the different geometrical arrangements sampled in the torsional scans for each considered  $\delta$  in equation (9). Concretely, the JOYCE parameterization was usually carried out minimizing<sup>31</sup> such an objective function through a two-step procedure, hereafter named **Route I**:

- i) a first step necessary to determine all harmonic parameters at once using the QM Hessian and minimizing the first term on the right hand side of the objective function (12)
- ii) only when flexible dihedrals are present in the target molecule, a second step consisting in constraining all harmonic parameters depending on  $\mathbf{r}^{stiff}$  (i.e.,  $E_s$ ,  $E_b$  and  $E_{st}$  in equation (5)), and parametrizing only the terms related to flexible dihedrals in  $E_{ft}$ . These terms use reference QM torsional relaxed energy scans and the algorithm exploits the so-called Frozen Internal Rotation Approximation (FIRA).<sup>31</sup>

At this point, it is important to notice that in **Route I**, the FIRA in step (ii), inevitably introduces a geometry mismatch in equation (12). In fact, for each geometry  $g$  considered in the relaxed scans, the the QM and QMD-FF descriptors can be respectively written as:

$$\Delta U_g = \Delta U(\delta_g) \quad (13)$$

and

$$E_{QMD-FF}^{intra}|_{g'} = E_{bonded}(\mathbf{r}_{g_0}, \boldsymbol{\theta}_{g_0}, \phi_{g_0}, \delta_g) + E_{nb}(\mathbf{r}_{g'}^{soft}) \quad (14)$$

where

$$E_{bonded}(\mathbf{r}_{g_0}, \boldsymbol{\theta}_{g_0}, \phi_{g_0}, \delta_g) = E_s(\mathbf{r}_{g_0}) + E_b(\boldsymbol{\theta}_{g_0}) + E_{st}(\phi_{g_0}) + E_{ft}(\delta_g) \quad (15)$$

This in turn implies that the energy difference in the second term on the right hand side in **Route I**'s objective function equation (12) is possibly evaluated at slightly different geometries, i.e.  $g$

(from the QM relaxed scan) and  $g'$  (obtained from  $g_0$  through the FIRA). This is formalized as:

$$\Delta U_g - E_{QMD-FF}^{intra}|_{g'} = \Delta U(\delta_g) - \left( E_{ft}(\delta_g) + E_{nb}(\mathbf{r}_{g'}^{soft}) \right) \quad (16)$$

In the above equation, all the harmonic terms in equation (15) vanished exploiting the FIRA and the remaining flexible torsions term  $E_{ft}$  is computed over the same  $\delta$  angle employed at QM level, thus limiting the mismatch due to the presence of the non-bonded intra-molecular term,  $E_{nb}$ . In absence of strong coupling terms between the selected ICs, the latter term is not required and **Route I** allows for assigning the whole QM relaxed energy sampled along the scans  $\Delta U_g$  to the flexible dihedral term  $E_{ft}(\delta_g)$ . Nonetheless, when dealing with large and very flexible molecules, where dihedral rotations might cause either specific internal interactions, as for instance intra-molecular HBs, or close contacts between far-lying atoms,<sup>21</sup> it becomes necessary to include specific intramolecular non-bonded terms. This introduces a possible geometrical mismatch. One of the main consequences of such a mismatch, is the general underestimation of the  $\mathbf{r}^{soft}$  distances evaluated at the unrelaxed  $g'$  FIRA geometry, which in turn leads to a significant overestimation of the intra-molecular LJ contribution in equation (11). To circumvent this issue, when including LJ intra-molecular terms, in previous work,<sup>21,66</sup> **Route I** consisted in a rather time-consuming user-defined inclusion and tuning of the selected interacting atom pairs and their  $\mathbf{r}^{soft}$  distances. This approach required the manual inclusion of specific pairs and their corresponding  $\mathbf{r}^{soft}$  distances after the initial parameterization. This process hence relied on a trial-and-error methodology, where yet adjustments were still guided by comparing the modified QMD-FF with the original QM database.

In the present work the parameterization procedure for the non-bonded terms has been revised and automated. This was achieved via the implementation of **Route II** in the JOYCE3.0 version of the code. For all flexible and semi-flexible molecules shown in Figure 1 and reported in Table 1 the parameterization is carried out by minimizing the JOYCE3.0 new objective function,

$$I^{intra} = \sum_{K \leq L}^{3N-6} \frac{2W''_{KL}}{C} \left[ H_{KL}^{corr} - \left( \frac{\partial^2 E_{bonded}}{\partial Q_K \partial Q_L} \right) \right]_{g_0}^2 + \sum_g^{N_{geom}} W_g [\Delta U^{corr} - E_{bonded}]_g^2 \quad (17)$$

where  $H_{KL}^{corr}$  is obtained from the original QM Hessian by subtracting the non-bonded con-



tribution contributions  $E_{nb}$  in equation (10) evaluated at the QM geometry  $g_0$ ,

$$H_{KL}^{corr} = \left[ H_{KL} - \frac{\partial^2 E_{nb}}{\partial Q_K \partial Q_L} \right]_{g_0} \quad (18)$$

and  $\Delta U^{corr}$  is retrieved from the torsional scans database by subtracting from the computed QM energy the  $E_{nb}$  contribution at the relaxed geometry  $g$ .

$$\Delta U^{corr} = [\Delta U - E_{nb}]_g \quad (19)$$

More specifically, molecules bearing flexible dihedrals, as test cases **II**, **III**, **IV**, **V** and **VII**, are parameterized with the JOYCE3.0 code, following the new automated **Route II** procedure:

- i) In addition to all bonded ICs, which are automatically generated by the JOYCE3.0 code, atom pairs required to represent specific intra-molecular non-bonded interactions—such as those involved in intra-molecular HBs or to prevent potential 'bad contacts' between atoms located far apart during dynamics—are included in the ICs collection.
- ii) based on the included pairs, the corrected QM database is computed by JOYCE3.0 through equations (18) and (19).
- iii) all harmonic parameters are obtained at once by minimizing the first term of the objective function (17)
- iv) the final QMD-FF, which now includes directly intramolecular LJ parameters, is eventually retrieved by applying FIRA and minimizing the second term of the objective function.

## 2.5 MD simulations

All classical MD simulations presented in this work were carried out with the specifically tailored QMD-FFs by using the GROMACS<sup>63</sup> package, unless otherwise stated. Additionally, to test and show the portability of JOYCE3.0 toward other popular MD codes, in selected cases (see Section **S3.1** in the Supporting Information) the LAMMPS<sup>64</sup> and CHARMM format (as implemented in CP2K program package<sup>65</sup>) engines were also employed. Three different types of MD runs were explored in this work: a) a NVT trajectory of the isolated molecule at 298 K

for each considered species; b) NPT runs in ambient conditions for condensed phase (liquid) homogeneous systems composed of one thousand molecules (here *n*-tetradecane, **VI**); c) NPT runs in ambient condition on solvated systems, composed by the either the target molecule **VI** or **VII**) embedded in  $\sim 1000$  solvent molecules. Further details on classical MD simulations can be found in Section **S3.2** of the Supporting Information. For two of the systems investigated in this work (namely, **I** and **II**), classical MD simulations were compared with ab initio MD trajectories computed using the DFT potential obtained by solving the electronic structure of the target molecule at each MD step in the Born-Oppenheimer approximation. This DFT-based approach is hereafter referred to as Born –Oppenheimer molecular dynamics (BOMD), and details on these calculations are given in section **S3.3** of the Supporting Information.

## 2.6 Computational Spectroscopy

The calculation of optical and spectroscopic properties of chromophores in solution or in complex environments requires methodologies and computational protocols that can account for the intrinsic electronic structure and vibrations of the excited molecule, as well as environmental effects. As mentioned previously, this necessitates reliable sampling of solute-environment configurations. To achieve this, sequential CL/QM methods with varying levels of accuracy have been developed.<sup>55,72–76</sup> These approaches can be broadly classified as classical and mixed quantum-classical (MQC) schemes. CL methods account for the effects of solute and environment fluctuations by obtaining the spectrum as the distribution of vertical energies. These can be computed at a reasonable computational cost exploiting the time dependent density functional theory (TDDFT), for example within a QM/MM approach, over an ensemble of frames extracted from a reliable MD trajectory. Such a protocol, which relies on the classical Franck-Condon principle, is referred to in this work as the Classical Ensemble Average of Vertical Energies (CEA-VE) and is used to compute the optical properties of molecule **V** to **VII**. Computational details for the CEA-VE method are given in the Supporting Information in Section **S4**. The main limitation is that the quantum nature of the vibronic features (and related broadening) typically due to high-frequency modes of the excited molecule are not accounted for in this classical approach. To go beyond CEA-VE, a QM treatment of high-frequency modes and vibronic

effects, which yet preserves a reliable description of the low frequency dynamical effects, is necessary.<sup>77</sup> MQC approaches, such as the cumulant method<sup>78</sup> and the Adiabatic-Molecular Dynamics Generalized Vertical Hessian Approach (Ad-MD|gVH),<sup>55</sup> have been developed to address this challenge. For target molecule **VII** (i.e. dithiophene), we utilize the Ad-MD|gVH, which partitions the coordinates of the entire system (solute + medium) into soft and stiff modes. The soft modes, including the flexible degrees of freedom (DoFs) of both the solute and solvent, are sampled using classical MD. In contrast, the stiff modes are treated at the vibronic level using harmonic potentials specific to each configuration of the soft modes. To this end, analytic formulations within the BO approximation of the vibronic spectrum have been derived, and are implemented in our open source code *ℱℒclasses3*.<sup>61</sup> The final Ad-MD|gVH spectrum is derived by averaging signals computed at the QM level for the stiff coordinates across the configurations obtained from the MD trajectory, which explicitly incorporates the equilibrated solvent. Further details on the methodology and related computations are given in Section **S4** of the Supporting Information.

### 3 Results and Discussion

The seven benchmark cases shown in Figure 1 has been chosen with the two-fold objective of illustrating the wide range of targets and applications that can be tackled with the JOYCE3.0 protocol, and discussing all new features implemented in this work. The selected target compounds give an overview about the JOYCE3.0 possible applications. Various chemical species has been chosen, ranging from simple organic molecules (**I**, **IV**, **V** and **VII**), to more complicated TMCs (**VI**), up to long and flexible polymers (**II** and **III**). These systems present a rather different chemical structure, going from stiff scaffolds (**I** and **VI**) to highly flexible chains and backbones (**III**, **IV** and **V**). Moreover, the resulting QMD-FFs, specifically obtained here for each target, are used to compute a plethora of distinct properties, which include structural, thermodynamic and spectroscopic features, in turn requiring a specific a well-tuned QM description, as reported in Table 1.

On the other hand, the target selection has been specifically made to discuss the new implementations. Compounds **I** to **III** exhibit an increasing degree of flexibility, requiring different

model FF functions. These compounds have been chosen to illustrate both the automated generation of ICs and the portability to other MD codes. Target molecule **IV** contains several flexible dihedrals and is thus characterized by a complex QM conformational energy landscape. Its FF representation necessitates careful tuning of non-bonded intramolecular interactions, which has been achieved through our new automated procedure. The 11-*cis* retinal protonated Schiff base (rPSB11) (compound **V**) is a well-known conjugated polyene, whose bond-length alternation (BLA) is crucial for accurately reproducing its spectroscopic behavior. This requires a precise representation of both stiff and flexible coordinates. The Ni complex (compound **VI**), on the other hand, was selected to demonstrate how the JOYCE3.0 procedure can be straightforwardly extended to transition metal complexes (TMCs) of any symmetry, such as the square planar geometry used here. Finally, to highlight the high degree of accuracy encoded in our QMD-FFs, dithiophene (compound **VII**) was chosen as an example of potential novel mixed quantum-classical approaches in computational spectroscopy.

### 3.1 I-III: Handling Stiff and Soft Torsions

In this section, we show how JOYCE3.0 handles the different levels of molecular flexibility that characterize medium to large size organic molecules. Table 2 summarizes the degree of flex-

Target	JOYCE description		
	Harmonic Torsions ( $\phi$ )	Flexible Torsions ( $\delta$ )	Intra-molecular LJ ( $\mathbf{r}^{soft}$ )
<b>I:</b> triperylene [3,3,3] propellane, <sup>44,45</sup>	yes	no	no
<b>II:</b> poly(2,7-pyrenylene) <sup>36,37</sup>	yes	yes	no
<b>III:</b> <i>n</i> -tetradecane <sup>46–48</sup>	no	yes	yes

Table 2: Summary of the different types of ICs (see also Figure 3) involved in the parameterization of the first three test cases.

ibility for the first three benchmark compounds, whose ICs are evidenced in Figure 3. In the following, a detailed discussion of the three resulting QMD-FFs and their quality will be given separately for each considered target.

#### **I: propellane**

The first target molecule, namely triperylene [3,3,3] propellane<sup>44,45</sup> (see Figure 4.a), is

characterized by a rather unusual shape and a rather stiff scaffold, constituted by three anthracene rings placed as wings and connected via three 5-atom rings. Such a peculiar structure makes it difficult to obtain accurate FF parameters from transferable databases. Nonetheless, for comparison purposes, we first assembled a standard GAFF potential<sup>79</sup> through the PrimaDO-RAC software<sup>69</sup> (see section **S5.1** in the supporting Information), which allows for the automatic generation of a GAFF topology file in GROMACS format. Successively, a QMD-FF was built by fitting QM data with the JOYCE3.0 protocol at PBE/6-31+G(d,p) level of theory as detailed in Table 1. Just by looking at target **I**'s structure, it is evident that large amplitude oscillations of the various rings are not possible because of the constrained arrangement of the bonds, whereas only small distortions around the equilibrium geometry should be expected in ambient conditions. For this reason, as reported in Table 2, besides the stretching and bending potentials defined in equations (6) and (7), only stiff harmonic torsions (8) were employed to build the QMD-FF. Harmonic torsion describe more appropriately the small amplitude oscillations distorting the aromatic rings. Conversely, the GAFF automated parameterization employs the standard Fourier series (9) to describe all the dihedrals. The intramolecular QMD-FF parameterized with JOYCE3.0 was validated by comparing selected features to both those obtained with the general purpose FF and at QM level, as summarized in Figure 4. First, propellane's QM and QMD-FF optimized structures are compared visually in Figure 4.a, whereas the vibrational

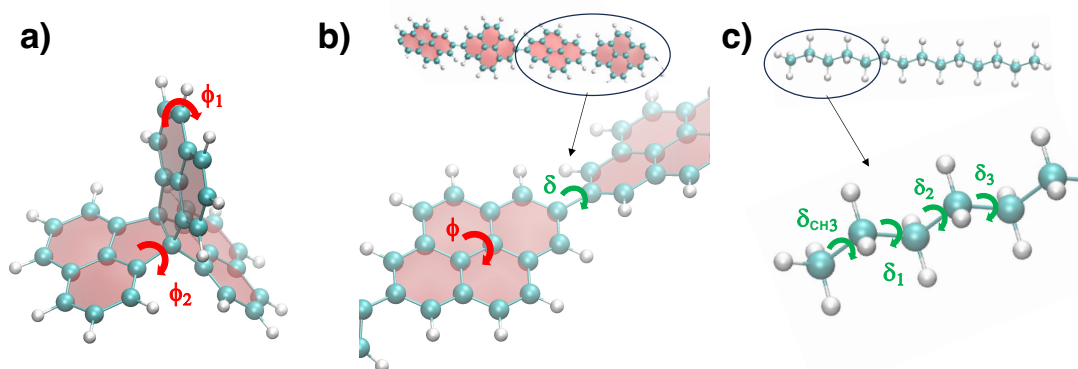


Figure 3: Selected targets to explore different kinds of dihedral flexibility: a) triperylene [3,3,3] propellane,<sup>44,45</sup> b) poly(2,7-pyrenylene)<sup>36,37</sup> and c) *n*-tetradecane.<sup>46–48</sup> In each panel and the insets, stiff ( $\phi$ ) and soft ( $\delta$ ) torsions are evidenced in red and green, respectively.

normal modes and related frequencies, again computed at both levels of theory, are compared in Figure S2. Both tests support the quality of our QMD-FF in reliably mimicking the equilibrium structure of the molecule and the distortions that are induced by small amplitude oscillations in a LHA. Notably, as reported in the top rows of Table 3, the RMSD of the MM optimized geometry with respect to the QM one is closer to zero when the JOYCE3.0 parameterization is used, indicating that GAFF parameters introduce a structural mismatch already at zero Kelvin.

To compare the performance of the two FFs at ambient temperature, we first carried out a BOMD as detailed in the Methods Section. Thereafter, the internal potential energy of several propellane conformations, extracted from the BOMD trajectory, was re-computed at MM level, either with the general purpose or the JOYCE3.0 FF. The energy distribution at the different level of theory is displayed in Figure 4.b, where it appears that QMD-FF nicely peaks close to

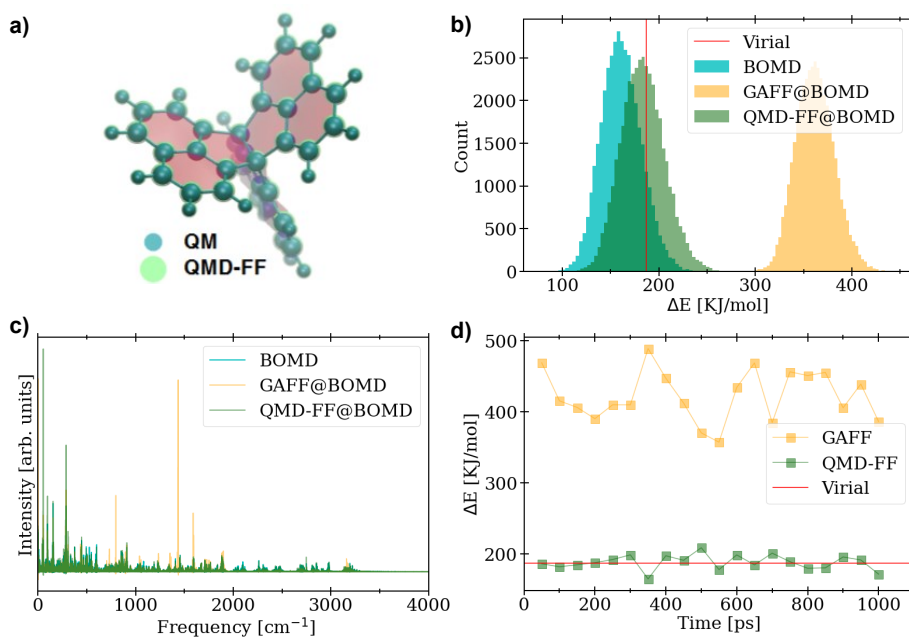


Figure 4: Validation of the JOYCE3.0 parameterized QMD-FF for target **I**. In all panels the same color palette is adopted: QM, GAFF<sup>69,79</sup> and QMD-FF results are displayed in cyan, orange and green, respectively. a) Overlap of the optimized structure at QM and QMD-FF level: b) internal energy distributions computed along the BOMD at QM or MM level, with either the general-purpose or the JOYCE3.0; c) spectral densities as obtained from the auto-correlation function of the internal energy fluctuations; d) internal energy computed by performing single point QM calculations on the structures obtained by MD runs carried out with the two FFs indicated. In panels c) and d), the internal energy expected from the virial theorem is reported as reference with a red line.

the BOMD reference, and both distributions spread near the value ( $\sim 200$  KJ/mol) expected by the equi-partition principle and the virial contribution. By contrast GAFF peak is found at remarkably higher energies. Although all distributions show seemingly similar widths, we report in Figure 4.c the cosine transform of the autocorrelation function of the fluctuations of the potential internal energy. Here, we see notable differences between the two FF models, where the largest discrepancies in the spectral densities are found again for the general purpose description, which, even in the case of this stiff molecule, is not completely able to capture the variation in potential energy connected to structural fluctuations.

This significant structural mismatch appears even more clearly from the last validation test illustrated in Figure 4.d and reported in Table 3. Unlike the previous test, the potential internal energy of compound **I** is now computed only at the QM level, yet using frames extracted from classical MD runs, performed in the gas phase with either GAFF<sup>69,79</sup> or the JOYCE3.0 FF. It is evident from Figure 4.d that while the QMD-FF generates geometries that are genuinely accessible under the imposed thermodynamic conditions, a general-purpose description produces excessively distorted geometries, yielding populations corresponding to much higher temperatures.

<b>0 K</b>	GAFF	JOYCE3.0
Bond lengths (Å)	0.003	0.000
Bending angles (degr)	1.212	0.004
Dihedrals (degr)	0.412	0.009
<b>298 K</b>	GAFF	JOYCE3.0
$\langle \Delta E_{QM}^{intra} \rangle$ (kJ/mol)	$422 \pm 35$	$187 \pm 10$
MUE (kJ/mol)	243	11
$D_{max}$ (kJ/mol)	308	28
$D_{min}$ (kJ/mol)	176	0.1

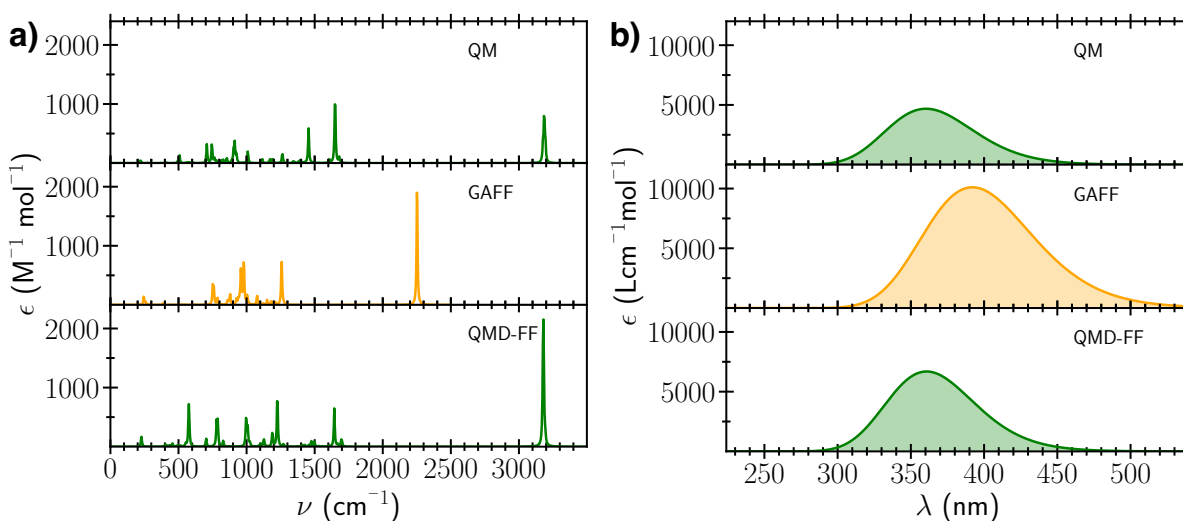
Table 3: Top (0 K): RMSD between QM and MM optimised geometries of compound **I** in terms of bond lengths, bending angles and dihedrals. Bottom (298 K): Average internal energies (kJ/mol) and statistical descriptors with respect to the expected internal energy at room temperature ( $\frac{k_B T}{2} (3N - 6) = 180$  kJ/mol): Maximum Unsigned Error (MUE), maximum and minimum deviation ( $D_{max}$  and  $D_{min}$ ).  $\langle \Delta E_{QM}^{intra} \rangle$  is obtained for triperylene [3,3,3] propellane at QM along the structures obtained with MD runs carried out at ambient conditions with either GAFF<sup>69,79</sup> or JOYCE3.0's QMD-FF.

## II: poly(2,7-pyrenylene)

With the aim of considering a progressive increase of the molecular flexibility, the second selected target is the 2,7-pyrenylene polymer, which is composed by stiff PAH units (pyrene), connected by rotatable bonds.<sup>36,37</sup> As detailed in Table 2, although the structure of each PAH unit can be safely handled through the harmonic term defined in equation (8) depending on the stiff dihedrals  $\phi$ , the rotation of two neighboring units around the C-C bond connecting them (see Figure 3.b) can undergo large amplitude oscillations and present several minima. To take such slow anharmonic motions into account, with respect to the previous test case, both the flexible dihedrals  $\delta$  and their contribution, accounted for by equation (9), are automatically added by JOYCE3.0 to the IC collection and to the set of employed model functions, respectively. A three unit chain of pyrenylene polymer has been recently<sup>36</sup> parameterized by some of us, following the JOYCE procedure. The resulting trimer QMD-FF was thereafter extended to a chain of 35 units, to perform MD simulations of the polymer on a gold surface, well reproducing the molecular adhesion/detaching forces as measured by cryogenic atomic force microscopy.<sup>36,37</sup> Given such successful validation, in the present work we build a QMD-FF (see Section S5.2 in the Supporting Information for the full list of parameters) for a four-unit polymer chain (target II) by directly transferring the JOYCE parameters previously obtained for the trimer, following the same procedure reported in Ref. [36]. Although the quality of the JOYCE tailored QMD-FF for poly(2,7-pyrenylene) chains of 35 and 10 units has been already assessed in previous work,<sup>36,37</sup> here we specifically validate the JOYCE3.0 procedure on target II by comparing the description of the molecular structure achieved with our QMD-FF, with that obtained with a standard one. This was done to further highlight the potential shortcomings of relying on a general-purpose description and the advantages offered by a specifically tailored approach.

Figure 5.a compares the the IR spectra obtained using the optimized geometries and Hessian matrices computed at QM (DFT) level and at MM level, through the two considered FFs: it is evident how the JOYCE3.0 parameterization yields a very good description of the poly-pyrene vibrational features with respect to QM. On the contrary, GAFF displaces many peaks in the medium and high frequency range, indicating a faulty representation of the aromatic scaffold dynamics. The effect of the structural mismatch introduced by the general purpose description





**Figure 5:** a) Tetra(2,7-pyrenylene) IR spectra computed from the isolated molecule's Hessian matrix and optimized geometry at full QM level (top, green line) or at MM level, either with the general purpose GAFF<sup>69,79</sup> (center, orange line) or the QMD-FF (bottom, green line); b) UV/VIS spectra computed for tetra(2,7-pyrenylene) at the geometry optimized at QM level (top, green line) or at MM level, using either GAFF<sup>69,79</sup> (center, orange line) or QMD-FF (bottom, green line).

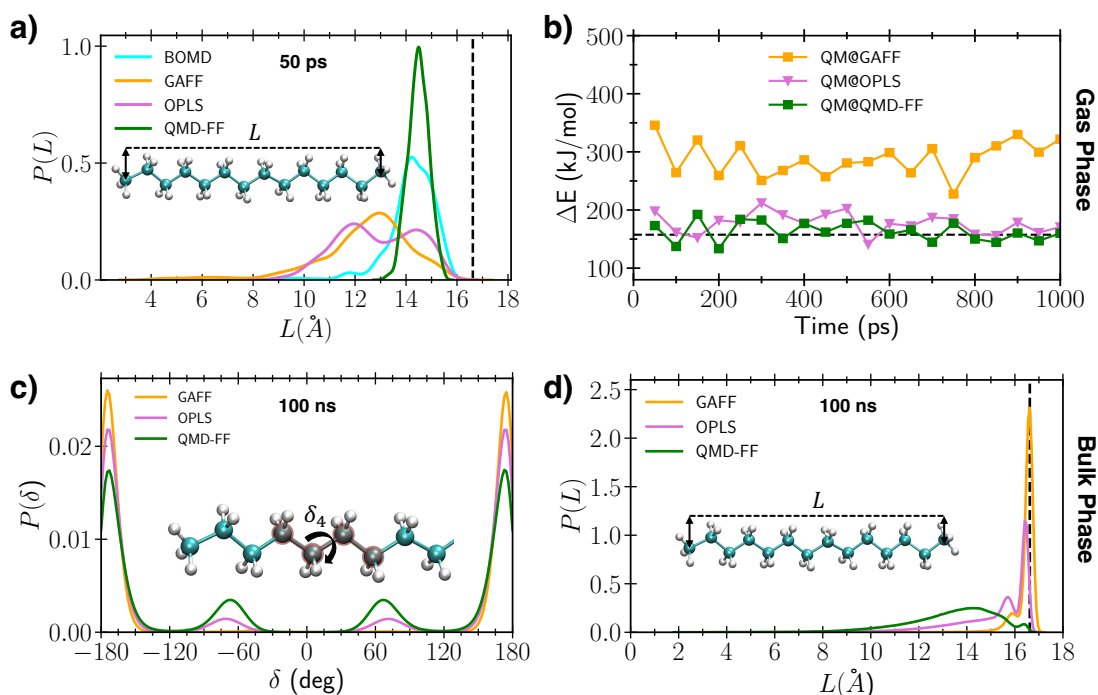
appear even more clearly in Figure 5.b, where the UV/Vis absorption spectra computed for the QM, GAFF and QMD-FF optimized geometries are displayed and compared. The lack of specificity in reproducing the bond length alternation (BLA) in the aromatic units when using GAFF results in a significantly red shifted spectrum. This confirms that, in addition to accurately reproducing properties governing molecular adhesion and detachment dynamics as shown in previous work,<sup>36,37</sup> a good reproduction of spectroscopic features also benefits from the tailored description provided by QMD-FF.

### III: *n*-tetradecane

The successful validation of the JOYCE3.0 procedure in handling both rigid molecules with “exotic” shapes and long, semi-flexible chains prompted us to complete these first three benchmarks with a highly flexible compound, i.e. target **III** (see Figure 3). *n*-tetradecane is indeed a long and very flexible linear alkane, characterized by a chain of fourteen Carbon atoms and several rotatable dihedrals, which play a crucial role in determining the overall conformation of the molecule.<sup>46–48</sup> Moreover, target **III** it undergo significant backbone distortion even under ambient conditions, resulting in head-to-tail or similar contacts during its conformational dy-

namics. Therefore, as for target molecule **II**, *n*-tetradecane was parameterized through the new JOYCE3.0 two-step procedure using both the QM Hessian matrix and the relaxed scans along the flexible dihedrals, yet introducing the use of internal Lennard-Jones (LJ) interactions, as described in equations (10) and (11). As detailed in Section **S5.3** of the Supporting Information, target **III** parameterization was carried out on an ensemble of 252 ICs, which comprises both stiff (bond stretching and angle bending) and soft (flexible dihedrals and 1-8 or further intramolecular interactions) ICs. This large set can be reduced to 19 independent force constants exploiting *n*-tetradecane symmetry (see also Figure **S5**), obtaining a final standard deviation of 0.3 kJ/mol, which confirms that the QMD-FF is able to accurately represent the QM PES both in a LHA and along the relaxed profiles of the scanned dihedrals (Figures **S6** and **S7**). Nonetheless, to achieve a more comprehensive validation, which also accounts for possible coupling among ICs during collective displacements, we also repeated some of the validation tests based on BOMD, previously discussed for target **I**.

Figure 6.a shows that, in contrast with general purpose FFs, the QMD-FF is capable to well reproduce the chain length distribution observed in the more accurate BOMD run. This good agreement between QMD-FF and QM data is confirmed by Figure 6.b: the QM internal potential energy computed along the *n*-tetradecane geometries extracted from QMD-FF trajectories again nicely oscillates around the value expected by theory at 300 K. Notably, as displayed in Figure **S8** of the Supporting Information, the chain length distribution obtained along this 1 ns run, substantially differs from the one obtained over 50 ps, either obtained at the QM level with BOMD or classically with the same QMD-FF. This finding confirms that while large amplitude and slow motions can occur at room temperature, inducing a significant distortion of the molecular shape, the long time scales required to sample such motions are clearly unreachable by applying BOMD techniques. Since the conformational dynamics is expected to be sensitive to the environment, the validation of any model for *n*-tetradecane should be thus extended to the condensed phase, hence requiring the use of the cost-effective CL-MD methods. Exploiting the partition of the total FF energy in inter- and intra-molecular terms, the JOYCE3.0 parameters can be straightforwardly merged with an inter-molecular set, for instance transferred from some tested database for alkane molecules. Here the QMD-FF intramolecular description



**Figure 6:** Validation of the JOYCE3.0 parameterized QMD-FF for target **III**. a) Distributions of the chain elongation ( $L$ ) computed in the gas phase for 50 ps along the BOMD at QM level (cyan line) or MM level, with either a general-purpose (GAFF,<sup>69,79</sup> orange or OPLS,<sup>68,80</sup> magenta) or the JOYCE3.0 (QMD-FF, green) FF. The maximum elongation expected by the all-*anti* conformer is also displayed for reference with a dashed black vertical line. b) Internal potential energy computed at QM level along the structures obtained by MD runs carried out in gas phase with either a general-purpose (GAFF,<sup>69,79</sup> orange or OPLS,<sup>68,80</sup> brown) or the JOYCE3.0 (green, QMD-FF) FF. The internal energy expected from the virial theorem is again reported as reference with a horizontal black line. c) Population distribution of the  $\delta_4$  dihedral (see Figure S5.a in the Supporting Information) achieved in condensed phase simulations with the different FFs. d) Distribution of the chain elongation  $L$  computed in condensed phase simulations with the different FFs.

was merged with the OPLS intermolecular parameters, allowing for simulation of systems with large numbers of interacting tetradecane polymers. For a final evaluation of the resulting FF, as reported in Table 4, selected thermodynamic properties, averaged for 100 ns at 1 atm and 298 K on a system of 1000 *n*-tetradecane molecules in of condensed phase, were compared with those obtained with the same protocol, yet using two transferable FFs, namely GAFF<sup>69,79</sup> and OPLS.<sup>68,80</sup>

The performances delivered by the three FFs are quite different: while OPLS and our QMD-FF (which share the same intermolecular term) yield an overall good agreement with the experimental measures, GAFF simulates a too dense phase, with a significant overestimation of the

FF	$\rho$ (kg/m <sup>3</sup> )	$\Delta H_{vap}$ (kJ/mol)	$c_p$ (J mol <sup>-1</sup> K)	$\eta$ (mPa s)
GAFF <sup>69,79</sup>	841.1 $\pm$ 5.2	125.6 $\pm$ 0.5	-	-
OPLS <sup>68,80</sup>	771.1 $\pm$ 2.2	80.4 $\pm$ 0.4	472 $\pm$ 2	1.3 $\pm$ 0.5
QMD-FF	769.3 $\pm$ 2.2	78.7 $\pm$ 0.4	460 $\pm$ 1	1.1 $\pm$ 0.1
Experiments	759 <sup>48</sup>	71.3 <sup>47</sup>	438.28 <sup>46</sup>	2.33 <sup>48</sup>

Table 4: Density ( $\rho$ ), vaporization enthalpy ( $\Delta H_{vap}$ ), specific heat at constant pressure ( $c_p$ ) and viscosity ( $\eta$ ) computed using the three FFs and compared to experiments. All values refer to 298 K and 1 atm.

vaporization enthalpy, suggesting too strong intermolecular interactions. Indeed, as revealed from Figure S9 in the Supporting Information, evident crystalline sub-domains start forming already at 298 K, confirming the GAFF deficiency in well representing long alkyl chains.<sup>81–83</sup> In particular, as displayed in the c) and d) panels of Figure 6, the overestimated interaction in GAFF alters the molecular conformation artificially favoring the all-*anti* conformers, hence resulting in too elongated molecules. As shown in more detail in Figures S10 and S11 in the Supporting Information, a conformational behavior more in line with the other two descriptions can be gained with GAFF by raising the temperature, suggesting that the latter FF could significantly overestimate the melting temperature. Such an incorrect representation of the *n*-

	all- <i>ati</i>	Single- <i>gauche</i>	Double- <i>gauche</i>	Triple- <i>gauche</i>	Other
GAFF (%)	84.3	12.7	1.9	0.5	0.6
OPLS (%)	32.6	11.7	28.1	18.7	8.9
QMD-FF (%)	4.4	15.7	26.3	26.4	27.2

Table 5: Comparison of dihedral distributions of *n*-tetradecane across the different force fields.

tetradecane conformational dynamics is further confirmed in Table 5, where the distribution of the chain conformers, averaged along the 100 ns dynamics were collected for the three models, from which interesting information can be retrieved also for the comparison between OPLS and the QMD-FF. In fact, despite they share the same inter-molecular term, the two FFs do not deliver identical thermodynamic properties, the QMD-FF being closer to the experiment in all descriptors, except viscosity. The increased accuracy can be traced back only to the intra-molecular re-parametrization, which allows the alkyl chains to follow a different (see Table

5) and more realistic conformational dynamics, leading to a better estimate of the molecular excluded volume and, in turn, of the intermolecular interaction energy.

### 3.2 IV: Specific Tuning of Intra-molecular Non-bonded Interactions

In the previous sections, we presented three examples demonstrating how JOYCE3.0 can address increasing levels of molecular flexibility using either harmonic functions or periodic series, while the contribution of the non-bonded intramolecular term (10) remained minimal. In fact, the only function of non-bonded interaction was preventing bad contacts between atoms laying far apart in the molecular backbone (e.g. head-to-tail or similar interactions within the alkyl chains of target **III**). There are, however, many cases in which a high degree of internal flexibility allows the molecule to explore a complex conformational space, often characterized by several different local minima, featuring strong and specific interactions, possibly settled between non-bonded, yet neighboring atoms. This is the case for instance of internal HBs,  $\pi$ - $\pi$  stacking or H- $\pi$  patterns. To tackle such kind of highly flexible target molecules, JOYCE 3.0 comes with an improved implementation for the parameterization of non-bonded parameters (**Route II**, see Methods sections), which allows the intra-molecular LJ interactions to be introduced between a carefully selected set of atoms.

Here, to demonstrate the effectiveness of the new approach, we consider target **IV**, the monomeric repeating unit of a well-known polymer, namely polyethylene terephthalate (PET), which consists of a benzene ring covalently bonded to a carboxyl group and an ethylene glycol moiety (see Figures 7.a and **S12** in the Supporting Information). As shown in the b) to d) panels of Figure 7, when compared to the QM reference (see Table 1) an accurate representation of the monomeric unit conformational features is not achievable by simply transferring parameters from general-purpose FFs (see also Figures **S13** and **S14** in the Supporting Information). Yet, a reliable description of the internal flexibility is crucial to PET's properties and functions.<sup>49,50</sup> It is thus evident that, in view of extending the model to longer polymeric chains, the intramolecular FF describing each unit should be accurately refined.

The first step of the new JOYCE3.0 parameterization procedure is to choose a number of potentially interacting non-bonded atoms, and assign a set of specific LJ parameters (namely

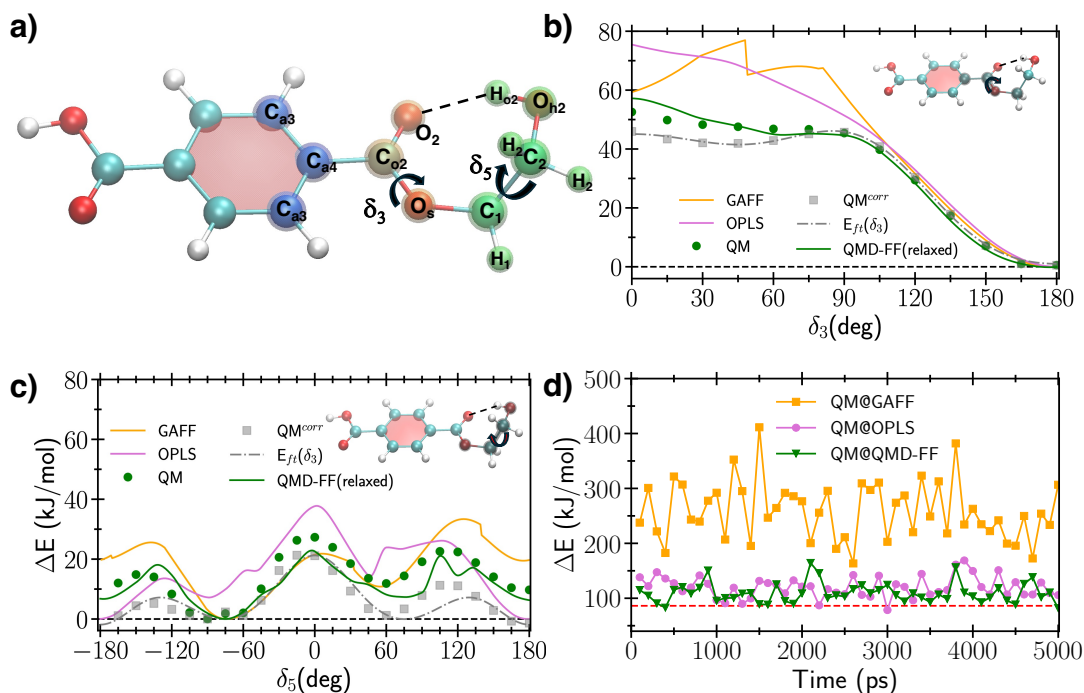


Figure 7: (a) QMD-FF atom types for which intramolecular LJ interactions (10) were included are highlighted with transparent spheres (blue for ring atoms, orange for carboxyl atoms, and green for the ethylene/hydroxyl chain). The dashed line displays the internal HB and the dihedrals scanned in panels b) and c) are also indicated; b) Relaxed torsional scans of dihedral  $\delta_3$  carried out either with transferable FFs (GAFF,<sup>69,79</sup> orange line, and OPLS,<sup>68,80</sup> magenta line), QMD-FF (green line) or at QM level (green circles). The QM energies corrected as in equation (19) for the LJ contributions and the potential terms for flexible torsion (9) are also displayed with gray squared and brown dashed line; c) relaxed torsional scans of dihedral  $\delta_5$ , lines symbols and colors are the same as in the previous panel; d) Conformational energies evaluated at QM level on 1 ns gas phase MD run, carried out with GAFF (orange), OPLS (magenta) or QMD-FF (green).

$\epsilon^{intra}$  and  $\sigma^{intra}$ ) for each of them. By looking at target **IV**'s molecular structure, displayed in Figure 7.a, two different kinds of interactions can be settled along its conformational dynamics: *i*) internal HBs (between the ethylene glycol moiety and the carboxyl group on the right side of the ring, and within the carboxyl group on the left); *ii*) steric repulsion between the ethylene-glycol atoms and the phenyl ring. As shown in Figure 7.a and detailed in Figure **S15** in the SI, both these interactions can be accounted for by introducing LJ parameters for few target atoms, namely the three phenyl carbons on the ethylene substituted side ( $C_{a3}$  and  $C_{a4}$  atom types) and their hydrogens ( $H_{a3}$ ), the left-side carboxyl group ( $C_{o2}$ ,  $O_2$  and  $O_5$ ) and all atoms of the flexible chain. Table 6 reports all non-bonded LJ parameters, whose values were assigned as illustrated in the Section S5.4 of the Supporting Information, together with additional details on

the parameterization procedure. The latter was carried out for a total of 160 internal coordinates,

phenyl–chain repulsion			carboxyl–chain repulsion		
Pair ( <i>ij</i> )	$\sigma_{ij}^{intra}$	$\epsilon_{ij}^{intra}$	Pair ( <i>ij</i> )	$\sigma_{ij}^{intra}$	$\epsilon_{ij}^{intra}$
C <sub>a3</sub> -C <sub>1</sub>	3.11	0.1	C <sub>o2</sub> -H <sub>o2</sub>	2.57	0.1
C <sub>a3</sub> -C <sub>2</sub>	3.52	0.1	C <sub>o2</sub> -C <sub>2</sub>	2.83	0.1
C <sub>a4</sub> -C <sub>2</sub>	3.52	0.1	C <sub>o2</sub> -C <sub>2</sub>	2.79	0.1
H <sub>a3</sub> -C <sub>1</sub>	2.80	0.1	O <sub>2</sub> -O <sub>h2</sub>	2.72	0.2
intra-molecular HB					
Pair ( <i>ij</i> )	$\sigma_{ij}^{intra}$	$\epsilon_{ij}^{intra}$	Pair ( <i>ij</i> )	$\sigma_{ij}^{intra}$	$\epsilon_{ij}^{intra}$
O <sub>2</sub> -H <sub>o2</sub>	1.75	10.0	O <sub>S</sub> -H <sub>o2</sub>	1.75	10.0

Table 6: QMD-FF parameters for non-bonded intramolecular interactions assigned to target **IV** atoms displayed in Figure 7.a;  $\sigma^{intra}$  are in Å and  $\epsilon^{intra}$  in kJ/mol.

parameterizing 75 independent parameters with a total standard deviation of 0.85 kJ/mol. A full list of the QMD-FF parameters is given in Tables **S40** to **S45**, together with a collection of Figures showing in some detail the successful validation tests (see also Table 7), whereas in the following the rest of the discussion focuses on the non-bonded interaction effect and on the comparison with general purpose FFs.

Panel b) and c) of Figure 7 display in some detail the effect of the inclusion of specific LJ interactions and how they are handled within the new **Route II**. In fact, by comparing the original QM profiles with those corrected for non-bonded interactions according to equation (11), the significant role of these interactions in reshaping the curves becomes evident. Moreover, as expected, the effect of the specific interactions is generally repulsive, as the corrected profiles consistently appear at lower energies compared to the original values when non-bonded terms are considered. Finally, it is worth noticing how the corrected curves are well reproduced (gray symbols and brown dashed lined in both panels) by the flexible torsional valence term (9), while the original QM profiles is recovered when the whole QMD-FF is relaxed (green symbols and lines, respectively). Turning to the comparison of the results obtained with those delivered from the two general purpose FFs, Figures 7, **S13**, **S14** and **S20** show how the QMD-FF outperforms the one computed for both transferable FFs, when they are compared with QM data. Figure 7 clearly shows that GAFF and OPLS miss or displace some minima, and provide generally overestimated barriers, while this in turn reflects in internal energy landscapes significantly dif-

ferent from those obtained at QM level (see for instance Figure S13). As a final validation, Figure 7.d shows the results of the same test conducted for the previous molecules, namely verifying that the molecular conformations sampled by the target molecule during gas-phase dynamics are indeed accessible under the given thermodynamic conditions. This is determined by checking if their QM internal energy falls within the range predicted by the virial theorem. The results, also summarized in Table 7, confirm the excellent performance of the QMD-FF, which notably minimizes the structural mismatch with respect to the QM data, compared to the tested general-purpose models.

### 3.3 V: Parameterizations with post-HF methods

Target V (rPSB11) is a natural chromophore that is covalently bound to opsins through a Schiff base linkage. It is extensively studied due to its role initiating the vision process upon light absorption in animals.<sup>51,52</sup> For all the four previous targets, the reference QM database was obtained exploiting DFT methods. However, to investigate the photoisomerisation of this chromophore, which takes place at the 11-*cis* double bond through a conical intersection, multiconfigurational strategies as CASSCF/CASPT2 methods are typically used on structures obtained from classical MD simulations.<sup>52,84,85</sup> Since the BLA and flexible torsions are known to gov-

<b>0 K</b>	GAFF	OPLS	JOYCE3.0
Bond lengths (Å)	0.024	0.014	0.000
Bending angles (degr)	4.8	2.6	0.1
Dihedrals (degr)	7.3	1.9	10.1
<b>298 K</b>	GAFF	OPLS	JOYCE3.0
$\langle \Delta E_{QM}^{intra} \rangle$ (kJ/mol)	$259 \pm 52$	$120 \pm 19$	$109 \pm 18$
MUE (kJ/mol)	173	34	23
$D_{max}$ (kJ/mol)	325	83	78
$D_{min}$ (kJ/mol)	77	1	2

Table 7: Top (0 K): RMSD between QM and MM optimised geometries of compound IV in terms of bond lengths, bending angles and dihedrals. Bottom (298 K): Average internal energies (kJ/mol) and statistical descriptors with respect to the expected internal energy at room temperature ( $\frac{k_B T}{2}(3N - 6) = 86.3$  kJ/mol): Maximum Unsigned Error (MUE), maximum and minimum deviation ( $D_{max}$  and  $D_{min}$ ).  $\langle \Delta E_{QM}^{intra} \rangle$  is obtained for ethylene terephthalate at QM along the structures obtained with MD runs carried out at ambient conditions with either general-purpose FFs (GAFF<sup>69,79</sup> and OPLS<sup>68,80</sup>) or JOYCE3.0's QMD-FF.



ern such photoisomerisations,<sup>86</sup> it is in fact crucial that the adopted FF reliably describes both these properties and their dynamics. Notably, DFT does not reproduce correct BLA, and post-HF calculations are therefore required to investigate photoisomerization.<sup>87</sup> For this reason, we carry out the JOYCE3.0 parameterisation of target **V** using MP2. The conjugated polyene scaffold present in the chromophore, responsible for light absorption initiating the photochemical process of typical interest, shows that at MP2 level a delocalised electronic structure, where single bonds are stronger than those in alkanes, and double bonds weaker than those in olefins. This fact not only influences equilibrium distances and force constants, but also torsional profiles, suggesting that double bonds in such a molecule should not be described by the harmonic potential terms, but rather by Fourier series, which enables a better description of the expected large torsional displacements. For these reasons, in addition to the equilibrium geometry and Hessian matrix, the QM database includes ten torsional relaxed scans (see also Section **S5.6** in the Supporting Information), divided equally between rotations about single bonds and double bonds ( $\delta_n$  and  $\phi_n$ , respectively, with  $n = 1 \dots 5$ ).

The JOYCE3.0 parameterization was carried out on a total of 282 internal coordinates and 142 independent parameters, yielding a standard deviation of 0.15 kJ/mol. A preliminary validation of the resulting QMD-FF is displayed in panel a) and b) of Figure 8. Indeed, from the overlap of the QM and QMD-FF optimised structures in Figure 8.a and from the values reported in Table 8 it appears that the QMD-FF reproduces the QM equilibrium geometry with negligible error on bond lengths and angles. The FF vibrational frequencies shown in Figure 8.b also well correlate with their QM counterparts, confirming that the chromophore is well described in a LHA. Turning to the expected large amplitude conformational dynamics, to further validate our FF and better ascertain the gain in accuracy with respect to general purpose descriptions, the QMD-FF obtained from JOYCE 3.0 is compared to a GAFF model.<sup>67</sup> First, as detailed in Table 8, we highlight the worse performance of the latter, especially in terms of dihedrals and BLA of the polyene system, which anticipates a less accurate description of the photophysics and photochemistry that one would study on geometries extracted from a CL trajectory. Concerning molecular flexibility, it is evident from the selected MM relaxed scans shown in Figure 8.c that a general purpose model provides lower  $\delta$  barriers with respect to JOYCE3.0 (see also Figures

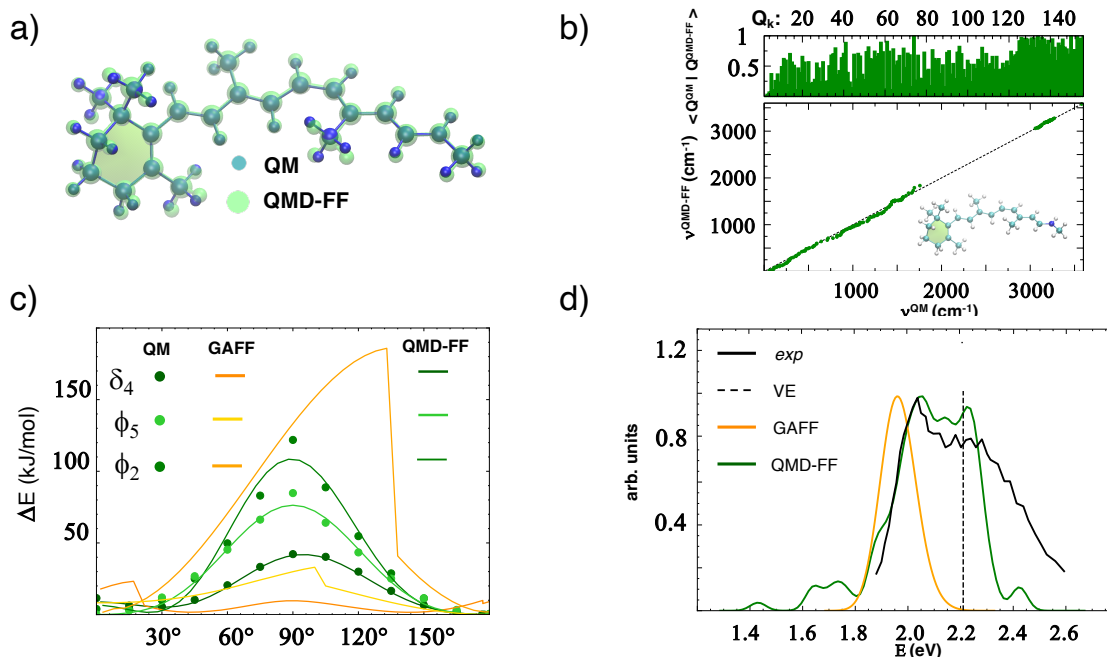


Figure 8: a) Overlap between QM (MP2) and QMD-FF optimized structures for target **VI**; b) validation of the JOYCE3.0 parameterization in terms of normal modes overlap (top) and vibrational frequencies (bottom); c) selected torsional profiles for GAFF (reddish dashed lines) and QMD-FF (greenish lines) with respect to MP2 profiles (green circles); d) computed (RMS-CASPT2/cc-pvDz) and experimental (digitalised from Ref. [88], black solid line) spectra, obtained for rPSB11 *in vacuo*. All spectra have been normalised with respect to their maximum intensity. Simulated spectra have been obtained either at a static level on sole optimized geometry (VE, vertical dashed line), or through the CEA-VE approach (see the Methods Section and **S4** in the Supporting Information), on both GAFF (orange) or QMD-FF (green) trajectories.

**S22–S26** in the Supporting Information), stemming from single bonds which are comparable to the one found in regular alkanes, yet weaker than what expected by QM theory. The opposite is true for double bonds, where GAFF's higher barriers indicate too stiff coordinates, comparable to standard double bonds in olefins. This results in an overestimated BLA, in contrast with expectations based on electronic nature of the polyene system, which are instead correctly described at QM level and, therefore, in the QMD-FF.

The availability of experimental UV-vis spectra *in vacuo* for this chromophore<sup>88</sup> makes it possible to further validate the two descriptions. To this end, absorption spectra were computed through the CEA-VE procedure (see Section **S4** in the Supporting Information for details), exploiting 100 geometries extracted from MD trajectories carried out with the two FFs for 20

FF	Bond lengths / Å	Bending angles / deg	Dihedrals / deg	BLA / Å
GAFF <sup>67</sup>	0.018	1.7	18.2	0.100
JOYCE	0.002	0.2	4.6	0.056

Table 8: MM Bond Length Alternation (BLA) and RMSD between QM and MM optimised geometries for rPSB11 in terms of bond lengths, bending angles, dihedrals.

ns. For each sampled snapshot, we used the OpenMolcas software<sup>60</sup> to compute, at the cc-pVDZ level, RMS-CASPT2(12,12) vertical transition energies for the first 3 states and SA(3)-CASSCF(12,12)  $S_0 \rightarrow S_1$  oscillator strengths, eventually retrieving the final UV-vis spectrum as a gaussian convolution of all data. The comparisons shown in Figure 8.d demonstrate that the JOYCE 3.0 description provides a much better agreement with experimental data. The position of the spectrum, attributable to the overall description of molecular geometry, is on top of experimental data, whereas GAFF is visibly more shifted. Moreover, JOYCE3.0 spectral width is also much closer to experiments, whereas, as expected from its too rigid structure, the general purpose description provides a significantly narrower spectrum. Based on these findings, the accuracy of the proposed QMD-FF appears as an appropriate starting point for subsequent studies of excited state dynamics.

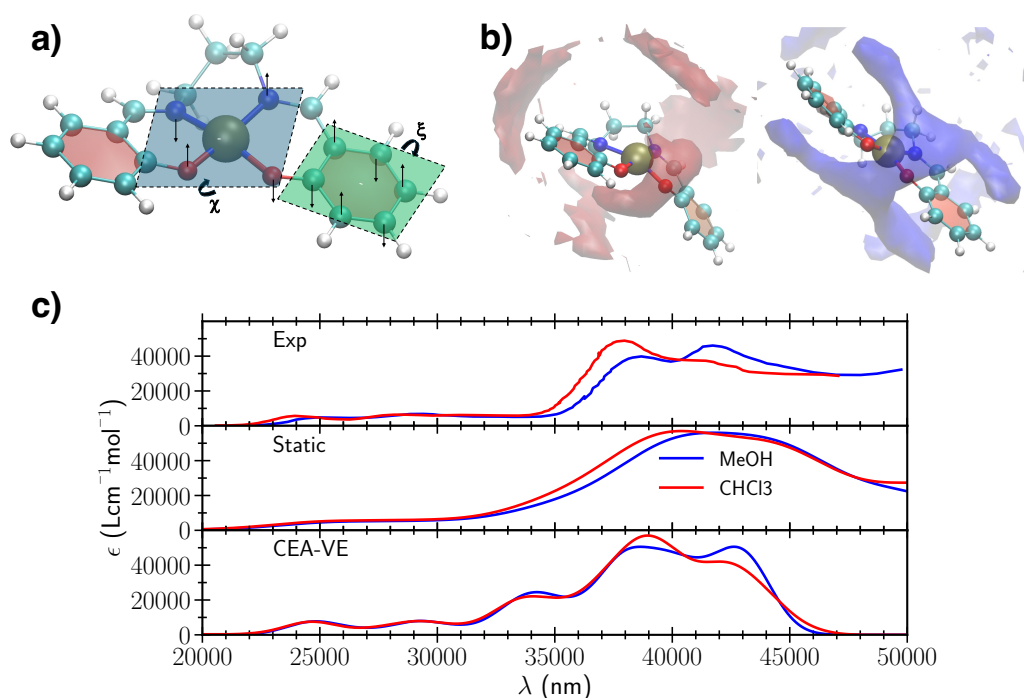
### 3.4 VI: FF parameterizations of transition metal complexes for computational spectroscopy

Transition metals complexed by organic ligands are important for energy applications and technologies, due to their ability to facilitate efficient electron transfer and catalytic processes both in the ground and excited states. Their versatile electronic structure makes them useful in key processes such as solar energy conversion, battery storage, and fuel cell reactions.<sup>11</sup> Often, various excited-state processes and photophysical phenomena occurring in these systems are influenced by interactions with the surrounding environment. These dynamic effects and non-equilibrium processes occur across multiple time scales, ranging from femtoseconds to nanoseconds.<sup>89</sup> To accurately capture these phenomena and the specific and mutual interaction between the TMC and the environment, it is essential to consider large systems to properly account for outer solvation shells. Reaching such extended size and time scales benefits from the use of classical force

fields (FF). However, there is a significant challenge: no universally accurate or transferable FF parameters currently exist for reliably describing the metal-ligand interactions or maintaining the correct coordination geometry in TMCs, especially in their electronic excited states. In response to this limitation, here we show how a system-specific and accurate parametrization is possible using the JOYCE3.0 protocol and offers a useful tool to describe the optical properties of the squared planar [Ni(salpn)] complex (compound **VI**).<sup>53</sup> As discussed for targets **II** and **V**, a correct prediction of the position and shape of the absorption band to electronically excited states heavily depends on the accuracy of the molecular geometries delivered by the FF. We have shown that the specific description yielded by JOYCE3.0 outperforms the one obtained by transferring general purpose-parameters that are affected by structural mismatch issues. For target **VI**, the comparison with transferable FFs is not even possible, because several parameters, as for instance those describing the coordination around the metal are missing in commonly available databases. For this reason, here we chose instead to validate our protocol by a direct comparison between computed and recently measured<sup>53</sup> experimental absorption spectra, in which target **VI** revealed a clear solvatochromism.

The JOYCE3.0 parameterization was carried out following the familiar assignment of specific atom types (see Figure **S27** for a detailed list) on a system of 206 internal coordinates. As shown in Figure 9.a, these included specific stiff dihedrals ( $\chi$  and  $\xi$ ), described through equation (8) and specifically defined for both Ni coordination atoms (two Oxygen and two Nitrogen), and the lateral aromatic rings. The resulting system of 61 independent parameters, was optimized with a final standard deviation of 0.1 kJ/mol. Given the fairly stiff scaffold and the absence of low vibrational modes, the parameterization was carried out in a single step, based on the optimized QM geometry and Hessian matrix only. Figure **S28** shows the standard preliminary validations tests, where it appears that both the optimized geometry and the vibrational features described by the QMD-FF are in good agreement with the reference QM data.

To further validate the QMD-FF accuracy and to show JOYCE3.0 capabilities in determining spectroscopic properties in different environments, we joined the parameterized intramolecular term with a set of intermolecular parameters (see Section **S5.7** in the Supporting Information for a full list) and carried out MD simulations in condensed phase. In contrast with *n*-tetradecane,



**Figure 9:** a) Optimized geometry of the squared planar [Ni(salpn)] complex (target compound **VI**) and definition of the stiff dihedrals  $\chi$  and  $\xi$ , ruling respectively the squared planar structure around the metal and the planarity of the aromatic rings; b) proton spatial density function in MeOH (left, red) or CHCl<sub>3</sub> (right, blue), c) comparison between experimental UV/vis absorption spectra (top) of compound **I** in methanol (red lines) and chloroform (blue lines)<sup>53</sup> and those computed in the present work through either a static approach (center) or exploiting MD simulations carried out with the QMD-FF (bottom)

where the bulk phase is considered, here the target compound is instead solvated into two different solvent molecules, namely methanol and chloroform. The analysis of the different solvation structure established after equilibration of the two solutions at ambient conditions was carried out both in terms of radial distribution functions ( $g(r)$ , see Figures **S29** to **S31** in the Supporting Information) and the spatial distribution functions displayed in Figures 9.b and **S32**. All data agree in indicating significant differences in the solvation shell settled by the two solvents: around the metal, it is much more structured in CHCl<sub>3</sub>, while a rather stable HB between the solvent protons and target **VI**'s oxygen atom is settled in both solvents, yet at a shorter distance in methanol. These local differences in the solvation patterns reflect in turn on the spectral features. Figure 9.d displays the experimental absorption spectra and compares them with the computed ones obtained with two different procedures, already employed for target **II** and **V**,

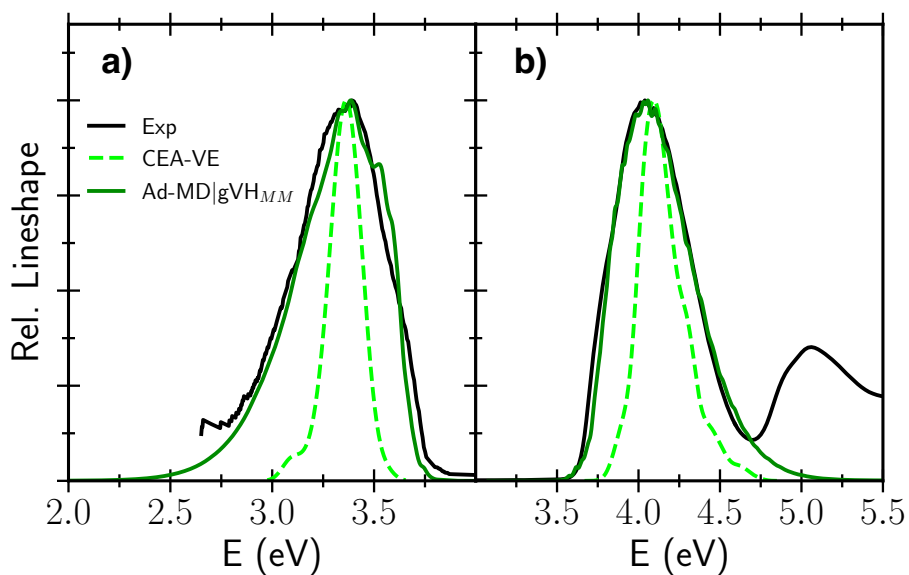
respectively: the "static" approach and the CEA-VE protocol. It is important to stress that in the "static" approach the bandshape is totally phenomenological, since it is obtained by applying a Gaussian broadening function, with an half width at half maximum (HWHM) of 0.33 eV, to the single stick transition, while the solvent is taken into account as a continuum, through the Polarizable Continuum Method (PCM).<sup>90</sup> Conversely, at the CEA-VE level, both the discrete nature of the solute-solvent interactions and the JOYCE3.0 driven conformational dynamics of the solute are taken into account. This results in a significantly smaller broadening to add to each VE. The approximations introduced in the static approach are evident in Figure 9.d, where the two solvents are almost indistinguishable, and the band structure at higher energies completely lost. On the other hand, by taking the system dynamics into account, the CEA-VE spectra show a more defined structure, in better agreements with the experimental signal. In particular, the local and specific solute-solvent interactions obtained by explicitly including in the VE calculations of the first neighbor shell of solvent molecules reveal the difference between the two solvents. As a final note, it might be worth noticing that the solvent molecules were included in the CEA-VE calculations at full QM level (in a QM/MM-like fashion), hence re-introducing in the spectra the effect of both specific interactions as HBs and mutual polarization effects.

### 3.5 VII: QMD-FF for excited states

One key feature of the JOYCE protocol is that FFs for excited states (ESs) can be straightforwardly derived on the same footing as those in the ground state (GS), provided a reliable QM dataset is computed for the ES. In this way, it is possible to simulate the steady-state fluorescence CEA-VE lineshapes<sup>20</sup> from the ensemble of geometries extracted from the MD simulation carried out in the ES, similarly as we have shown for absorption of the systems above. Moreover, the possibility of building consistent FFs for GS and ES paves the way to detailed descriptions of excited state dynamics.<sup>38,41</sup> In this section, we resort to the parameterization of the GS and ES FFs both at the ground and the first excited state of dithiophene, already reported in a previous work by some of us,<sup>55</sup> and here re-performed with the JOYCE3.0 code. Furthermore, in the case of the ES, we considered two different minima (*cis* and *trans* configurations around the central bond), each with a dedicated FF. As shown in the original publication, the usual analyses

on the quality of the FF reveal a similar accuracy for both GS and ES, compared to the reference data at DFT and TDDFT levels, respectively, confirming that the method is able to provide consistent parameterizations at both states. The CEA-VE calculation of the emission spectra, averaged for both *cis* and *trans* configurations, along with the one for absorption, are shown in Figure 10 (green dashed lines). Even applying a phenomenological Gaussian broadening with  $\text{HWHM} = 0.05$  to each transition, the comparison with the experiments reveals a considerable underestimation of the overall width for absorption and emission. This is, however, an expected effect,<sup>55</sup> as the conformation ensemble arising from a classical sampling generally underestimates the spectral width, especially when the role of progressions along high-frequency modes is significant. One way to address a proper simulation of the spectra, which considers the quantum nature of the high-frequency modes, while retaining an extended CL sampling for both the soft modes of the solute and those of the environment, consists in adopting a MQC strategy, as the Ad-MD|gVH method, recently proposed by some of us.<sup>55</sup> As discussed in the Methods Section, in this strategy the flexible degrees of freedom of the solute (specifically the torsion between the two thiophene rings) and the solvent are treated classically, while the stiff modes of the solute are treated at QM level. The electronic structure calculations carried out at each configuration extracted along the MD trajectory are usually performed at DFT and TDDFT levels for both states, which renders the Ad-MD|gVH protocol computationally demanding.

To circumvent this issue, here we exploit the availability of accurate and reliable QMD-FFs, consistently parameterized against the same level of QM theory at both GS and ES, to enable a more computationally effective calculation of such vibronic models. This is achieved by evaluating the gradient and Hessian matrices required by the VH model to build the quadratic PESs for both states using instead of TDDFT, the QMD-FFs derived with JOYCE3.0. In practice, we use GROMACS to evaluate gradients and Hessians with the QMD-FF. Concretely, the gradient is computed analytically at each snapshot, directly in the simulation box, including all solvent (ethanol) molecules and using PBC. In turn, the Hessian elements for all solute atoms are computed numerically from such analytical gradients, using a step of  $10^{-3}$  Å. The transition properties, i.e., the vertical energy and the transition dipole moment, are evaluated using the TDDFT method. In Figure 10, we present absorption and emission spectra computed



**Figure 10:** Experimental (black lines) and simulated (green) spectra. The latter are obtained with the CEA-VE (dashed lines) or the Ad-MD|gVH (solid lines) approaches, using QMD-FF potentials to evaluate gradients and Hessians. a) emission spectra; b) absorption spectra. Note that for a better comparison with the experimental line shapes, as in the original paper,<sup>55</sup> emission and absorption computed signals were shifted respectively by 0.12 and 0.03 eV. For the latter, the second band observed in the experiment is not reproduced by our calculations because only transitions to the lowest excited state were considered.

through such new implementation of the original Ad-MD|gVH procedure,<sup>55</sup> hereafter referred as Ad-MD|gVH<sub>MM</sub>. It is here important to recall that in the original publication Hessian and gradients were evaluated with computationally expensive QM/MM models, which yet lead to Ad-MD|gVH absorption and emission spectra in very good agreement with the experiment. By comparing the original spectra<sup>55</sup> with those here generated through the new procedure (see also Figure S34 in the Supporting Information), Ad-MD|gVH<sub>MM</sub> ones exhibit a similar broadening and well align with experimental data, yet at a significantly reduced computational cost.<sup>91</sup> Notably, one key difference between the two procedures is that, after projecting out the torsion, no imaginary frequencies were observed in Ad-MD|gVH<sub>MM</sub> at any snapshot. In contrast, the original Ad-MD|gVH protocol produced a significant number of imaginary frequencies,<sup>55</sup> requiring the projection of the pyramidalization coordinate to reduce the number of snapshots with such frequencies. This difference indicates mismatches between the potential energy surface computed at QM (TDDFT) level and that delivered by the QMD-FF. Such differences may arise



from the limited coupling between internal coordinates within the FF, which is connected to the simple functional forms adopted. Interestingly, another source for such a difference is the inherently diabatic nature of our FFs, which neglect potential coupling, especially between excited states, that could lead to additional imaginary frequencies. While these couplings might need to be reintroduced in subsequent calculations,<sup>41</sup> the diabatic FF approach offers a clean zero-order description that helps maintain control over the effects incorporated into the spectral simulation. Moreover, a key advantage of using QMD-FFs is that they allow for a very fast evaluation of gradients and Hessians, significantly facilitating the application of the Ad-MD|gVH<sub>MM</sub> method to more complex systems.

## 4 Conclusions

In this work, we have presented JOYCE3.0, a software package designed for the automated generation of specific and accurate intramolecular force fields suitable for classical simulations, reliably derived from reference QM descriptions. The most relevant features of the new JOYCE3.0 parameterization protocol have been summarized and discussed through seven distinct applications, each characterized by a remarkably diverse molecular structure and chemical properties. In addition to the successful validation of each parameterization against a target QM description, the quality of the QMD-FFs generated by the JOYCE3.0 protocol was evaluated by comparing their outcomes with both general-purpose models and experimental measurements. The results showed consistently good agreement with experimental data while outperforming in all cases general-purpose models.

Concretely, the ability of JOYCE3.0 to handle molecular scaffolds with increasing flexibility through the automated identification of stiff and soft modes was successfully demonstrated with targets **I** to **III**. In all cases, and leveraging the enhanced specificity provided by the new atom type definition, the QMD-FFs accurately captured the target conformational dynamics, both within a local harmonic approximation and across larger displacements. Even when more complex potential energy surfaces are considered - such as those modulated by strong and specific nonbonded intramolecular interactions, as in target **IV**, or strongly influenced by fluctuations in electronic density, as in target **V** - the robustness of our protocol enabled reliable descrip-

tions. The excellent accuracy observed for the first four benchmarks in reproducing realistic conformations prompted us to test JOYCE3.0 within protocols designed for computational spectroscopy. Also in this context, the QMD-FFs obtained for targets **V** to **VII** delivered in all cases a better agreement with experimental results with respect to transferred general purpose descriptions. Moreover, for the last two targets (**VI** and **VII**), where transferable databases are unavailable, the JOYCE3.0 protocol successfully provided a complete set of force field parameters, exploiting its capability of handling cases such as transition metal complexes with organic ligands or chromophores in their excited state.

From a broader perspective, the results presented in this work confirm the wide applicability of the JOYCE3.0 protocol, which can be reliably used to generate force field parameters that accurately describe the structure and dynamics of nearly any molecular target. This motivated us to enhance the code's usability: new automated procedures, an improved interface with widely used electronic structure programs and popular MD engines, as well as detailed support provided on a dedicated website,<sup>43</sup> will facilitate users in developing their own QMD-FFs to simulate their systems of interest. On the one hand, the intramolecular term of JOYCE3.0 can be combined with an intermolecular component, making it easily adaptable for the simulation of simple solutions, complex fluids,<sup>33</sup> or multi-component devices.<sup>18,39,92,93</sup> On the other hand, the ability to generate consistent ground- and excited-state FFs (using TD-DFT, post-HF or multi-reference techniques) enables the integration of JOYCE3.0 FFs into cost-effective classical CL/QM protocols, allowing for realistic simulations of optoelectronic properties. The integration of QMD-FFs in the novel Ad-MD|gVH<sub>MM</sub> procedure proposed in the present work is probably the most striking example of JOYCE3.0 potentialities in this framework, where overcoming the structural mismatch allows for accurately predicting the spectral shape without resorting to any phenomenological broadening. As demonstrated in this work, the capabilities of this new release of JOYCE have matured to a level of general applicability, that allows it to handle a wide range of systems, encompassing both ground and excited states. This versatility opens the door to numerous exciting applications, from the rational design of advanced materials design to the study of photoinduced processes in biological as well as complex heterogeneous systems.

On the development side, future efforts may instead focus on establishing standardized protocols for efficient and reproducible parameterization schemes that rely solely on the structural information of the target molecule. Undoubtedly, significant steps have been taken in this direction, including the creation of a website<sup>43</sup> with comprehensive information and tutorials that present a unified view of the parameterization strategy and help its user base to adopt common protocols. Additionally, tools developed for the (semi-)automatic generation of reasonable internal coordinates (ICs) contribute to this effort. Still, the selection of the set of ICs remains a critical step, and some degree of arbitrariness persists, particularly in the selection of non-bonded terms or the set of redundant dihedral coordinates. These aspects call for further investigation of different settings, which should be carefully analyzed and classified in terms of their efficiency and reliability. In this sense, ongoing work in the IC selection tools include the design of different standard strategies, that could be reliably compared, including a systematic way to describe more exotic ICs sometimes required in specific FFs. Such transparent protocols would greatly facilitate the creation of databases where new force fields are categorized not only based on the target molecule and quantum mechanical level but also on the set of ICs selected.

## Acknowledgments

G.P., S.G., A.P., A.S. and D.P. thank the financial support from ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by European Union – NextGenerationEU – PNRR, Missione 4 Componente 2 Investimento 1.4. J.C. and J.P.G. thank the Spanish Ministerio de Ciencia e Innovación, Project PID2023-151719NA-I00, for financial support. J.P.G. also thanks Universidad de Murcia for a predoctoral fellowship. D.P. gratefully acknowledges funding from the Italian Ministry of University and Research (MUR, PRIN grant 2022CXHY3A) and computational resources kindly provided by Università di Siena (hpc@dbcf). A.L. gratefully acknowledges funding from the Italian Ministry of University and Research (MUR, PRIN grant 2022XSC9P5). Dr. Fabrizio Santoro is gratefully acknowledged by all authors for the many useful discussions, the continuous support and stimulating ideas. P.M.M. and J.G.V acknowledge the assistance and computing resources from Red Española de Supercomputación (RES-BSC) HPC computational facilities, funding from

the Spanish CM “Talento Program” (Project No. 2020-T1/ND-20306), the Spanish Ministerio de Ciencia e Innovación (Grant Nos. PID2020-113722RJ-I00, TED2021-132219A-I00 and CNS2023-144011) and the Spanish Ministerio de Educación y y Formación Profesional (Grant No. FPU21/06224).

## Supporting Information

Additional computational details on general FF models and MD implementations. Further information on the generation of the QM database and its file format. Details on the employed computational spectroscopy techniques. Supplementary validation results. A complete list of QMD-FF parameters for each considered target. This information is available free of charge via the Internet at <http://pubs.acs.org>

## References

- [1] Committee on Challenges for the Chemical Sciences in the 21st Century, Board on Chemical Sciences and Technology, *Beyond the Molecular Frontier*; National Academies Press, 2003.
- [2] Cragg, P. J. *Supramolecular Chemistry - From Biological Inspiration to Biomedical Applications*; Springer Netherlands, 2010.
- [3] Lehn, J.-M. Supramolecular chemistry: Where from? Where to? *Chem. Soc. Rev.* **2017**, *46*, 2378–2379.
- [4] Chen, Q.; Zhang, X. Nanoparticle self-assemblies with modern complexity *MRS Bulletin* **2024**, *49*, 310–318.
- [5] Yi, J.; Zhang, G.; Yu, H.; Yan, H. Advantages, challenges and molecular design of different material types used in organic solar cells *Nat. Rev. Mater.* **2023**, *9*, 46–62.
- [6] Lee, H.; Park, C.; Sin, D. H.; Park, J. H.; Cho, K.; Recent advances in morphology optimization for organic photovoltaics; 2018.

- [7] Heinzmann, C.; Weder, C.; de Espinosa, L. M. Supramolecular polymer adhesives: advanced materials inspired by nature *Chem. Soc. Rev.* **2016**, *45*, 342–358.
- [8] Cao, Z.; Tolba, S. A.; Li, Z.; Mason, G. T.; Wang, Y.; Do, C.; Rondeau-Gagné, S.; Xia, W.; Gu, X. Molecular Structure and Conformational Design of Donor-Acceptor Conjugated Polymers to Enable Predictable Optoelectronic Property *Adv. Mat.* **2023**, *35*.
- [9] Chen, J.; Zhu, M.; Shao, M.; Shi, W.; Yang, J.; Kuang, J.; Wang, C.; Gao, W.; Zhu, C.; Meng, R.; Yang, Z.; Shao, Z.; Zhao, Z.; Guo, Y.; Liu, Y. Molecular Design of Multifunctional Integrated Polymer Semiconductors with Intrinsic Stretchability, High Mobility, and Intense Luminescence *Adv. Mat.* **2024**, *36*.
- [10] Hogan, B. T.; Kovalska, E.; Craciun, M. F.; Baldycheva, A. 2D material liquid crystals for optoelectronics and photonics *J. Mat. Chem. C* **2017**, *5*, 11185–11195.
- [11] Gao, H.; Yu, R.; Ma, Z.; Gong, Y.; Zhao, B.; Lv, Q.; Tan, Z. Recent advances of organometallic complexes in emerging photovoltaics *J. Pol. Sci.* **2022**, *60*, 865–916.
- [12] Odínokov, A.; Yakubovich, A.; Son, W.-j.; Jung, Y.; Choi, H. Exploiting the quantum mechanically derived force field for functional materials simulations *Comput. Mat.* **2021**, *7*, 155.
- [13] Prampolini, G.; Silveira, L. G. D.; Vilhena, J. G.; Livotto, P. R. Predicting Spontaneous Orientational Self-Assembly: In Silico Design of Materials with Quantum Mechanically Derived Force Fields *J. Phys. Chem. Lett.* **2022**, *13*, 243–250.
- [14] Csizi, K.-S.; Reiher, M. Universal QM/MM approaches for general nanoscale applications *WIREs Comp. Mol. Sci.* **2022**, doi:<https://doi.org/10.1002/wcms.1656>.
- [15] Frenkel, D.; Smith, B. *Understanding Molecular Simulations*; Academic Press: San Diego, 1996.
- [16] Coutinho, K.; Rivelino, R.; Georg, H. C.; Canuto, S. *The Sequential qm/mm Method and its Applications to Solvent Effects in Electronic and Structural Properties of Solutes*; Springer, Dordrecht, 2008; pp 159–189.

- [17] Singh, V.; Patra, S.; Murugan, N. A.; Toncu, D.-C.; Tiwari, A. Recent trends in computational tools and data-driven modeling for advanced materials *Materials Advances* **2022**, *3*, 4069–4087.
- [18] Landi, A.; Padula, D.; Peluso, A. Fast Nonradiative Decay Paths in Organic Solar Cells: Implications for Designing More Efficient Photovoltaic Systems *ACS Appl. Energy Mater.* **2024**, *7*, 707–714.
- [19] Padula, D.; Landi, A.; Prampolini, G. Assessing alkyl side chain effects on electron transport properties of Y6-derived non-fullerene acceptors *Ener. Adv.* **2023**, *2*, 1215–1224.
- [20] De Mitri, N.; Monti, S.; Prampolini, G.; Barone, V. Absorption and Emission Spectra of a Flexible Dye in Solution: A Computational Time-Dependent Approach *J. Chem. Theory Comput.* **2013**, *9*, 4507–4516.
- [21] Dellai, A.; Naim, C.; Cerezo, J.; Prampolini, G.; Castet, F. Dynamic effects on the nonlinear optical properties of donor acceptor stenhouse adducts: insights from combined MD plus QM simulations *Phys. Chem. Chem. Phys.* **2024**, *26*.
- [22] König, G.; Riniker, S. On the faithfulness of molecular mechanics representations of proteins towards quantum-mechanical energy surfaces *Interface Focus* **2020**, *10*, 20190121.
- [23] Padula, D.; Lee, M. H.; Claridge, K.; Troisi, A. Chromophore-Dependent Intramolecular Exciton–Vibrational Coupling in the FMO Complex: Quantification and Importance for Exciton Dynamics *J. Phys. Chem. B* **2017**, *121*, 10026–10035.
- [24] Bolis, G.; Clementi, E. Analytical potentials from ab initio computations for interaction between biomolecules .3. Reliability and transferability of pair potentials. *J. Am. Chem. Soc.* **1977**, *99*, 5550–5557.
- [25] 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.

- [26] Vanduyfhuys, L.; Vandenbrande, S.; Verstraelen, T.; Schmid, R.; Waroquier, M.; Van Speybroeck, V. QuickFF: A Program for a Quick and Easy Derivation of Force Fields for Metal-Organic Frameworks from Ab Initio Input. *J. Chem. Theory Comput.* **2015**, *36*, 1015–27.
- [27] Riniker, S. Fixed-Charge Atomistic Force Fields for Molecular Dynamics Simulations in the Condensed Phase: An Overview *J. Chem. Inf. Mod.* **2018**, *58*, 565–578.
- [28] Horton, J. T.; Allen, A. E. A.; Dodda, L. S.; Cole, D. J. QUBEKit: Automating the Derivation of Force Field Parameters from Quantum Mechanics *J. Chem. Inf. Mod.* **2019**, *59*, 1366–1381.
- [29] Ghanavati, R.; Escobosa, A. C.; Manz, T. A. An automated protocol to construct flexibility parameters for classical forcefields: applications to metal–organic frameworks *RSC Advances* **2024**, *14*, 22714–22762.
- [30] 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.
- [31] Cacelli, I.; Prampolini, G. Parametrization and Validation of Intramolecular Force Fields Derived from DFT Calculations *J. Chem. Theory Comput.* **2007**, *3*, 1803–1817.
- [32] Prampolini, G.; Livotto, P. R.; Cacelli, I. Accuracy of Quantum Mechanically Derived Force-Fields Parameterized from Dispersion-Corrected DFT Data: The Benzene Dimer as a Prototype for Aromatic Interactions. *J. Chem. Theory Comput.* **2015**, *11*, 5182–96.
- [33] Vilhena, J. G.; Greff da Silveira, L.; Livotto, P. R.; Cacelli, I.; Prampolini, G. Automated Parameterization of Quantum Mechanically Derived Force Fields for Soft Materials and Complex Fluids: Development and Validation *J. Chem. Theory Comput.* **2021**, *17*, 4449–4464.

- [34] Giannini, S.; Cerda, J.; Prampolini, G.; Santoro, F.; Beljonne, D. Dissecting the nature and dynamics of electronic excitations in a solid-state aggregate of a representative non-fullerene acceptor *J. Mat. Chem. C* **2024**.
- [35] Prampolini, G.; Andersen, A.; Poulter, B. I.; Khalil, M.; Govind, N.; Biasin, E.; Pastore, M. Integrated Quantum-Classical Protocol for the Realistic Description of Solvated Multi-nuclear Mixed-Valence Transition-Metal Complexes and Their Solvatochromic Properties *J. Chem. Theory Comput.* **2023**, *20*, 1306–1323.
- [36] Pawlak, R.; Vilhena, J. G.; D’Astolfo, P.; Liu, X.; Prampolini, G.; Meier, T.; Glatzel, T.; Lemkul, J. A.; Häner, R.; Decurtins, S.; Baratoff, A.; Pérez, R.; Liu, S.-X.; Meyer, E. Sequential Bending and Twisting around C–C Single Bonds by Mechanical Lifting of a Pre-Adsorbed Polymer *Nano Lett.* **2020**, *20*, 652–657.
- [37] Vilhena, J. G.; Pawlak, R.; D’Astolfo, P.; Liu, X.; Gnecco, E.; Kisiel, M.; Glatzel, T.; Pérez, R.; Häner, R.; Decurtins, S.; Baratoff, A.; Prampolini, G.; Liu, S.-X.; Meyer, E. Flexible Superlubricity Unveiled in Sidewinding Motion of Individual Polymeric Chains *Phys. Rev. Lett.* **2022**, *128*, 216102.
- [38] Cerezo, J.; Gao, S.; Armaroli, N.; Ingrosso, F.; Prampolini, G.; Santoro, F.; Ventura, B.; Pastore, M. Non-Phenomenological Description of the Time-Resolved Emission in Solution with Quantum-Classical Vibronic Approaches-Application to Coumarin C153 in Methanol *Molecules* **2023**, *28*, 1–24.
- [39] Cerezo, J.; Gierschner, J.; Santoro, F.; Prampolini, G. Explicit Modelling of Spectral Bandshapes by a Mixed Quantum-Classical Approach: Solvent Order and Temperature Effects in the Optical Spectra of Distyrylbenzene *ChemPhysChem.* **2024**.
- [40] Segalina, A.; Aranda, D.; Green, J. A.; Cristino, V.; Caramori, S.; Prampolini, G.; Pastore, M.; Santoro, F. How the Interplay among Conformational Disorder, Solvation, Local, and Charge-Transfer Excitations Affects the Absorption Spectrum and Photoinduced Dynamics of Perylene Diimide Dimers: A Molecular Dynamics/Quantum Vibronic Approach *J. Chem. Theory Comput.* **2022**, *18*, 3718–3736.



- [41] Cigrang, L. L. E.; Green, J. A.; Gomez, S.; Cerezo, J.; Improta, R.; Prampolini, G.; Santoro, F.; Worth, G. A. Non-adiabatic direct quantum dynamics using force fields: Toward solvation *J. Chem. Phys.* **2024**, *160*.
- [42] Prampolini, G.; Cerezo, J.; Giannini, S.; De Mitri, N.; Cacelli, I.; JOYCE3.0, intra-molecular force field parameterization software, available free of charge at <http://www.iccom.cnr.it/en/joyce-2/>, last consulted Dec 2024; 2024.
- [43] Martinez, P. M.; Piras, A.; Giannini, S.; Semmeq, A.; Galvez, J. P.; Padula, D.; Cerezo, J.; Vilhena, J. G.; Prampolini, G.; JOYCE website: <https://joyce-documentation.gitlab.io/>, last consulted December; 2024.
- [44] Lv, L.; Roberts, J.; Xiao, C.; Jia, Z.; Jiang, W.; Zhang, G.; Risko, C.; Zhang, L. Triperylene[3,3,3]propellane triimides: achieving a new generation of quasi- D 3h symmetric nanostructures in organic electronics *Chem. Sci.* **2019**, *10*, 4951–4958.
- [45] Dyker, G.; Kerl, T.; Körning, J.; Bubenitschek, P.; Jones, P. G. An Improved Synthesis of Trisnaphtho[3.3.3]propellane *Tetrahedron* **2000**, *56*, 8665–8668.
- [46] Trejo, L. M.; Costas, M.; Patterson, D. Effect of molecular size on the W-shaped excess heat capacities: oxaalkane–alkane systems *Journal of the Chemical Society Faraday Transactions* **1991**, *87*, 3001–3008.
- [47] Haynes, W. M. CRC Handbook of Chemistry and Physics; CRC Press eBooks, 2014.
- [48] Luning Prak, D. J. Densities, Viscosities, and Speeds of Sound of n-Tetradecane and n-Alkylcyclohexane Binary Mixtures within the Temperature Range (288.15–333.15) K *Journal of Chemical & Engineering Data* **2023**, *68*, 1610–1623.
- [49] Schmidt-Rohr, K.; Hu, W.; Zumbulyadis, N. Elucidation of the Chain Conformation in a Glassy Polyester, PET, by Two-Dimensional NMR *Science* **1998**, *280*, 714–717.
- [50] Joo, S.; Cho, I. J.; Seo, H.; Son, H. F.; Sagong, H.-Y.; Shin, T. J.; Choi, S. Y.; Lee, S. Y.; Kim, K.-J. Structural insight into molecular mechanism of poly(ethylene terephthalate) degradation *Nat. Comm.* **2018**, *9*, 382.

- [51] Ernst, O. P.; Lodowski, D. T.; Elstner, M.; Hegemann, P.; Brown, L. S.; Kandori, H. Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms *Chem. Rev.* **2014**, *114*, 126–163.
- [52] Sen, S.; Kar, R. K.; Borin, V. A.; Schapiro, I. Insight into the isomerization mechanism of retinal proteins from hybrid quantum mechanics/molecular mechanics simulations *WIREs Comput. Mol. Sci.* **2021**, *12*, e1562.
- [53] Gonciarz, A.; Żuber, M.; Zwoździak, J. Spectrochemical Properties and Solvatochromism of Tetradentate Schiff Base Complex with Nickel: Calculations and Experiments *ChemistryOpen* **2018**, *7*, 677–687.
- [54] Becker, R. S.; Seixas de Melo, J.; Maçanita, A. L.; Elisei, F. Comprehensive Evaluation of the Absorption, Photophysical, Energy Transfer, Structural, and Theoretical Properties of  $\alpha$ -Oligothiophenes with One to Seven Rings *J. Phys. Chem.* **1996**, *100*, 18683–18695.
- [55] Cerezo, J.; Aranda, D.; Avila Ferrer, F. J.; Prampolini, G.; Santoro, F. Adiabatic-Molecular Dynamics Generalized Vertical Hessian Approach: A Mixed Quantum Classical Method to Compute Electronic Spectra of Flexible Molecules in the Condensed Phase *J. Chem. Theory Comput.* **2020**, *16*, 1215–1231.
- [56] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski,

J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J.; Gaussian~16 Revision D.02; 2016; Gaussian Inc. Wallingford CT.

- [57] Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA quantum chemistry program package *J. Chem. Phys.* **2020**, *152*, 224108.
- [58] Zahariev, F.; Xu, P.; Westheimer, B. M.; Webb, S.; Vallejo, J. G.; Tiwari, A.; Sundriyal, V.; Sosonkina, M.; Shen, J.; Schoendorff, G.; Schlinsog, M.; Sattasathuchana, T.; Ruedenberg, K.; Roskop, L. B.; Rendell, A. P.; Poole, D.; Piecuch, P.; Pham, B. Q.; Mironov, V.; Mato, J.; Leonard, S.; Leang, S. S.; Ivanic, J.; Hayes, J.; Harville, T.; Gururangan, K.; Guidez, E.; Gerasimov, I. S.; Friedl, C.; Ferreras, K. N.; Elliott, G.; Datta, D.; Cruz, D. D. A.; Carrington, L.; Bertoni, C.; Barca, G. M. J.; Alkan, M.; Gordon, M. S. The General Atomic and Molecular Electronic Structure System (GAMESS): Novel Methods on Novel Architectures *J. Chem. Theory Comput.* **2023**, *19*, 7031–7055.
- [59] Smith, D. G. A.; Burns, L. A.; Simmonett, A. C.; Parrish, R. M.; Schieber, M. C.; Galvelis, R.; Kraus, P.; Kruse, H.; Di Remigio, R.; Alenaizan, A.; James, A. M.; Lehtola, S.; Misiewicz, J. P.; Scheurer, M.; Shaw, R. A.; Schriber, J. B.; Xie, Y.; Glick, Z. L.; Sirianni, D. A.; O'Brien, J. S.; Waldrop, J. M.; Kumar, A.; Hohenstein, E. G.; Pritchard, B. P.; Brooks, B. R.; Schaefer, H. F.; Sokolov, A. Y.; Patkowski, K.; DePrince, A. E.; Bozkaya, U.; King, R. A.; Evangelista, F. A.; Turney, J. M.; Crawford, T. D.; Sherrill, C. D. PSI4 1.4: Open-source software for high-throughput quantum chemistry *J. Chem. Phys.* **2020**, *152*, 184108.
- [60] Li Manni, G.; Galván, I. F.; Alavi, A.; Aleotti, F.; Aquilante, F.; Autschbach, J.; Avagliano, D.; Baiardi, A.; Bao, J. J.; Battaglia, S.; Birnoschi, L.; Blanco-González, A.; Bokarev, S. I.; Broer, R.; Cacciari, R.; Calio, P. B.; Carlson, R. K.; Couto, R. C.; Cerdán, L.; Chibotaru, L. F.; Chilton, N. F.; Church, J. R.; Conti, I.; Coriani, S.; Cuéllar-Zuquin, J.; Daoud, R. E.; Dattani, N.; Decleva, P.; de Graaf, C.; Delcey, M. G.; De Vico, L.; Dobrautz, W.; Dong, S. S.; Feng, R.; Ferré, N.; Filatov(Gulak), M.; Gagliardi, L.; Garavelli, M.; González, L.; Guan, Y.; Guo, M.; Hennefarth, M. R.; Hermes, M. R.; Hoyer, C. E.; Huix-

Rotllant, M.; Jaiswal, V. K.; Kaiser, A.; Kaliakin, D. S.; Khamesian, M.; King, D. S.; Kochetov, V.; Krośnicki, M.; Kumaar, A. A.; Larsson, E. D.; Lehtola, S.; Lepetit, M.-B.; Lischka, H.; Ríos, P. L.; Lundberg, M.; Ma, D.; Mai, S.; Marquetand, P.; Merritt, I. C. D.; Montorsi, F.; Mörchen, M.; Nenov, A.; Nguyen, V. H. A.; Nishimoto, Y.; Oakley, M. S.; Olivucci, M.; Oppel, M.; Padula, D.; Pandharkar, R.; Phung, Q. M.; Plasser, F.; Raggi, G.; Rebolini, E.; Reiher, M.; Rivalta, I.; Roca-Sanjuán, D.; Romig, T.; Safari, A. A.; Sánchez-Mansilla, A.; Sand, A. M.; Schapiro, I.; Scott, T. R.; Segarra-Martí, J.; Segatta, F.; Sergentu, D.-C.; Sharma, P.; Shepard, R.; Shu, Y.; Staab, J. K.; Straatsma, T. P.; Sørensen, L. K.; Tenorio, B. N. C.; Truhlar, D. G.; Ungur, L.; Vacher, M.; Veryazov, V.; Voß, T. A.; Weser, O.; Wu, D.; Yang, X.; Yarkony, D.; Zhou, C.; Zobel, J. P.; Lindh, R. The OpenMolcas *Web*: A Community-Driven Approach to Advancing Computational Chemistry *J. Chem. Theory Comput.* **2023**, *19*, 6933–6991.

[61] Cerezo, J.; Santoro, F. FCclasses3: Vibrationally-resolved spectra simulated at the edge of the harmonic approximation *J. Comput. Chem.* **2023**, *44*, 626–643.

[62] Cerezo, J.; Tools to interface *FCclasses3* with quantum chemistry codes, visit: [https://github.com/jcerezochem/fcc\\_tools](https://github.com/jcerezochem/fcc_tools), last consulted September; 2022.

[63] Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GRO-MACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers *SoftwareX* **2015**, *1-2*, 19 – 25.

[64] Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales *Comp. Phys. Comm.* **2022**, *271*, 108171.

[65] Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schütt, O.; Schiffmann, F.; Golze, D.; Wilhelm, J.; Chulkov, S.; Bani-Hashemian, M. H.; Weber, V.; Borštnik, U.; Taillefumier, M.; Jakobovits, A. S.; Lazzaro,

- A.; Pabst, H.; Müller, T.; Schade, R.; Guidon, M.; Andermatt, S.; Holmberg, N.; Schenter, G. K.; Hehn, A.; Bussy, A.; Belleflamme, F.; Tabacchi, G.; Glöß, A.; Lass, M.; Bethune, I.; Mundy, C. J.; Plessl, C.; Watkins, M.; VandeVondele, J.; Krack, M.; Hutter, J. CP2K: An electronic structure and molecular dynamics software package -Quickstep: Efficient and accurate electronic structure calculations *J. Chem. Phys.* **2020**, *152*.
- [66] Cerezo, J.; Prampolini, G.; Cacelli, I. Developing accurate intramolecular force fields for conjugated systems through explicit coupling terms *Theor. Chem. Acc.* **2018**, *137*, 80.
- [67] Wang, J.; Wang, W.; Kollman, P.; Case, D.; Tirado-Rives, J. Automatic atom type and bond type perception in molecular mechanical calculations *J. Mol. Graph. Model.* **2006**, *25*, 247260.
- [68] Dodda, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands *Nucleic Acids Res.* **2017**, *45*, W331–W336.
- [69] Procacci, P. PrimaDORAC: A Free Web Interface for the Assignment of Partial Charges, Chemical Topology, and Bonded Parameters in Organic or Drug Molecules *J. Chem. Inf. Mod.* **2017**, *57*, 1240–1245.
- [70] Padula, D.; A program to automatically select and group internal coordinates given a geometry. The output is a GROMACS topology ready to be used with the Joyce program, to parameterise a Force Field., <https://github.com/dpadula85/SelIntCoords>; 2024.
- [71] Galvez, J.; Automated generation of topologies intended for JOYCE parametrization procedure, [https://github.com/jpablogalvez/gen\\_FF](https://github.com/jpablogalvez/gen_FF); 2024.
- [72] D’Abramo, M.; Aschi, M.; Amadei, A. Theoretical modeling of UV-Vis absorption and emission spectra in liquid state systems including vibrational and conformational effects: Explicit treatment of the vibronic transitions *J. Chem. Phys.* **2014**, *140*, 164104.

- [73] Cerezo, J.; Ferrer, F. J. A.; Prampolini, G.; Santoro, F. Modeling Solvent Broadening on the Vibronic Spectra of a Series of Coumarin Dyes. From Implicit to Explicit Solvent Models *J. Chem. Theory Comput.* **2015**, *11*, 5810–5825.
- [74] Loco, D.; Cupellini, L. Modeling the absorption lineshape of embedded systems from molecular dynamics: A tutorial review *Int. J. Quant. Chem.* **2019**, *119*, e25726.
- [75] Zuehlsdorff, T. J.; Montoya-Castillo, A.; Napoli, J. A.; Markland, T. E.; Isborn, C. M. Optical spectra in the condensed phase: Capturing anharmonic and vibronic features using dynamic and static approaches *J. Chem. Phys.* **2019**, *151*, 074111.
- [76] Shedge, S. V.; Zuehlsdorff, T. J.; Khanna, A.; Conley, S.; Isborn, C. M. Explicit environmental and vibronic effects in simulations of linear and nonlinear optical spectroscopy *J. Chem. Phys.* **2021**, *154*, 084116.
- [77] Khanna, A.; Shedge, S. V.; Zuehlsdorff, T. J.; Isborn, C. M. Calculating absorption and fluorescence spectra for chromophores in solution with ensemble Franck–Condon methods *J. Chem. Phys.* **2024**, *161*, 044121.
- [78] Mukamel, S. Fluorescence and absorption of large anharmonic molecules - spectroscopy without eigenstates *J. Phys. Chem.* **1985**, *89*, 1077–1087.
- [79] Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713; PMID: 26574453.
- [80] 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.
- [81] Ogata, K.; Nakamura, S. Improvement of Parameters of the AMBER Potential Force Field for Phospholipids for Description of Thermal Phase Transitions *J. Phys. Chem. B* **2015**, *119*, 9726–9739.

- [82] Lindberg, G. E.; Baker, J. L.; Hanley, J.; Grundy, W. M.; King, C. Density, Enthalpy of Vaporization and Local Structure of Neat N-Alkane Liquids *Liquids* **2021**, *1*, 47–59.
- [83] Yang, X.; Zhang, M.; Gao, Y.; Cui, J.; Cao, B. Molecular dynamics study on viscosities of sub/supercritical n-decane, n-undecane and n-dodecane *J. Mol. Liq.* **2021**, *335*, 116180.
- [84] Pedraza-González, L.; Barneschi, L.; Marszałek, M.; Padula, D.; De Vico, L.; Olivucci, M. Automated QM/MM Screening of Rhodopsin Variants with Enhanced Fluorescence *J. Chem. Theory Comput.* **2022**, *19*, 293–310.
- [85] González, L. P.; Barneschi, L.; Padula, D.; Vico, L. D.; Olivucci, M. Evolution of the Automatic Rhodopsin Modeling (ARM) Protocol; Springer International Publishing, 2022; p 105–152.
- [86] Barneschi, L.; Marsili, E.; Pedraza-González, L.; Padula, D.; Vico, L. D.; Kaliakin, D.; Blanco-González, A.; Ferré, N.; Huix-Rotllant, M.; Filatov, M.; Olivucci, M. On the fluorescence enhancement of arch neuronal optogenetic reporters *Nat. Commun.* **2022**, *13*, 6432.
- [87] Jacquemin, D.; Adamo, C. Bond Length Alternation of Conjugated Oligomers: Wave Function and DFT Benchmarks *J. Chem. Theory Comput.* **2010**, *7*, 369–376.
- [88] Nielsen, I. B.; Lammich, L.; Andersen, L. H. S1 and S2 Excited States of Gas-Phase Schiff-Base Retinal Chromophores *Phys. Rev. Lett.* **2006**, *96*.
- [89] Prampolini, G.; Ingrosso, F.; Cerezo, J.; Iagatti, A.; Foggi, P.; Pastore, M. Short- and Long-Range Solvation Effects on the Transient UV–Vis Absorption Spectra of a Ru(II)–Polypyridine Complex Disentangled by Nonequilibrium Molecular Dynamics *J. Phys. Chem. Lett.* **2019**, *10*, 2885–2891.
- [90] Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models *Chem. Rev.* **2005**, 2999–3094.

- [91] A systematic comparison of the computational cost for each level of calculation has not been performed but, as reference, it can be noted that the CPU time required to compute the QM Hessian for a single snapshots is around 100 minutes (run on an Intel Xeon E5-2640), while the same calculation at MM level a took around 2 minutes.
- [92] Kraner, S.; Prampolini, G.; Cuniberti, G. Exciton Binding Energy in Molecular Triads *J. Phys. Chem. C* **2017**, *121*, 17088–17095.
- [93] Landi, A.; Padula, D. Multiple charge separation pathways in new-generation non-fullerene acceptors: a computational study *J. Mater. Chem. A* **2021**, *9*, 24849–24856.