JCTC Journal of Chemical Theory and Computation

Spectroscopy and Excited States

Assessment of multi-scale approaches for computing UV-Vis spectra in condensed phases: toward an effective yet reliable integration of variational and perturbative QM/MM approaches.

Sara Del Galdo, Balasubramanian Chandramouli, Giordano Mancini, and Vincenzo Barone

J. Chem. Theory Comput., **Just Accepted Manuscript** • DOI: 10.1021/acs.jctc.9b00120 • Publication Date (Web): 05 Apr 2019

ACS Publications

is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Assessment of multi-scale approaches for computing UV-Vis spectra in condensed phases: toward an effective yet reliable integration of variational and perturbative QM/MM approaches.

Sara Del Galdo,[†,‡] Balasubramanian Chandramouli,[¶,‡] Giordano Mancini,[‡,§] and Vincenzo Barone*[,‡,§]

†*Consiglio Nazionale delle Ricerche, Istituto di Chimica dei Composti OrganoMetallici (ICCOMCNR), UOS di Pisa, Area della Ricerca CNR, Via G. Moruzzi 1, I-56124 Pisa, Italy*

‡*Scuola Normale Superiore di Pisa, Piazza dei Cavalieri 7 I-56126, Pisa, Italy*

¶*Compunet, Istituto Italiano di Tecnologia (IIT), Via Morego 30, I-16163 Genova, Italy*

§*Istituto Nazionale di Fisica Nucleare (INFN) sezione di Pisa, Largo Bruno Pontecorvo 3, 56127 Pisa, Italy*

E-mail: vincenzo.barone@sns.it

Abstract

Computational simulation of UV/vis spectra in condensed phases can be performed starting from converged molecular dynamics simulations and then performing QM/MM computations for a statistically significant number of snapshots. However, the need of

variational solutions (e.g. ONIOM/EE) for a huge number of snapshots makes unpractical the use of state-of-the-art QM Hamiltonians. On the other hand, the effectivity of perturbative approaches (e.g. PMM) comes at the price of poor convergence for configurations strongly different from the reference one. In this paper we introduce an integrated strategy based on a cluster analysis of the MD snapshots: next a representative configuration for each cluster is treated at the ONIOM/EE level, whereas local fluctuations within each cluster are described at the PMM level. Some representative systems (uracil in dimethylformamide and in water and tyrosine zwitterion in water) are analysed to show the effectivity and flexibility of the proposed strategy.

1 Introduction

The accurate reproduction of the spectroscopic properties of molecular systems in condensed phase is certainly among the main aims of contemporary theoretical-computational chemistry[1–4]. The challenges in this context are essentially related to the description of the key features involved in such phenomena that, at variance with gas-phase processes, rely not only on the accurate characterization of the Quantum Mechanical (QM) properties of the chromophore, but also on the description of its interaction with the embedding environment. In fact, the simulation of the absorption/emission spectra in realistic conditions should in principle include (i) the study of the electronic states of the most stable conformers of medium-large size molecular systems, (ii) the characterization of the vibrations of each chromophore conformer, and (iii) the proper treatment of the interaction between the chromophore and the surrounding environment. Indeed, in the cases of complex systems (such as condensed phase systems), a brute force first-principle strategy in which the forces of the whole system are computed on the fly can be performed only employing oversimplified QM models[5]. In this context, multi-scale approaches are still the methods of choice for reliable and feasible calculations[6–11]. According to these multiscale methods, complex systems and the inherent occurring phenomena can be factorized to employ theoretical-computational methodologies

of different accuracy and cost to tackle the complete problems. To this end, the system of interest is partitioned into multiple fragments, each treated with a different theoretical approach. When quantum mechanics is combined with inexpensive classical Molecular Mechanics (MM) these approaches are known as QM/MM methods. QM/MM methods have been successfully applied to investigate many condensed phase problems ranging from chemical reactions in solution or complex media, to spectroscopic features of complex systems. Typically, the total energy of the (whole) system is computed according to either one of two ways, namely i) summing the energy of the system sub-parts treated at high level of theory, the energy of the remainder being described at classical level with a further classical term describing the interactions between the two parts (additive scheme) or ii) summing the energy of the high-level portion to the classical energy of the complete system then subtracting the classical energy of the system sub-part already accounted for at QM level of theory (extrapolative scheme). Besides, for computing the coupling between the quantum and classical sub-parts several different approaches have been proposed[10,12–16]. If the QM level calculation is performed on the isolated QM region, then the model is referred to as "mechanical embedding". Accordingly, the electrostatic interactions between the two regions are treated only at the classical level. Conversely, the "electronic embedding" (EE) scheme is exploited when the high-level calculations are performed incorporating the partial charges of the MM region into the QM Hamiltonian. The latter approach avoids the approximation of the QM charge distribution by means of point charges and allows the QM wave-function to be polarized by the MM region charges, thus providing a better description of the interaction between the two regions. Most of the present QM/MM applications to chemical reactivity and spectroscopy are based on electrostatic embedding schemes[8].

Within this general context, the Perturbed Matrix Method (PMM) has been developed[17–21]. Analogously to more conventional QM/MM procedures, the basics of the method consists in the partitioning of the system to define a sub-part to be treated at the QM level (the so-called Quantum Center, or QC) that interacts with the surrounding environment (like, e.g., the

solvent molecules) described at the MM level. At variance with more standard approaches, the perturbing effects on the QC properties due to the interaction with the environment is not directly included in the Hamiltonian operator, but it is obtained by diagonalizing the perturbed Hamiltonian matrix constructed in the basis set of the Hamiltonian eigenstates computed in the absence of perturbation. Accordingly, the unperturbed QM properties of the QC are computed *in vacuum* and the perturbation operator is built as a function of the electrostatic potential generated by the atomistic environment on the QC for each system configuration of the MM simulation. The method has been implemented in the Gaussian package[18] and very recently it has been expanded to include different levels of theory to deal with the perturbation term in the perturbed Hamiltonian matrix[20]. If the QC is a rigid or semi-rigid species (the simplest case) the unperturbed eigenstates are evaluated only at a single semi-classical QC geometry. When a flexible QC is studied, a set of reference structures is extracted from the MD simulation (typically by means of some clustering-procedure). Indeed, even when applying these hybrid methods, the subtle balance between accuracy and computational feasibility has to be tuned according to system/user needs. For instance, when applying conventional QM/MM procedures (like the well-known ONIOM procedure) it is hardly feasible to exploit the complete classical sampling for the successive QM calculations. In fact, depending on the size of the system, a huge number of snapshots are required to classically sample the system phase space. This implies that a reduced sampling has to be employed to assess the electronic properties of the overall system, going from several to hundreds of snapshots distributed on the whole configuration space. The more the configurations utilized for the computations, the more accurate the final result will be. On the contrary, when applying the PMM procedure, a very limited number of QM calculations is performed, corresponding to the QC conformations chosen as references, while the complete MM sampling can be utilized for evaluating the perturbation. This indeed comes at the price of a reduced accuracy in the treatment of the electrostatic interaction between the two regions (see Methods section for additional details). These considerations

suggest that effective hybrid models combining ONIOM/EE and PMM approaches could offer significantly improved cost/reliability ratios. The development and validation of such hybrid approaches is the main aim of the present contribution. To this end, we computed the electronic absorption spectrum of three systems, exploiting both the ONIOM procedure applied with the Electronic Embedding scheme (ONIOM/EE) and the PMM procedure. We studied Uracil in dimethylformamide (DMF), Uracil in water and Tyrosine zwitterion in water. The study carried out by applying the ONIOM/EE procedure represents the reference for subsequent accuracy comparisons between more approximate methods. Given the size of the systems studied, we could fairly consider that the use of hundreds of snapshots (we used 500, see Methods section) should be sufficient for an accurate evaluation of the spectroscopic features of the systems. By applying the different computational procedures, we move from use of the complete classical sampling for evaluating the quantum properties of the system to the limit of using just one reference structure in combination with the standard PMM procedure. In this way, starting from the very simple case of a semi-rigid solute in an aprotic solvent and proceeding to the (more complex) case of a zwitterionic flexible species in aqueous solution, we are able to analyse the conditions in which the standard PMM procedure can be convenient in terms of the accuracy/computational cost ratio. On these grounds, we introduce a combined approach in which the full ONIOM/EE method is used for a single representative configuration within each cluster found in a preliminary analysis of the snapshots issuing from the classical MD simulation. Next, the faster PMM model is employed to take into account the local fluctuations within each cluster.

2 Methods

2.1 The effects of the environment on the quantum properties of the system: zoom in the PMM

The well-known and (more) usual methods for treating the effects of the presence of the environment on the quantum properties of a system (either polarizable continuum or QM/MM methods) have been the object of a number of reviews[2,8,22–24]. Thus, herein we provide a brief summary only of the key aspects of PMM theoretical basis (details can be found elsewhere[25,26]).

As briefly outlined in the Introduction, the basis of the PMM relies on the definition of the system sub-part to be treated at the QM level, classically interacting with its environment described at atomistic level, which is a common feature among all the most widely used QM/MM methods. In the case of PMM, instead of including directly in the Hamiltonian operator the perturbation term, the effect of the perturbation is obtained by diagonalizing *a posteriori* the perturbed Hamiltonian matrix constructed in the basis set of the unperturbed Hamiltonian eigenstates. The unperturbed electronic properties of the QC are provided by QM calculations, whereas classical molecular dynamics (MD) simulations provide the statistical-mechanical ensembles for evaluating the perturbation. The unperturbed QM properties of the QC are evaluated *in vacuum*, thus defining the perturbation operator through the electrostatic potential generated by the atomistic environment on the QC for each step of the MD simulation. According to the original version of the method, the perturbation operator is expressed in terms of the electric potential acting on the QC, utilizing a single expansion for the whole QC (QC-based expansion) typically centred in the QC centre of mass. Depending on the atomistic/structural features of the QC, the expansion can be truncated either at the second or the third order (dipolar or quadrupolar approximation). Then, the higher order terms can be approximated by short-range potentials independent of the electronic coordinates. A new development has been recently proposed[20], in which

the perturbation is expressed in terms of its effect on each QC atom. That is, the perturbing electrostatic potential is expanded within the atomic region around the corresponding atomic centre (atom-based expansion). Such an expansion involves the definition of atomic charge and dipolar operators for each atomic region of the QC. In this case, the terms beyond the second order can be safely described by effective two-body potentials independent of the electronic coordinates. Moreover, the atomic dipoles can be neglected as the electronic density of each unperturbed electronic eigenstate within each atomic region is basically symmetric around the nucleus. Thus, in principle, the evaluation of the unperturbed electronic eigenstate atomic charges should be sufficient to build the perturbed Hamiltonian matrix. However, the unperturbed eigenstate atomic charges are very difficult to obtain when different eigenstates are involved. For such a reason, the adopted strategy consists of using the atom-based expansion only for the Hamiltonian matrix diagonal elements while the other (non-diagonal) elements are obtained by using the QC-based expansion.

2.1.1 Integrating PMM with ONIOM/EE

Here, a new strategy to apply the PMM is proposed. The quantum properties of the QC are computed explicitly accounting for the effect of the embedding environment via the Electronic Embedding scheme of the ONIOM method, at variance with the original version, which employs the QC in the gas phase. Some relevant QC-environment instantaneous configurations (or just one in the simplest case) are utilized as the reference conditions. Accordingly, for each MD frame the perturbed Hamiltonian matrix is expressed utilizing the eigenstates computed through the ONIOM/EE method as the basis set. Thus, the MM simulation provides the ensemble for evaluating the electrostatic potential fluctuations with respect to the one explicitly accounted for in the QM calculation. That is, for each step of the simulation, the perturbation operator is defined as the difference between the instantaneous electrostatic potential generated by the environment and the corresponding reference configuration. This new strategy was applied only for the aqueous solutions, for

which the solutes are expected to experience the most intense interactions with the solvent.

2.2 Application of clustering procedures and dimensionality reduction techniques

For the estimation of the spectroscopic properties of the systems of interest, we made use of a number of configurations sampled from the MD trajectory to apply the QM methods. MD data are both high dimensional and of relevant size: thousands of configurations are sampled in a simulation and for each of them many variables may be defined. Hence, we selected a number of configurations to apply QM methods both i) by reducing the size of the data set, applying clustering procedures to some property sampled by the MD, and ii) reducing the overall dimensionality of the trajectory, using the Principal Component Analysis (PCA), which is a dimensionality reduction technique employed in previous PMM studies[18,21]. Clustering procedures aim at finding inherent relationships between the members of a data set and, as a consequence, select a limited number of representative members to describe its properties. Basically, these procedures work maximizing the similarities between members of the same clusters and minimizing the one between members of different clusters. PCA works by performing a variable transformation from a set of correlated variables to a new set of uncorrelated variables. By selecting those uncorrelated variables that contain a predefined percentage of the variance (the so called Principal Components) a space of reduced dimensionality can be managed instead of the complete one, minimizing the information loss.

In this framework, clustering allows one to select a limited number of solute or solute plus solvent configurations to be used in computationally demanding QM calculations with (hopefully) minimal loss of detail, while the principal component analysis of the atomic fluctuations permits to predict a low-dimensional subspace in which the main QC motion is expected to occur. Further details about clustering and PCA are given in the next sections.

2.2.1 Selection of the relevant reference structures to apply the standard PMM procedure

A PCA of the QC atomic fluctuations over the MD trajectory is initially performed to assess the flexibility of the QCs[27,28]. In particular, the largest eigenvalue of the covariance matrix (corresponding to the largest internal motion of the QC) offers a criterion to assign the QC as either semi-rigid or flexible. In fact, depending on the semi-rigidity or flexibility of the QC, different computational routes to apply the PMM can be followed (note that the procedures described in this paragraph refer only to the PMM application based on the evaluation of the QC unperturbed properties in vacuum, whereas they do not apply to the newly developed strategy merging PMM and ONIOM/EE). In the case of a semi-rigid QC, the unperturbed quantum properties of only one semi-classical QC geometry are sufficient for describing the perturbed quantum states and properties. This is the case for the Uracil solute, whose QM properties were thus evaluated only on the geometry corresponding to the energy minimum. Indeed, the PCA showed that the QC did not undergo any dramatic structural deformation during the simulation (see Figure 1). The resulting eigenvalues are, in fact, tiny (10^{-3} nm^2 order of magnitude) and the projections of the MD QC conformations onto the plane defined by the first two eigenvectors of the covariance matrix (Essential plane) clearly correspond to small structural fluctuations around the most stable structure.

On the other hand, if the QC is flexible, several reference structures are needed to describe the conformational dependence of the perturbed quantum properties. This is the case of the Tyrosine zwitterion, which is characterized by 5 highly flexible dihedral angles. For this case, a number of MD configurations are chosen in order to provide a set of relevant reference structures to be used in a successive interpolation procedure to reconstruct the unperturbed QC quantum properties at any sampled MD configuration. The reference structures were selected from the QC configuration projections onto the two-dimensional space defined by the first two principal components (the Essential plane). To choose the reference structures the following two-step procedure was applied: i) an ordered grid was built on the essential plane

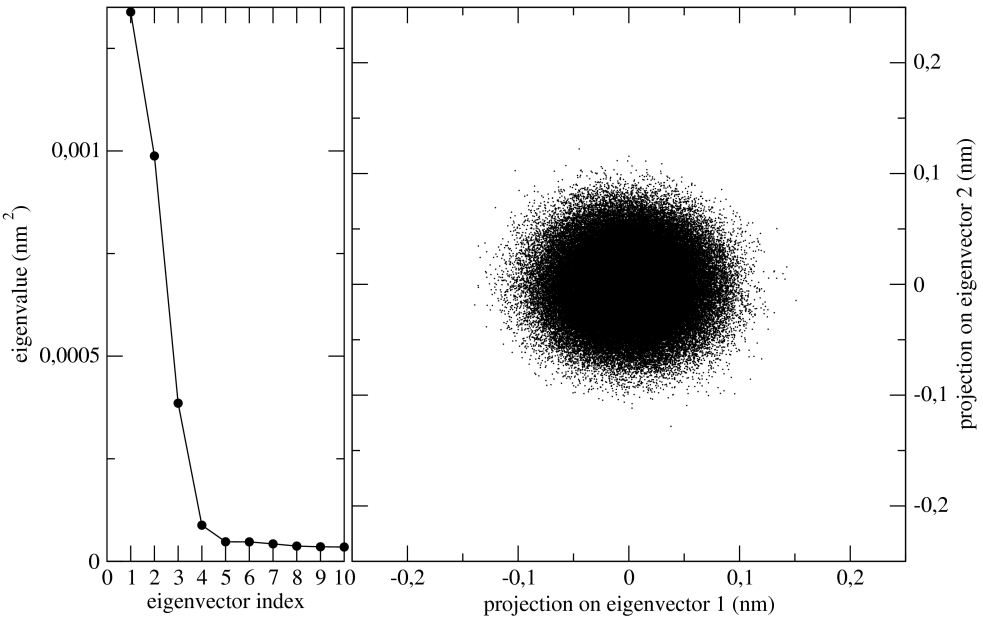

Figure 1: Left Panel: Spectrum of the first 10 eigenvalues of the covariance matrix of the QC geometrical fluctuations. Right Panel: Projections of the MD QC conformations onto the essential plane.

enclosing all the QC projections, taking a reference structure from the MD point closest to each grid node; ii) a density based spatial clustering method (Density Based Spatial Clustering of Applications with Noise, DBSCAN[29]) was applied to identify the relevant MD point clusters within the essential subspace. We set 40 and 0.005 as minimum number of points and radius threshold, respectively. Then, the MD structures closest to each centroid in the essential subspace accounted for new reference structures. On each structure the unperturbed QM properties of interest were computed and then interpolated by using the essential plane positions of the corresponding QC configurations to obtain the corresponding values for each MD time frame. The (cubic) Shepard's inverse distance weighted interpolation method[30–32] was used to achieve the goal. The two-step procedure implemented for the structures extraction was designed to get the most accurate results from the interpolation procedure. In fact, by means of the grid construction, a first stratified sampling is performed to properly sample the edges of the QC projections. Then, we applied the density based clustering procedure since we expect that some regions of the essential subspace can be

more populated than others, thus selecting a new reference structure from each highly dense region of the subspace. It is worth noting that the procedure currently developed does not deal with cases in which the QC changes its structure dramatically. In such cases, the approximation of describing the essential subspace only by the first two principal components can be not sufficiently accurate. Hence, the complete procedure (structures extraction + QM properties interpolation) should be performed considering a higher dimensional subspace. In light of such procedural limitation, in the present work an *ad hoc* selection of the atoms to be considered for the PCA was made, as shown in Figure 2. Such selection was aimed at giving the highest relevance to the proper dihedral angles, especially the one describing the rotation of the hydroxyl group around the benzene ring. As a matter of fact this dihedral strongly affects the phenol properties in the gas-phase, as shown in our recent work on the Tyrosine UV-vis spectrum in water[21]. To increase the computational efficiency of the procedure, a reduced MD sampling was utilized ($\approx$ 9000 frames, obtained by sampling the complete trajectory skipping each 13 MD frames). The projections of the QC configurations (either the complete trajectory sampling or the reduced one) onto the first two principal components, as well as the points chosen as reference structures are shown in Figure 2.

2.2.2 Selection of the relevant reference structures to apply the PMM procedure in conjunction with the Electronic Embedding scheme

A crucial point when applying the new strategy (i.e. combining PMM with ONIOM/EE) is the choice of the reference configuration(/s). Indeed, while the QM properties evaluated in vacuum depend only on the QC conformation chosen for the calculation, when dealing with ONIOM/EE it is the overall QC plus environment configuration that determines the QM properties of the QC. In the present case, the Uracil being a semi-rigid molecule, we could disregard the dependence of the QM properties from the QC conformational motions, thus focusing on the properties of the Uracil-SPC water interaction.

We analysed the probability of Hydrogen Bonding (HB) between the Uracil donor/acceptor

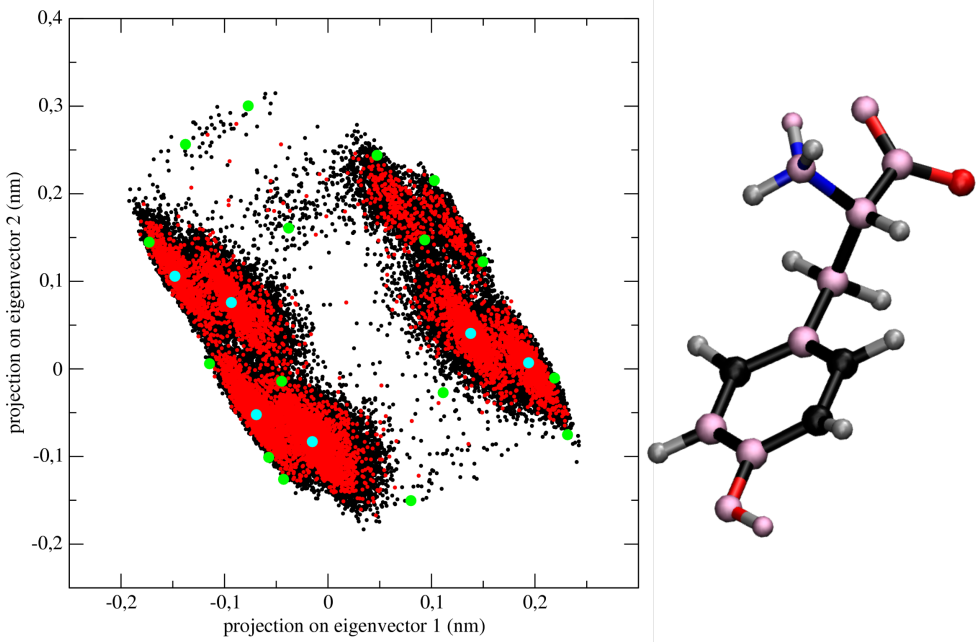

Figure 2: Left Panel: Projections of the QC configurations onto the Essential plane, as resulting from the the complete trajectory (black points) and from the reduced sampling (red points). The circles represent the projections corresponding to the chosen reference structures, either found with the rectangular grid (green circles) or corresponding to the centroids of the DBSCAN clusters (cyan circles).
Right Panel: Zwitterionic Tyr structure. The atoms selected for the PCA are highlighted in pink.

atoms and the SPC water, as a descriptor of the QC-environment interaction. In principle, the HB formation should be assessed accounting for the electronic structure and corresponding energy of the two partners engaged in the interaction. Nevertheless, we made use of our classical MD simulation exploiting the fulfilment of geometrical criteria reflecting the number of possible HB per atom for each MD time frame. Such geometrical criteria take into account both the distance between the HB donor(/acceptor) in Uracil and the HB acceptor(/donor) in the SPC molecule and the angle formed between these and the Hydrogen(/Oxygen) in the SPC molecule. We utilized a function (hereafter called F function), developed by Pagliai *et al.*, which is equal to 1 if the deviation of this distance and angle from suitable reference values are smaller than a threshold value, while it decreases exponentially for increasing deviations[33,34]. The radial and angular distribution functions are utilized to extract the exact parameters to be inserted into the function. In fact, the posi-

tion (a distance or an angle value) of the first peak in these curves, if any, corresponds to the geometrical threshold value and the half-width at half-maximum of the peak is utilized for weighting the geometrical deviations. The radial distribution functions were computed (see Figure 3) between the carbonyl oxygens, the aminic hydrogens and the nitrogens in Uracil and the corresponding HB acceptor/donor atom in an SPC molecule. From these data it is evident that the two carbonyl oxygens are the most likely atoms to form HBs with the SPC molecules, the corresponding curves showing a well-defined peak of increased density for short O-H distance. Thus, only for these two atoms we computed the angular distribution function (see Figure 3) and the F function[33,34]. We utilized the space defined by the two computed F functions to identify different sub-populations (clusters) within the Uracil-SPC MD ensemble. To this purpose, we applied the Kmeans clustering procedure. This method belongs to the class of clustering algorithms based on partition. The aim of the method is to partition the starting data set into k clusters, where k is a fixed parameter. Application of the Kmeans method with the elbow criteria (as done for instance in[35–38]) yielded 4 clusters, corresponding to the sub-populations where the Uracil molecule forms about one or two HBs with water through each of the two carbonyl oxygens (see also Figure 4). Each of these sub-trajectories was then utilized as a different ensemble for applying the PMM procedure according to the new strategy. Finally, for each sub-trajectory the reference state was chosen as the MD frame characterized by the electrostatic field components closest to the average values (averaged over each sub-trajectory). The identified sub-trajectories were used as the ensembles to evaluate the perturbation. The final spectrum was then obtained by summing the spectra computed for each of the sub-trajectories, weighting the parts by the corresponding sub-trajectory population.

We also applied the PMM procedure tailored with the EE scheme utilizing only one frame from the Uracil in SPC water MD trajectory. This configuration was chosen according to the value of the three components of the electric field acting on the center of mass of the solute. In fact, we utilized the MD frame characterized by the electric field components closest to the

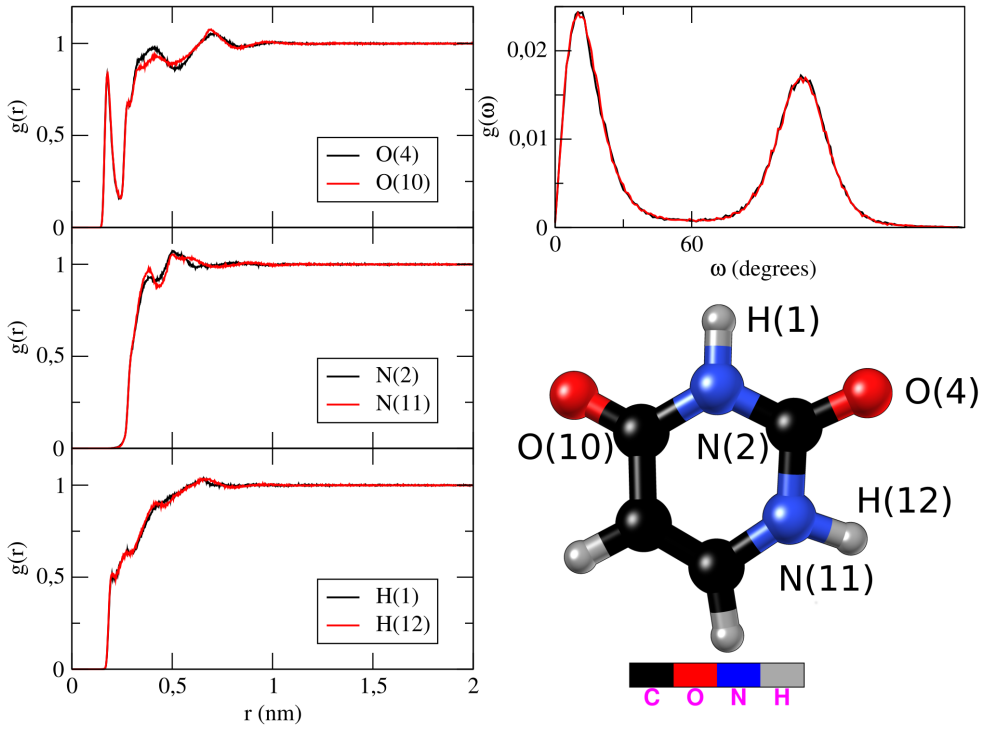

Figure 3: Left panels: Radial distribution functions between the O/H in SPC and the relevant HB donor/acceptor in the Uracil molecule (bottom right). Right panel: Angular distribution functions of the angle formed by Uracil carbonyl oxygens, H in SPC and O in the SPC.

average values as the reference configuration to be employed in the ONIOM/EE calculations (mean configuration). In the Supporting Informations (Figure S1) the distributions of the electric field components along with the correlations between them are shown. The presence of Uracil in the SPC solution does not induce any remarkable correlation between the electric field components, thus allowing us to straightforwardly identify (and extract) from the MD trajectory the mean configuration for the subsequent calculations.

Regarding the Zwitterionic Tyr-SPC water system, when applying the PMM strategy tailoring it with the Electronic Embedding scheme, the computational procedure proposed to deal with the flexibility of the QC (see section 2.2.1) cannot be applied. Nevertheless, in our previous work[21], we highlighted the role of the dihedral angle describing the rotation of

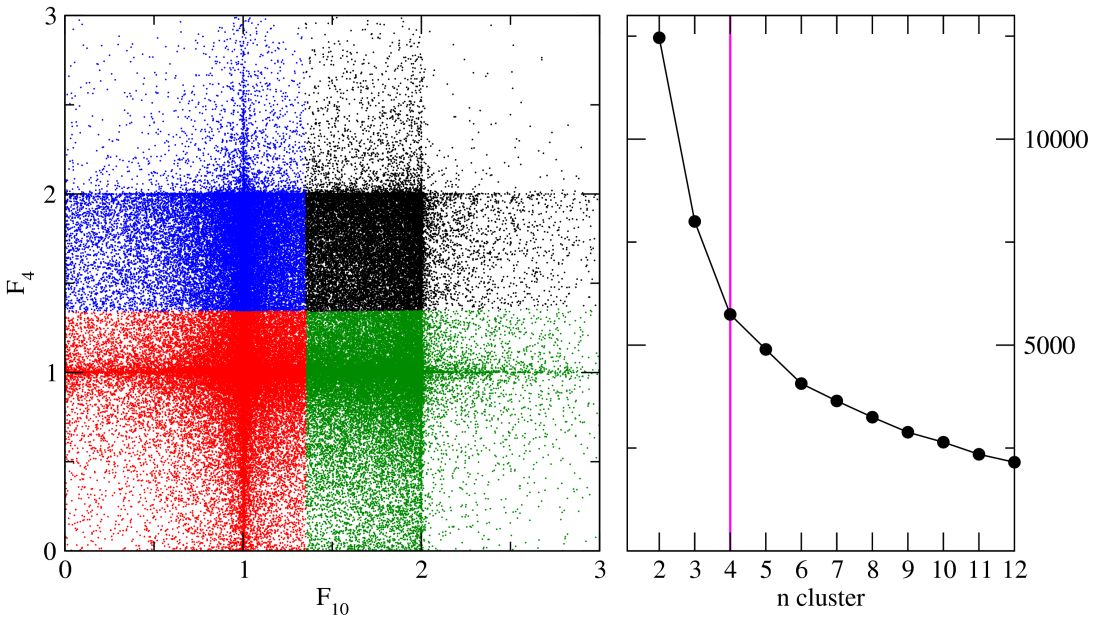

Figure 4: Left panel: The space defined by the F functions computed for the two carbonyl oxygens in the Uracil molecule (atoms 10 and 4, refer to Figure 3). A different color identifies each cluster. Right panel: Sum of squared distances of cluster samples to their closest center as a function of the number of clusters. The chosen number of cluster (4) is highlighted.

the hydroxyl group around the benzene ring in affecting the phenol QM properties. For this reason, in the present work, we divided the overall trajectory into 4 sub-trajectories according to the value of such a dihedral angle θ ($0.0 \leq \theta < 90.0$; $90.0 \leq \theta \leq 180.0$; $-180.0 \leq \theta <$ -90.0; $-90.0 \leq \theta \leq$ -0.0). The identified sub-trajectories were then utilized as the ensembles to evaluate the perturbation. In analogy with the previous case, for each ensemble, the reference frame to be employed for the ONIOM/EE calculations was chosen by selecting the frame characterized by the electric field on the QC center of mass components closest to the average. The identified sub-trajectories were then used as the ensembles to evaluate the perturbation, resulting in a different spectrum. Hence, by summing these spectra, weighted by the corresponding sub-trajectory population, the final spectrum was obtained.

2.3 Computational Details

2.3.1 Molecular Dynamics simulations

MD simulations of Uracil in N,N-dimethylformamide (DMF), Uracil in water and zwitterionic Tyrosine in water were performed with the Gromacs software package[39]. The force field (ff) utilized to model the Uracil molecule was parameterized by using QM data, according to the procedure implemented in the JOYCE program[40,41]. The target geometries, the corresponding energies, gradients and Hessian matrices, were obtained with the Gaussian16 package[42], by using Density-Functional Theory[43](DFT) with the hybrid B3LYP functional[44](B3LYP) in conjunction with 6-311 G(d) basis set. The atom types and Lennard Jones parameters were assigned according to the AMBER description[45], while the atomic charges were computed by using the RESP method[46]. For the Tyrosine zwitterion, the ff parameters derived in our previous study were utilized[21]. The SPC model was utilized for the aqueous solvent[47] while for the DMF an in-house derived force field[48] was utilized. All the systems were simulated in the isothermal-isochoric ensemble (NVT) with periodic boundary conditions, using an integration step of 2 fs. The temperature was kept constant (300 K for Uracil and Tyrosine zwitterion in water, and 298 K for Uracil in DMF) by the velocity-rescaling temperature coupling[49]. The bonds were constrained using the LINCS algorithm[50]. The particle mesh Ewald method[51] was used to compute long range interactions with grid search and cut-off radii of 1.1 nm. The simulation box densities were calibrated to obtain in the NVT MD simulations a pressure identical, within the noise, to the one provided by a corresponding reference MD simulation of the pure solvent. The reference simulations of the pure solvents were carried out in the NVT ensemble, using the same (room) temperature utilized for the production runs and imposing a box density equal to the experimental density of the corresponding pure solvent at room temperature and atmospheric pressure (that is, 55.32 mol per l for water and 12.91 mol per l for DMF[48]). The production runs of all the systems consist of 100 ns.

2.3.2 Polarizable continuum

500 MD time frames were extracted sequentially from the Uracil in DMF and Uracil in water MD trajectories. The corresponding Uracil conformations were utilized for subsequent QM calculations. We employed Density-Functional Theory with the CAM-B3LYP functional in conjunction with the 6-311G+(d) basis set. We mimicked the solvent effects by means of the Conductor-like Polarizable Continuum Model (C-PCM)[52,53] setting, accordingly, water and DMF as reference solvents. We used the Gaussian suite of programs for the calculations. To compute the spectra, the resulting excitation energies and oscillator strengths of the first 7 excited states were extracted from each calculation. We convoluted the electronic transitions by using Gaussian functions setting the sigma value empirically to reproduce the experimental spectral full width at half maximum (0.0015 a.u. of frequency).

2.3.3 Electronic Embedding calculations

500 MD time frames were extracted sequentially from the Uracil in DMF, Uracil in water and Tyrosine zwitterion in water MD trajectories to apply the ONIOM procedure. For the Uracil-water and Uracil-DMF systems, the complete MD box was utilized for the TDDFT calculations, treating the solute at the CAM-B3LYP/6-311+G(d) level of theory and employing all the remaining solvent molecules for the Electronic Embedding (EE). For the Zwitterionic Tyr-water system, a sphere of ≈ 2 nm was cut from each MD extracted configuration to be utilized for the ONIOM/EE calculations. The TDDFT calculations were then performed treating the complete zwitterionic Tyr at the B3LYP/6-311+G(d) level of theory.

For all the systems studied, we extracted the resulting excitation energies and oscillator strengths of the first 7 excited states from each calculation. The sigma values of the Gaussian functions used to convolute the transitions were set empirically in order to reproduce the experimental spectral full width at half maximum (0.0015 a.u. of frequency was used for both the Uracil simulated spectra and 0.00058 a.u. of frequency was utilized in the case of Tyrosine in aqueous solution).

2.3.4 PMM

For all the systems studied, we defined the complete solute as the Quantum Center (the same portion of the system utilized for the high level treatment in the ONIOM procedure). Computationally, to build the perturbed Hamiltonian matrix, the unperturbed energies, electric dipole expectation values, and transition moments of a number of electronic states were computed. For the Uracil in DMF and water systems, on the QC conformation corresponding to the vacuum minimum of energy, obtained at the B3LYP/6-311+G(d) level of theory, the first 10 unperturbed electronic states and the complete matrix of the corresponding dipole moments were evaluated using Time-Dependent DFT calculations[54], at the CAM-B3LYP/6-311+G(d) level of theory.For Tyrosine zwitterion in water, a set of reference structures was extracted from the MD simulation according to the procedure reported in the section 2.2.1. These reference structures were energy minimized with DFT at the B3LYP/6-31+G(d) level of theory. The optimization was performed constraining the dihedral angles and accounting for the solvent effects by means of C-PCM (thus preventing the atomic rearrangement that leads to the neutral form with no charge separation of the amino acid, as it occurs in vacuum). Then, by fitting the reference structures to one of them (mass weighted least square fitting procedure), a common orientation was assigned to the structures. On the (fitted) optimized geometries 11 unperturbed (i.e. *in vacuum*) electronic states have been evaluated (TDDFT, B3LYP/6-31+G(d) level of theory). Then, the unperturbed properties of interest (namely, the vacuum ground to excited state frequency transition for all the computed electronic states, and the complete matrix of the corresponding dipole moments) were interpolated within the essential plane. The complete MD trajectory was utilized as statistical ensemble for evaluating the perturbing effect of the environment.

For the aqueous systems (both Uracil and Tyrosine), where the solutes are expected to experience stronger interactions with the solvent with respect to the DMF case, we also employed the new hybrid strategy combining PMM and ONIOM/EE models (see section 2.1.1). For Uracil in aqueous solution, on the QC plus environment configuration/s selected

(see section 2.2.2) the first 10 unperturbed electronic states and the complete matrix of the corresponding dipole moments were evaluated using Time-Dependent DFT calculations[54], at the CAM-B3LYP/6-311+G(d) level of theory exploiting the Electronic Embedding scheme of the ONIOM procedure. For Tyrosine zwitterion in aqueous solution, on the QC plus environment configuration/s selected (see section 2.2.2) the first 10 unperturbed electronic states and the complete matrix of the corresponding dipole moments were evaluated using Time-Dependent DFT calculations[54], at the B3LYP/6-31+G(d) level of theory exploiting the Electronic Embedding scheme of the ONIOM procedure. For the same two systems, we also applied the recently proposed development[20], based on expressing the perturbation in terms of its effect on each QC atom[20] (see 2.1.1 section). To this end, we computed the atomic charges to be used for the diagonal elements of the Hamiltonian while the non-diagonal ones were expressed according to the QC-based expansion by means of the transition dipole moments. For Uracil in aqueous solution we utilized the RESP model for computing the atomic charges[46], whereas for Tyrosine zwitterion in water we utilized both the RESP and the CM5[55] models in order to compare the results between the two computational procedures. For all the electronic states, the CM5 charges were rescaled to be consistent with the diagonal electric dipole moments. We utilized the level of theory consistent with the previous computations.

For each of the calculated spectra, to take into account the effects of the chromophore semi-classical structural fluctuations, for all the MD frames each transition was modeled as a Gaussian function centred on the perturbed electronic transition frequency. For the Uracil in water spectrum, to compute the sigma value to be used in the convolution, four MD frames were randomly extracted from the MD trajectory of the system. On each of the extracted configurations the first 7 electronic state energies were computed by using the ONIOM/EE scheme. On the corresponding QC conformations, the same energies were computed in vacuum. Then, the standard deviations of the electronic energies obtained by both the computational procedures were evaluated. The average between the two values was

utilized as the sigma for the Gaussian convolution of the electronic transitions for all the Uracil spectra reported in the Results section (σ=0.0015 au of frequency). Instead, for the Tyrosine zwitterion case, the standard deviation utilized for the Gaussian convolution is the one estimated in the previous study on Tyrosine (σ=0.0008 a.u. of frequency)[21].

3 Results

3.1 A semi-rigid solute in aprotic solvent: Uracil in dimethylformamide absorption spectrum

For Uracil in DMF, we computed the UV-Vis absorption spectra through i) the continuum polarizable embedding approach, ii) the ONIOM electronic embedding scheme and iii) the standard PMM procedure, as shown in Figure 5 (see also Computational Details section). We could not find experimental data to compare our results. However, it can be noted that

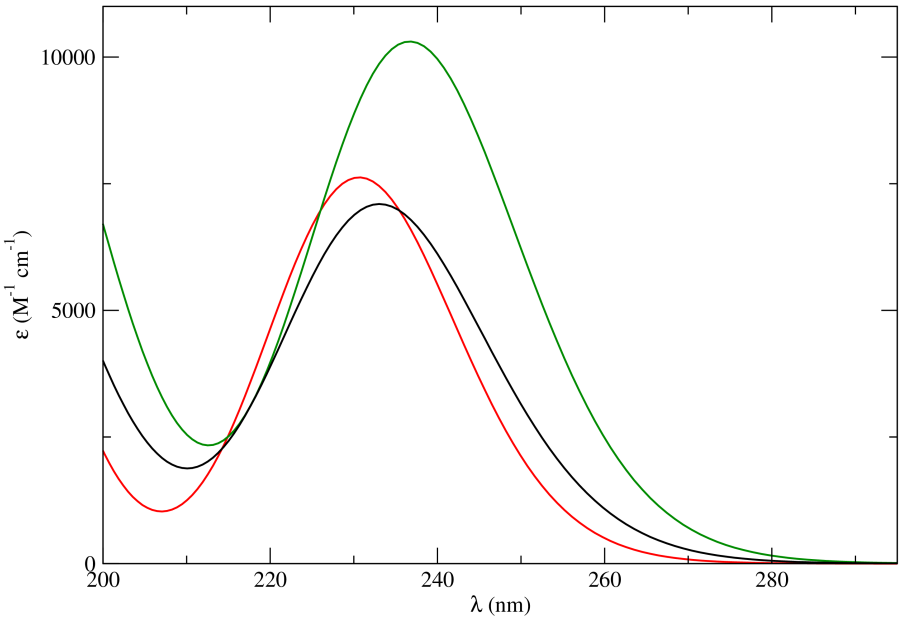

Figure 5: Uracil in DMF UV absorption spectra computed by applying the ONIOM/EE approach (black curve), the PMM procedure (red curve) and the C-PCM (green curve) approach.

the three computed spectra are in remarkable agreement. This is especially the case for the ONIOM/EE and PMM spectra. Indeed, the absorption maxima shift is only 2 nm (ie less then 0.05 eV) and the intensity difference is ≈ 600 M^{-1} cm^{-1}. Our data clearly show that for the case of a semi-rigid solute in aprotic solvent, moving from treating at the the QM level the "full" (according to our scope) MD sampling to employing only a single QC conformation does not relevantly affect the accuracy of the computed spectra.

3.2 A semi-rigid solute in water: Uracil in water absorption spectrum

In Figure 6 (refer to the black line) the computed spectrum obtained through the application of the ONIOM/EE procedure is shown for Uracil in aqueous solution. The spectrum i) is characterized by an absorption peak in the 220–280 nm region, as the experimental spectrum; ii) well reproduces the experimental intensity (≈ 9000 M^{-1} cm^{-1}) and iii) reproduces the experimental full width at half maximum of ≈ 30 nm (see also Table 1)[56]. The slight shift (28 nm, i.e. less than 0.6 eV) of the computed spectrum with respect to its experimental counterpart is well within the expected error bar of the underlying TDDFT calculations. These results, exploiting most of the MD sampling for the QM calculations, are the most accurate ones and will be used to benchmark less sophisticate models.

By means of HB analysis (see section 2.2.2) we identified four relevant sub-populations in the MD sampling according to the distribution of the HBs engaged by the Carboxyl Oxygens in the Uracil molecule with the SPC molecules. These sub-trajectories provided the four different ensembles to be used for applying the hybrid PMM/ONIOM-EE procedure. In Figure 7 the structures of Uracil in water corresponding to the MD frames utilized for applying the PMM procedure are shown. The solute molecules are shown with a 5 Å thick hydration shell and the water molecules closest to the Uracil are highlighted. For each sub-trajectory, the application of the PMM resulted in a different spectrum. Thus, the final spectrum reported in Figure 6 (red line) was obtained by summing the spectra obtained

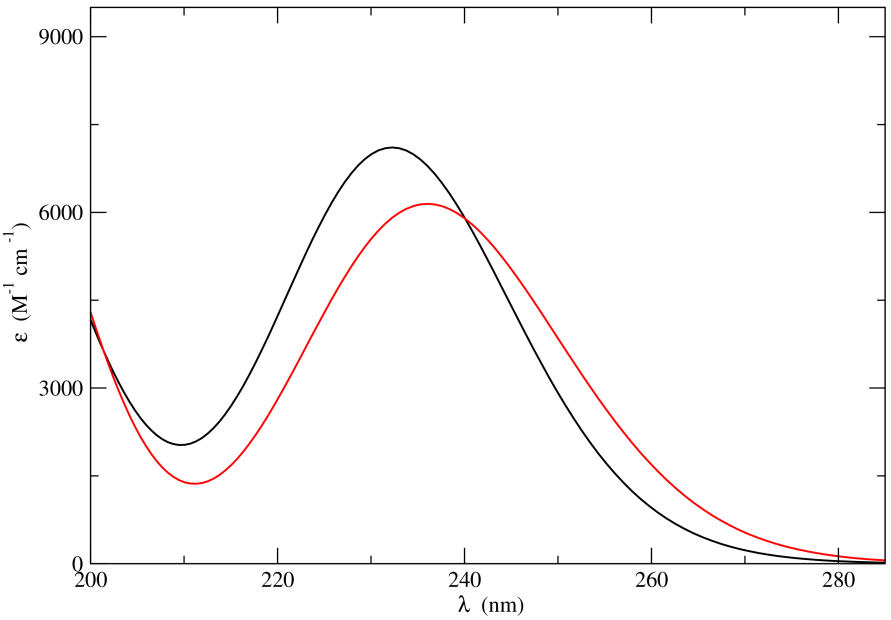

Figure 6: UV spectrum of Uracil in water obtained by applying the ONIOM procedure with the Electronic Embedding scheme (black line) and the PMM procedure according to the hydrogen bond analysis (*vide infra*, red line).

from each sub-trajectory, weighted by the corresponding cluster size.

Also in this case the computed spectrum is in good agreement with the experimental results. In Table 1 the main properties of all the computed spectra are compared with the corresponding experimental ones.

We employed only 4 MD snapshots for the QM calculations. It is worth noting that such computational gain is not accompanied by any relevant accuracy loss. The next step was the test of the results issuing from the use of a single snapshot for the QM calculations. To this end we applied both the original version of PMM and the new strategy for selecting the reference structure (see section 2.1.1). In the first case we utilized the QC conformation corresponding to the minimum of energy for the QM in vacuum calculations, whereas for the second case we selected the QC+environment configuration to be used for the ONIOM/EE calculation by looking at the electric field generated by the water molecules on the QC center of mass (see section 2.2.2). The results of such computations are reported in the Supporting Informations, figure S2. Note that for both cases the chosen level of theory for treating

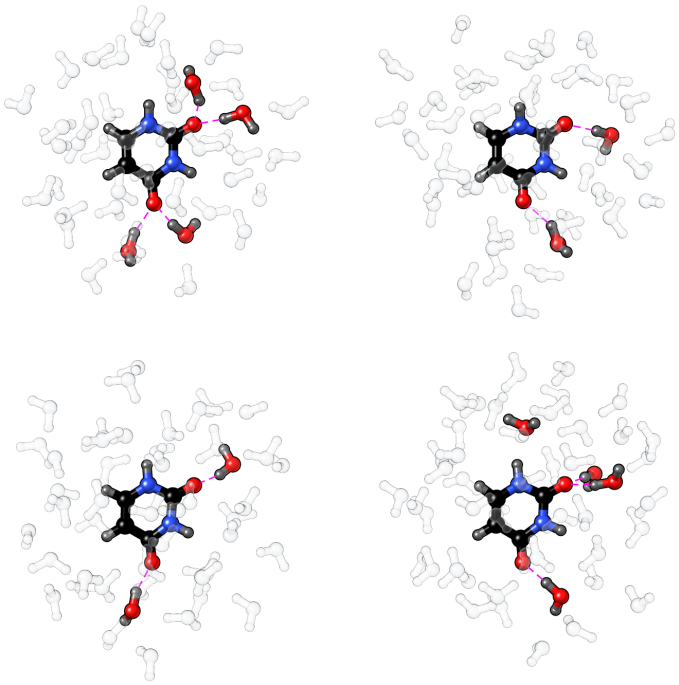

Figure 7: Representative structures of the different sub-populations identified through the HB analysis for the Uracil in water system. These structures were extracted from the MD frames utilized to apply the PMM. The solute molecule is shown along with a 5 Å thick hydration shell and highlighting the water molecules closest to the Uracil (i.e. less then 2 Å distance).

the perturbation operator in the PMM approach is the QC-based expansion in the dipolar approximation. When exploiting the new strategy (i.e. considering the solvated uracil as the reference system), we expressed the perturbation operator both in its original formulation and in terms of the perturbing field at each atom of the QC. The latter approach (atom-based expansion, see section 2.1) does not show any significant improvement with respect to the original QC-based expansion. The first three ground to excited state transitions characterizing the UV spectrum of Uracil in water obtained according to this procedure are shown in the Supporting Information as figure S3 . We utilized the RESP method for computing the uracil charges in different electronic states. In view of the close similarity between one-center and multi-center expansions in the QC, we did not test other computational procedures for the charge estimation. The two spectra obtained by employing only a single QM calculation are quite similar and both satisfactorily reproduce the experimental data[56] (see Table 1).

Indeed, also for this case, our results suggest that all approaches employed here are nearly equivalent in reproducing the essential features of the absorption spectrum.

Table 1: Main spectroscopic features of the uracil spectrum in DMF: position (λ_{max}) and intensity (ε_{max}) of the absorption maximum and full width at half maximum ($fwhm$).

	λ_{max} (nm)	ε_{max} ($M^{-1}cm^{-1}$)	$fwhm$ (nm)
Experimental	259	9000	30
ONIOM/EE	232	7100	30
PMM standard strategy	230	7800	26
PMM new strategy	227	8000	26
PMM new strategy HB analysis	236	6100	32

3.3 A flexible and complex solute in water: the absorption spectrum of tyrosine zwitterion in aqueous solution

3.3.1 The goal

The UV absorption spectrum of Tyr zwitterion in aqueous solution computed employing the ONIOM/EE procedure (see Methods section), is shown in Figure 8. The spectrum we obtained is characterized by an absorption maximum occurring at 256 nm (λ_{max}) with a molar extinction coefficient (ε_{max}) of 2200 M^{-1} cm^{-1} and a fwhm of 18 nm, in good agreement with the experimental results[57–61] ($\lambda_{max,exp}$=275 nm, $\varepsilon_{max,exp}$=1400 M^{-1} cm^{-1} and fwhm$_{exp}$=22, the of λ_{max} shift being again consistent with the expected TD-DFT error range). Moreover, both the ONIOM/EE and experimental spectra show an high energy absorption peak located at 220 nm (literature inconsistencies about the intensity of the absorption band prevented a more detailed comparison[57,60,61]).

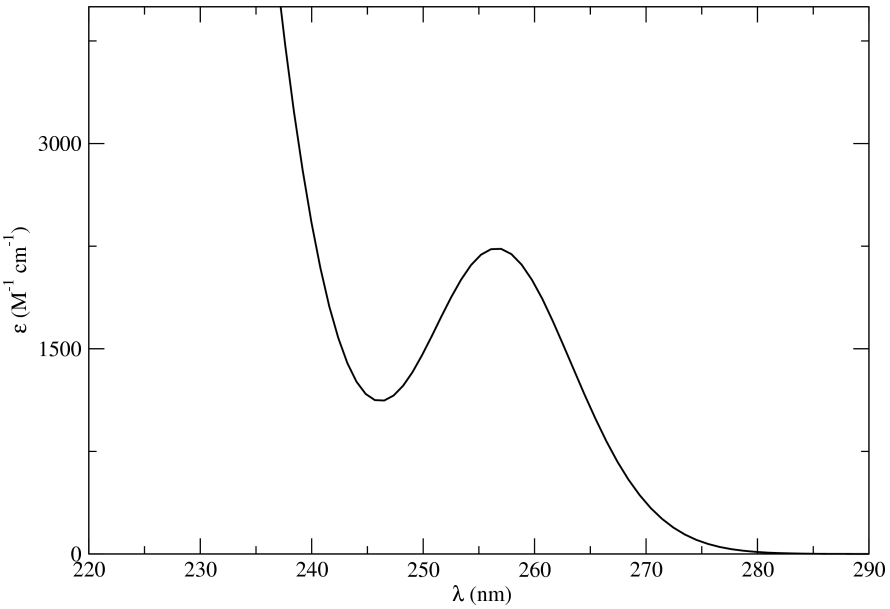

Figure 8: Zwitterionic Tyrosine in water UV absorption spectrum obtained through the ONIOM/EE procedure.

3.3.2 The (unsatisfactory) standard PMM procedure

The standard PMM strategy was then applied by following the procedure developed to deal with flexible QCs, as reported in section 2.2.1. This procedure made use of 22 conformations to compute the QC QM unperturbed properties and exploited the complete MD classical sampling, through the interpolation procedure, to model the QM perturbed properties of the system. At variance with previous cases, this approach gave disappointing results. The computed spectrum is reported in Figure 9 together with the contributions to the overall result due to the transitions from the ground to the first six perturbed electronic states. The computed spectrum is characterized by a single broad absorption band in the 180-400 nm region (fwhm=61 nm corresponding to $\approx$ 1 eV), given by the overlap of the broad ground to excited states transition peaks occurring even at very low energies. A Gaussian convolution of the electronic transitions was performed accounting for the effects of the semiclassical structural fluctuations to obtain the final spectrum reported in the figure. We utilized the σ value previously computed[21] (see section 2.3.4), but its exact value plays a

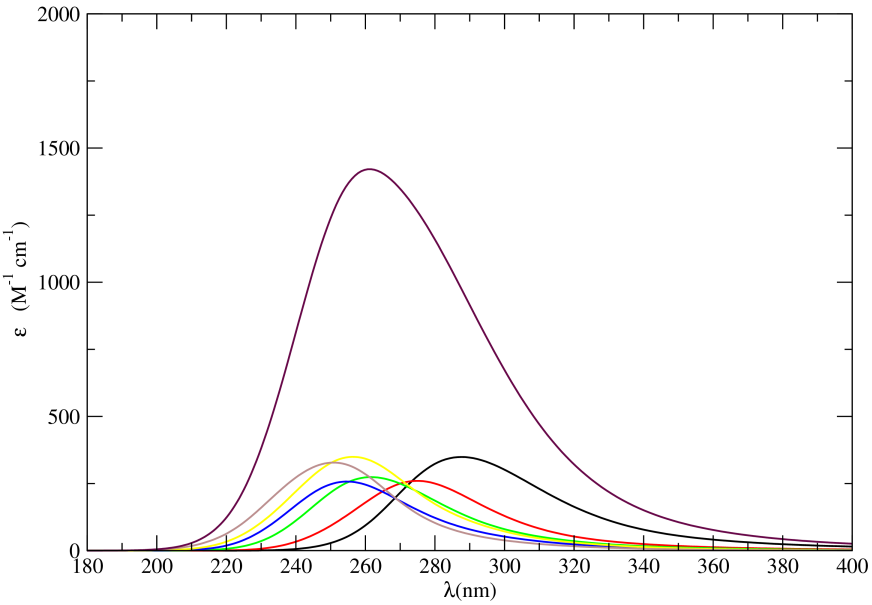

Figure 9: UV absorption spectrum of tyrosine zwitterion in aqueous solution computed via the standard PMM procedure (maroon curve) and contributions issuing from transitions to the first six perturbed electronic states (black, red, green, blue, yellow and brown curves, respectively).

negligible role in determining the final result. In fact, even decreasing the σ value by an order of magnitude the overall shape of the spectrum did not change (data not shown). Afterwards, we computed the same electronic transitions of Figure 9 in the absence of the perturbing environment, i.e. putting the charge of the SPC water atoms equal to zero while the remainder of the computation was performed in the same way (refer to the Supporting Informations, figure S4). Only the QC structural fluctuations contributed to this spectrum. Hence, from the comparison between the spectra reported in Figures 8, 9 and S4 , one may estimate the contribution of the environmental perturbing effects to the final spectrum. It is apparent that the blue-shift of the electronic transition induced by the perturbation is not sufficient to satisfactorily reproduce the experimental spectrum. From these data it emerges that the description of the interaction between the Tyr zwitterion and the embedding environment is not accurate enough to achieve a reasonable reproduction of the experimental spectrum. This may be due to a lack of accuracy of either the computational treatment of

the QC (high) flexibility, or the modelling of the QC-environment interaction. Namely, the description of the interaction by means of the electrostatic perturbation operator truncated at the second order in the QC-based expansion may be not accurate enough for the present case. Reasonably, given the dimension of the QC and the strong dipole moment of the backbone, the approximation of homogeneous or nearly-homogeneous perturbing field is probably too rough. Note that even in the atom-based expansion approximation, the coupling between the unperturbed states given by the non-diagonal elements of the perturbed Hamiltonian matrix exploits the QC-based expansion truncated at the dipolar term, thus being based on the same homogeneous or nearly-homogeneous perturbing field approximation.

3.3.3 The (succesfull) hybrid PMM/ONIOM-EE procedure

We then applied to Tyr zwitterion in aqueous solution the new hybrid strategy merging the PMM procedure with the ONIOM/EE scheme. To cope with the flexibility of the molecule, we divided the overall trajectory in 4 sub-trajectories according to the value of the dihedral angle formed by the hydroxyl group with the benzene ring (see section 2.2.2). The identified sub-trajectories were used as the ensembles to evaluate the perturbation. Within this framework, the PMM procedure was applied by means of both the QC based and the all atom based expansions, truncating the QC based expansion to the dipolar term. For the all atom approach, we utilized both the RESP and the CM5 models for computing the atomic charges to be used for the diagonal elements of the Hamiltonian (while the non-diagonal ones were expressed according to the QC-based expansion by means of the transition dipole moments). When applying these procedures, the quality of the PMM results improved greatly. In Figure 10 the UV spectra of Zwitterionic Tyr in water obtained by applying the PMM procedure by means of the QC based expansion in the dipolar approximation or by means of the atom-based expansion (employing both the RESP and the CM5 models) are shown. Indeed, following these procedures we were able to obtain a spectrum satisfactorily reproducing the experimental one, being characterized by two absorption peaks with the

lower energy one occurring around 255 nm. It is interesting to note that the choice of the level of theory for the expansion of the perturbation term (QC based or atom based), as well as the specific computational model for deriving the atomic charges does not significantly affect the final results.

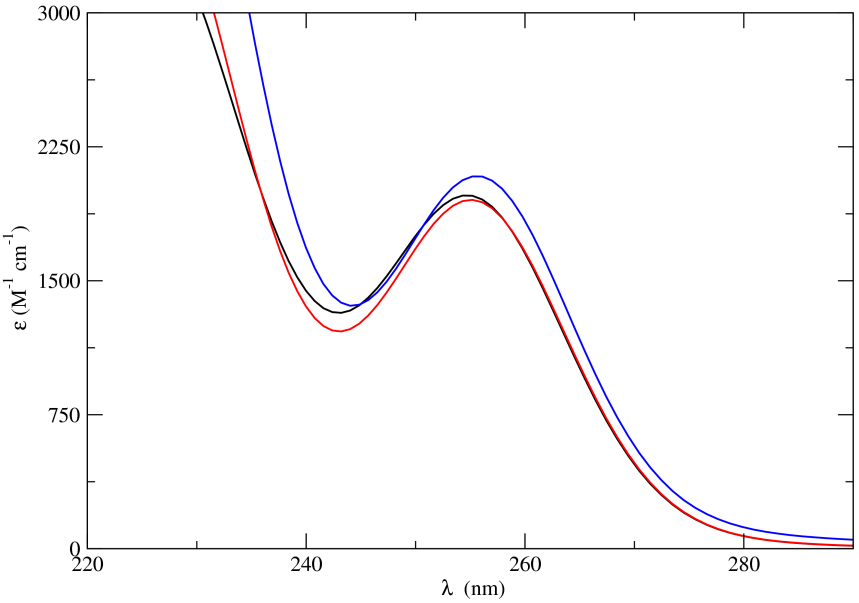

Figure 10: UV spectra of Zwitterionic Tyrosine in water obtained by applying the PMM procedure according to the QC based expansion in the dipolar approximation (blue line) or by means of the atom-based expansion, utilizing the RESP (black line) or the CM5 models (red line) to compute the atomic charges. Note that in this latter case the atom-based expansion is used only for the Hamiltonian matrix diagonal elements while the other (non-diagonal) elements are obtained by using the QC-based expansion (see Methods section).

This procedure allowed us to significantly reduce the number of QM calculations required to reproduce the experimental spectrum with respect to the full ONIOM/EE procedure. In fact, we moved from hundreds of QM ONIOM/EE calculations to only four. However, for the present case it is interesting to note that the use of the ONIOM/EE scheme is mandatory to accurately model the electrostatic coupling between the Tyr zwitterion and the water molecules. Recently, we studied the UV-vis spectrum of aqueous Tyrosine and its response to different backbone protonation states by applying the standard PMM with MD simulations[21]. In that case we considered as the QC only the side chain of the amino acid (i.e.

the p-Cresol moiety) while the backbone atoms and the water molecules were treated as the perturbing environment. We applied the standard PMM procedure utilizing the QC-based expansion to express the perturbation operator, truncating the expansion at the dipolar term. The approximation of modelling the chromophore as the tailored QC worked because the low-energy absorption band observed in the (experimental) spectrum is basically due to the p-Cresol electronic degrees of freedom while the backbone atoms play a consistent role only in defining the higher energies Tyr electronic states. The p-Cresol (i.e. the side-chain moiety in the structure of Figure 2) is characterized by only one highly flexible dihedral angle and both the molecular size and dipole are comparable with the Uracil molecule. Hence, it is not surprising that the procedure allowed us to fairly well reproduce the Tyr in water spectroscopic signal. Conversely, when considering the complete amino acid as the QC, given the high charge separation characterizing the backbone moiety in the Tyr zwitterion, the molecular dipole moment is ruled by the backbone contribution. It follows that a description of the molecule-environment interaction effects based on a nearly-homogeneous perturbing field approximation fails in accurately reproducing the effects of the environment on the side chain electronic properties. The aforementioned factorization of the molecule permitted to overcome the problems arising from the presence of the strong backbone dipole moment, but at the price of using a strongly reduced model system.

Moreover, it is also worth noting that a zwitterionic molecular system is highly interacting with a polar environment, like water. Thus, the process of solvation of these molecular species from a vacuum condition is expected to strongly perturb the QM properties of the system. From our data it emerged that for the present case the standard PMM procedure is not able to reproduce such effects. Conversely, when the starting point is an eigenstate basis set already accounting for the electrostatic effects of the environment, the electronic states quantum mixing due to the fluctuation of the perturbing electric field can be well reproduced by the PMM method even through the treatment of the perturbing operator in the dipolar approximation.

In conclusion, our data suggest that for small molecular systems interacting with the environment, all the proposed strategies and expansions based on the PMM method provide equivalently reliable tools for characterizing the effects of the embedding environment on the QM properties of the QC. Hence, in this context, the PMM is a reliable alternative that makes worthless the computational effort of performing QM calculations on the "full" MD sampling. Conversely, the standard PMM strategy becomes less accurate when dealing with large, flexible systems with non-homogeneous charge distributions. This is the case whenever the QC contains polar (or even charged) moieties, which strongly interact with polar solvents, but only marginally affect the main spectroscopic properties of the chromophore. An effective route to solve the problem, while remaining within the standard PMM context, can be based on the use of a reduced QC model, excluding those atoms that do not perturb the electronic properties of interest. However, when this factorization is not possible (or intentionally avoided), the use of the proposed PMM strategy merged with the ONIOM/EE scheme can be the method of choice in place of more conventional QM/MM methods, which require an extremely high number of QM calculations.

4 Concluding remarks

Several QM/MM approaches to compute electronic absorption spectra in condensed phases have been analyzed.

The common basis of the methods we applied is the partitioning of the system of interest into fragments leading to a portion of the system treated at the QM level that interacts with the remainder described at the MM level. In order to apply each method, a number of configurations of either the QM level fragment or the complete system is extracted from the MM simulations and employed in the subsequent computation of the spectrum. Starting from the MD simulations of systems of increasing complexity we applied multi-scale models based on different approximations to analyze the conditions under which each method performs at

its best. Both the theoretical foundation and the computational feasibility were taken into account, thus testing the limits of the accuracy/computational cost ratio.

The first method employed to describe the interaction between the QM level fragment and its environment exploits a polarizable continuum description of the latter (see section 3.1). This type of procedures, disregarding the solvent atomistic details, gives inaccurate results especially when strong solute-solvent interactions occur[62,63]. For this reason, the method was applied only to the very simple case of a semi-rigid solute (Uracil) in DMF (that is, we did not use it for the aqueous systems). Even in this case, this approach gave the worst results in comparison with the other methodologies. Next, the electronic absorption spectrum of the system was simulated also by applying i) the PMM approach, by following the standard procedure developed to treat semi-rigid quantum dyes and ii) the ONIOM procedure coupled with the Electronic Embedding scheme (ONIOM/EE). According to the first procedure the interaction between the chromophore and the solvent is described through a perturbative approach (see Methods section) making use of a single QM calculation to provide results that show no loss of accuracy if compared with the theoretically more accurate ONIOM/EE approach. These results allow one to envisage that the level of description provided by the (standard) PMM approach is sufficient to accurately describe the spectroscopic features of semi-rigid solutes in aprotic solvents. This makes worthless the computational effort required to apply the ONIOM/EE procedure (we employed hundreds of snapshots for our ONIOM/EE spectrum calculation).

All the previous considerations remain valid even when taking into account a slightly more complex system such as Uracil in aqueous solution. The PMM procedure, exploiting only one Uracil configuration for the QM calculations (either in vacuum or in a polarizable continuum) and the complete MM trajectory to evaluate the instantaneous electric field acting on it, is the method of choice in order to minimize the computational effort without any significant accuracy reduction. We also developed an hybrid procedure to merge the main strengths of both the PMM and the ONIOM/EE procedures to treat aqueous solutions.

While according to the standard PMM procedure the quantum properties of the reference fragment are computed in vacuum, in the latter case the same properties are computed explicitly accounting for the effect of the embedding environment via the EE scheme of the ONIOM method. The effects of the fluctuation of the electric field acting on the QM level fragment are then included in the QM property estimation via the PMM approach (see section 2.1.1). A delicate issue of this strategy is the choice of the reference configuration(/s) to be used for the ONIOM/EE calculations. Indeed, differently from the vacuum condition, both the solute and the solvent instantaneous configurations determine the corresponding QM properties. For this reason, it is troublesome, if not impossible, to identify a standard procedure suitable for all the systems one wants to inspect. Depending on the specific system, different clustering procedures can be applied in order to analyze relevant solute or solvent properties sampled during the MM trajectory, on which properly rationalized configuration(/s) extraction can be done. Since we were studying a semi-rigid chromophore in solution, we focused on the electric field generated by the environment on the QM fragment and the (classical) probability of Hydrogen Bond formation, thus focusing on the main factors governing the interaction between the fragments. As a matter of fact, application of this strategy, in the two different versions, gave reliable and robust results in agreement with those obtained through the other strategies employed (i.e. full ONIOM/EE calculation employing hundreds of QM calculations and the standard PMM procedure). At the same time, the computational advantage of PMM is preserved, since we performed at most 4 QM calculations.

Subsequently, we proceeded to study the absorption spectrum of a flexible and more complex solute in aqueous solution (namely, the Tyrosine amino acid in its zwitterionic form). We applied the ONIOM/EE procedure utilizing 500 configurations sequentially extracted from the MM sampling. In addition, the standard PMM strategy was applied *via* the procedure previously developed to deal with flexible chromophores. Accordingly, one starts by evaluating the properties of the QM fragment in vacuum for a number of relevant configurations

selected from the classical MD sampling by means of clustering and dimensionality reduction techniques. Next, by interpolating these data, a "trajectory" of QM properties is produced. However, this protocol is based on the reduction of the dimensionality of the configurational space explored by the chromophore from 3xN (with N the number of atoms of the fragment) to 2, which is likely to be a too rough approximation to treat highly flexible molecules. In fact, this procedure provided very satisfactory results[18,21] when applied to systems less flexible than Tyrosine (which is characterized by 5 dihedral angles). On the contrary, it failed in reproducing the spectroscopic features of the tyrosine zwitterion in aqueous solution. It is also noteworthy that the solute under investigation is characterized by a strong charge separation localized on a group of atoms of the molecule (the backbone of the amino acid), which is expected to be engaged in strong interactions with water molecules. It follows that the interaction energy, being defined through the electrostatic interaction between the QM fragment electric dipole and the external electric field, is expected to be mainly driven by the strong dipole induced by the localized charge separation. However, comparing the present results with the ones in reference[21], it emerges that the electronic absorption spectrum we want to reproduce can be well approximated by treating at the QM level only the amino acid side chain. This implies that a proper description of the interaction of this moiety with the external electric field is mandatory to reproduce the perturbing effects on the electronic properties under investigation. Irrespective of the role of the above two issues in determining the failure of the standard approach, the new hybrid PMM/ONIOM-EE strategy was fully successful in solving the problem. As previously tested for Uracil in aqueous solution, we explicitly accounted for the presence of the environment point charges by means of the ONIOM/EE scheme applied on a number of configurations of the system, and then we treated the fluctuations of the electric field with respect to these reference structures by the PMM approach. To this end, we divided the complete MD trajectory into 4 sub-trajectories so that each of these represents a configurational basin of the amino acid within which the chromophore could be fairly considered as a semi-rigid molecule. Next, each sub-trajectory

was analyzed focusing on the electric field generated by the environment on the Tyrosine to identify the corresponding configuration to be used in the QM calculations. The obtained results show that the use of only 4 QM calculations performed at the ONIOM/EE level combined with the PMM treatment of the fluctuations of the perturbing electric field allowed us to reproduce the system absorption spectrum with the same accuracy as the one obtained by means of hundreds of QM calculations employed when the full ONIOM/EE treatment is followed.

In summary, the current work aims at providing a reliable comparative analyses of different methods for computing electronic absorption spectra in condensed phases by means of multi-scale approaches. The present study is not exhaustive in terms of the number of showcase systems studied or of available protocols and algorithms (e.g., clustering methods) that can be applied for analyzing the system features. Nevertheless, our study provides guiding insights for users suggesting the choice and application of suitable tools to characterize absorption spectra. Special attention has been paid to the application of the very effective and robust PMM procedure suggesting some extensions and generalizations, which can further extend its reliability without any strong increase of computational requirements.

Acknowledgement

This work is financially supported by the "Ministero dell' Istruzione, dell' Università e della Ricerca" MIUR through the PRIN2015 project (Prot. no. 2015XBZ5YA) entitled: toward quantum photovoltaic: ultrafast energy and charge transport in hybrid nanomaterials. The high performance computer facilities of the SMART Laboratory (http://smart.sns.it) are acknowledged for providing computer resources.

References

(1) Morzan, U. N.; Alonso de Armiño, D. J.; Foglia, N. O.; Ramírez, F.; González Lebrero, M. C.; Scherlis, D. A.; Estrin, D. A. Spectroscopy in Complex Environments from QM-MM Simulations. *Chem. Rev.* **2018**, *118*, 4071–4113.

(2) Brunk, E.; Rothlisberger, U. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chem. Rev.* **2015**, *115*, 6217–6263.

(3) Barone, V. The virtual multifrequency spectrometer: a new paradigm for spectroscopy. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2016**, *6*, 86–110.

(4) Grimme, S. *Rev. Comput. Chem.*; John Wiley & Sons, Ltd, 2004; Chapter 3, pp 153–218.

(5) Kirchner, B.; di Dio, P.; Hutter, J. *Multiscale Molecular Methods in Applied Chemistry*; Springer: Netherlands, Dordrecht, 2012; Chapter 4, pp 109–153.

(6) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed* **2009**, *48*, 1198–1229.

(7) Kirchner, B., (Eds.), J. V., Eds. *Multiscale Molecular Methods in Applied Chemistry*; Springer-Verlag Berlin Heidelberg, 2012; Vol. 307.

(8) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678–5796.

(9) Dapprich, S.; Komàromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives1Dedicated to Professor Keiji Morokuma in celebration of his 65th birthday.1. *J. Mol. Struct. (theochem)* **1999**, *461-462*, 1 – 21.

(10) Vreven, T.; Byun, K. S.; Komàromi, I.; Dapprich, S.; Montgomery Jr., J. A.; Morokuma, K.; ; Frisch, M. J. Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J. Chem. Theory Comput.* **2006**, *2*, 815–826.

(11) Clemente, F. R.; Vreven, T.; Frisch, M. J. *Quantum Biochemistry*; John Wiley & Sons, Ltd, 2010; Chapter 2, pp 61–83.

(12) Fox, S. J.; Pittock, C.; Fox, T.; Tautermann, C. S.; Malcolm, N.; Skylaris, C. Electrostatic embedding in large-scale first principles quantum mechanical calculations on biomolecules. *J. Chem. Phys.* **2011**, *135*, 224107.

(13) Barone, V.; Baiardi, A.; Bloino, J. New Developments of a Multifrequency Virtual Spectrometer: Stereo-Electronic, Dynamical, and Environmental Effects on Chiroptical Spectra. *Chirality* **2014**, *26*, 588–600.

(14) Macchiagodena, M.; Frate, G. D.; Brancato, G.; Chandramouli, B.; Mancini, G.; Barone, V. Computational study of the DPAP molecular rotor in various environments: from force field development to molecular dynamics simulations and spectroscopic calculations. *Phys. Chem. Chem. Phys.* **2017**, *19*, 30590–30602.

(15) Pagliai, M.; Mancini, G.; Carnimeo, I.; De Mitri, N.; Barone, V. Electronic absorption spectra of pyridine and nicotine in aqueous solution with a combined molecular dynamics and polarizable QM/MM approach. *J. Comput. Chem.* **2017**, *38*, 319–335.

(16) Barone, V.; Bloino, J.; Monti, S.; Pedone, A.; Prampolini, G. Theoretical multilevel approach for studying the photophysical properties of organic dyes in solution. *Phys. Chem. Chem. Phys.* **2010**, *12*, 10550–10561.

(17) Aschi, M.; Spezia, R.; Di Nola, A.; Amadei, A. A first-principles method to model perturbed electronic wavefunctions: the effect of an external homogeneous electric field. *Chem. Phys. Lett.* **2001**, *344*, 374–380.

(18) Carrillo-Parramon, O.; Del Galdo, S.; Aschi, M.; Mancini, G.; Amadei, A.; Barone, V. Flexible and Comprehensive Implementation of MD-PMM Approach in a General and Robust Code. *J. Chem. Theory Comput.* **2017**, *13*, 5506–5514.

(19) Del Galdo, S.; Aschi, M.; Amadei, A. In silico characterization of bimolecular electron transfer reactions: The ferrocene–ferrocenium reaction as a test case. *Int. J. Quantum Chem.* **2016**, *116*, 1723–1730.

(20) Zanetti-Polzi, L.; Del Galdo, S.; Daidone, I.; D'Abramo, M.; Barone, V.; Aschi, M.; Amadei, A. Extending the perturbed matrix method beyond the dipolar approximation: comparison of different levels of theory. *Phys. Chem. Chem. Phys.* **2018**, *20*, 24369–24378.

(21) Del Galdo, S.; Mancini, G.; Daidone, I.; Zanetti Polzi, L.; Amadei, A.; Barone, V. Tyrosine absorption spectroscopy: Backbone protonation effects on the side chain electronic properties. *J. Comput. Chem.* **2018**, *39*, 1747–1756.

(22) Mennucci, B. Polarizable continuum model. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 386–404.

(23) Mennucci, B., Cammi, R., Eds. *Continuum Solvation Models in Chemical Physics: From Theory to Applications*; John Wiley & Sons, Ltd, 2007.

(24) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. New developments in the polarizable continuum model for quantum mechanical and classical calculations on molecules in solution. *J. Chem. Phys.* **2002**, *117*, 43–54.

(25) Svezia, R.; Aschi, M.; Di Nola, A.; Amadei, A. Extension of the perturbed matrix method: application to a water molecule. *Chem. Phys. Lett.* **2002**, *365*, 450–456.

(26) Amadei, A.; D'Alessandro, M.; D'Abramo, M.; Aschi, M. Theoretical characterization of electronic states in interacting chemical systems. *J. Chem. Phys.* **2009**, *130*, 084109.

(27) Amadei, A.; Linssen, B. M.; Berendsen, H. J. C. Essential dynamics of proteins. *Proteins* **1993**, *17*, 412.

(28) Daidone, I.; Amadei, A. Essential dynamics: foundation and applications. *WIREs Comput. Mol. Sci.* **2012**, *2*, 762.

(29) Rodriguez, A.; Laio, A. Clustering by fast search and find of density peaks. *Science* **2014**, *344*, 1492–1496.

(30) Shepard, D. A two-dimensional interpolation function for irregularly-spaced data. *Proc. 23rd Nat. Conf. ACM* **1968**, 517–523.

(31) Franke, R. Scattered data interpolation: tests of some methods. *Math. Comp.* **1982**, *38*, 181–199.

(32) Renka, R. Algorithm 790: CSHEP2D: cubic Shepard method for bivariate interpolation of scattered data. *ACM Trans. Math. Soft.* **1999**, *25*, 70–73.

(33) Pagliai, M.; Cardini, G.; Righini, R.; Schettino, V. Car-Parrinello molecular dynamics on the S_N2 reaction Cl^-+CH_3Br in water. *J. Mol. Struct. (Theochem)* **2003**, *630*, 141–149.

(34) Pagliai, M.; Raugei, S.; Cardini, G.; Schettino, V. Hydrogen bond dynamics in liquid methanol. *J. Chem. Phys.* **2003**, *119*, 6655.

(35) Ketchen, D. J.; Shook, C. L. The application of cluster analysis in strategic management research: an analysis and critique. *Strategic Management Journal* **1996**, *17*, 441–458.

(36) Abramyan, M. T.; Snyder, J. A.; Thyparambil, A. A.; Stuart, J. S.; Latour, R. A. Cluster analysis of molecular simulation trajectories for systems where both conformation and orientation of the sampled states are important. *J. Comput. Chem.* **2016**, *37*, 1973–1982.

(37) Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *J. Chem. Theory and Comput.* **2007**, *3*, 2312–2334.

(38) Wolf, A.; Kirschner, K. N. Principal component and clustering analysis on molecular dynamics data of the ribosomal L11·23S subdomain. *J. Mol. Model.* **2013**, *19*, 539–549.

(39) Berendsen, H.; van der Spoel, D.; van Druned, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.

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

(41) Barone, V.; Cacelli, I.; Mitri, N. D.; Licari, D.; Monti, S.; Prampolini, G. Joyce and Ulysses: integrated and user-friendly tools for the parameterization of intramolecular force fields from quantum mechanical data. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3736–3751.

(42) 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. GaussianËIJ16 Revision B.01. 2016; Gaussian Inc. Wallingford CT.

(43) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phy. Rev. B* **1988**, *37*, 785–789.

(44) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(45) Weiner, S.; Kollman, P.; Nguyen, D.; Case, D. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* **1986**, *7*, 230.

(46) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(47) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Chem Phys.* **1993**, *98*, 10089–10092.

(48) Macchiagodena, M.; Mancini, G.; Pagliai, M.; Barone, V. Accurate prediction of bulk properties in hydrogen bonded liquids: amides as case studies. *Phys. Chem. Chem. Phys.* **2016**, *18*, 25342–25354.

(49) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*.

(50) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije., J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(51) Darden, T.; Tork, D.; Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(52) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.

(53) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669–681.

(54) Gross, E.; Dobson, J.; Petersilka, M. Density functional theory of time-dependent phenomena. *Top. Curr. Chem.* **1996**, *181*, 81.

(55) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 5: An extension of Hirshfeld Population Analysis for the accurate description of molecular interactions in gaseous and condensed phases. *J. Chem. Theory Comput.* **2012**, *8*, 527–541.

(56) Gustavsson, T.; Bányász, A.; Lazzarotto, E.; Markovitsi, D.; Scalmani, G.; Frisch, M. J.; V.,; Improta, R. Singlet Excited-State Behavior of Uracil and Thymine in Aqueous Solution: A Combined Experimental and Computational Study of 11 Uracil Derivatives. *J. Am. Chem. Soc.* **2006**, *126*, 607–619.

(57) Creed, D. The photophysics and photochemistry of the near-UV absorbing amino acids-II. Tyrosine and its simple derivatives. *Photochem. and Photobiol.* **1984**, *39*, 563–575.

(58) Fornander, L. H.; Feng, B.; Beke-Somfai, T.; Nordèn, B. UV Transition Moments of Tyrosine. *J. Phys. Chem. B* **2014**, *118*, 9247–9257.

(59) Antosiewicz, J. M.; Shugar, D. UV–Vis spectroscopy of Tyrosine side-groups in studies of protein structure. Part 1: basic principles and properties of tyrosine chromophore. *Biophys. Rev.* **2016**, *8*, 151–161.

(60) Du, H.; Fuh, R. C. A.; Li, J.; Corkan, L. A.; Lindsey, J. S. PhotochemCAD: A computer-aided design and research tool in photochemistry. *Photochem. Photobiol.* **1998**, *68*, 141–142.

(61) Dixon, J. M.; Taniguchi, M.; Lindsay, J. S. PhotochemCAD: A computer-aided design and research tool in photochemistry. *Photochem. Photobiol.* **2005**, *81*, 212–213.

(62) Tomasi, J.; Cammi, R.; Mennucci, B.; Cappelli, C.; Corni, S. Molecular properties in solution described with a continuum solvation model. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5697–5712.

(63) Egidi, F.; Barone, V.; Bloino, J.; Cappelli, C. Toward an Accurate Modeling of Optical Rotation for Solvated Systems: Anharmonic Vibrational Contributions Coupled to the Polarizable Continuum Model. *J. Chem. Theory Comput.* **2012**, *8*, 585–597.

Graphical TOC Entry