

Development, Validation, and Pilot Application of a Generalized Fluctuating Charge Model for Computational Spectroscopy in Solution

Vincenzo Barone,* Ivan Carnimeo, Giordano Mancini, and Marco Pagliai

Cite This: <https://doi.org/10.1021/acsomega.2c01110>

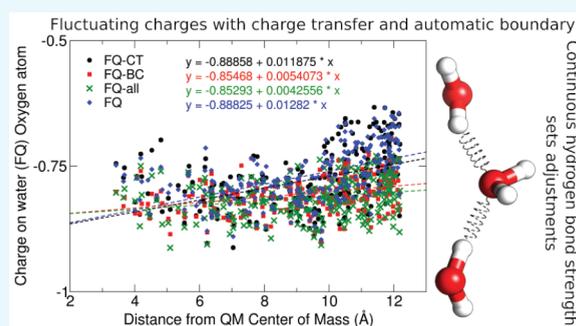
Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: A general approach enforcing nonperiodic boundary conditions for the computation of spectroscopic properties in solution has been improved including an effective description of charge-transfer contributions and coordination number adjustment for explicit solvent molecules. Both contributions are obtained from a continuous description of intermolecular hydrogen bonds, which has been employed also for an effective clustering of molecular dynamics trajectories. Fine tuning of the model has been performed for several water clusters, and then its efficiency and reliability have been demonstrated by computing the absorption spectra of different creatinine tautomers in aqueous solution.



INTRODUCTION

The computation of spectroscopic parameters in solution can be formally separated in two steps. The first step is the generation of a sufficient number of representative structures by means of stochastic simulations (e.g., Monte Carlo or molecular dynamics), which are then used in the second step to compute averaged values of the sought property. While in principle both steps can be performed at the same computational level,¹ the accuracy required in the second step is usually much higher than that needed in the generation of representative snapshots. As a consequence, quite simple force fields are now available for describing solvent dynamics and also solute–solvent interactions can be modeled quite effectively by nonpolarizable force fields or, at most, by quite simple QM-MM approaches. The situation is different for the computation of spectroscopic parameters where solvent polarization can have a non-negligible effect. Polarizable MM and QM/MM approaches can be based on distributed multipoles,² induced dipoles,^{3,4} Drude oscillators,⁵ or fluctuating charges (FQ).⁶ We have followed the FQ route in view of its remarkable efficiency, close connection with the polarizable continuum description of bulk solvent effects, and straightforward improvement by means of virtual sites. In this connection, we have developed in recent years an overall computational protocol enforcing nonperiodic boundary conditions, which usually provides remarkably accurate results.^{7–9} Some not fully satisfactory aspects are, however, still present, which concern a more effective treatment of reaction field effects by bulk solvent together with the lack of charge transfer and proper coordination number for solvent

molecules near the outer boundary. In the present paper, we will be concerned with the inclusion of both contributions in the stage of property evaluation, whereas a companion paper is devoted to the stochastic simulation step. In particular, reaction field effects are treated by the very powerful ddCOSMO model,¹⁰ whereas charge transfer and coordination number saturation are included by a new recipe described in detail in the next section. Then, after validation of the whole procedure for water clusters, the absorption spectra of amine and imine tautomers of creatinine in aqueous solution are used to show the potentialities of the new procedure. In this connection, effective strategies for clustering the snapshots generated by the stochastic simulation are discussed, which allow to strongly reduce the number of expensive QM/MM computations of spectroscopic parameters.

THEORY AND COMPUTATIONAL DETAILS

Theory. In the following, a generalized FQ model will be presented in order to deal with charge transfer and outer boundary effects in polarizable QM/MM calculations. The basic idea behind such developments is that the geometry of the molecular systems under study allows the estimate of the

Received: March 7, 2022

Accepted: March 23, 2022

intermolecular bonding by means of a suitable function (in the present case, the hydrogen bond function proposed by Pagliai et al.¹¹), from which the coordination number of the fragments can be estimated and subsequently used in order to balance (i) the truncation effects occurring at the outer boundary of the cluster and (ii) the charge transfer effects occurring between the fragments.

The first problem (i) has been solved by endowing each atom with a reference charge (q^0), whereas the second issue (ii) has been faced by imposing that the total charge on each fragment (Q) depends on the geometry of the system. Noted is that the charge transfer effects have been included only in the subsystem treated with classical approaches, while no charge transfer has been taken into account across the QM and MM boundaries.

Energy and Charges. The variation of the energy of an isolated atom with respect to its value for a reference charge q_0 is approximated by a Taylor series truncated at the second order:¹²

$$\begin{aligned} \epsilon(q) &= \epsilon(q^0) + \left. \frac{\partial \epsilon}{\partial q} \right|_{q=q^0} (q - q^0) + \frac{1}{2} \left. \frac{\partial^2 \epsilon}{\partial q^2} \right|_{q=q^0} (q - q^0)^2 \\ &= \epsilon_0 + \chi(q - q^0) + \frac{1}{2} \eta (q - q^0)^2 \end{aligned} \quad (1)$$

The total electrostatic energy (\mathcal{E}) of a system of N charges can be partitioned into self-energy ($\mathcal{E}_{\text{self}}$) and interaction (\mathcal{E}_{int}) contributions,^{12,13} with the former term being just the sum of atomic energies defined according to eq 1 and the latter term being an effective charge–charge interaction:

$$\mathcal{E} = \mathcal{E}_{\text{self}} + \mathcal{E}_{\text{int}} = \sum_i \epsilon(q_i) + \sum_{i>j} q_i J_{ij}(R_{ij}) q_j \quad (2)$$

The interaction potential $J_{ij}(R_{ij})$ is a function of the module of the interatomic distance $R_{ij} = |\vec{R}_i - \vec{R}_j|$, depending on some atomic parameters of the i th and j th atoms.

By substituting eq 1 in the last equation, we obtain

$$\begin{aligned} \mathcal{E} &= \sum_i \epsilon_i^0 + \sum_i \chi_i (q_i - q_i^0) + \frac{1}{2} \sum_i \eta_i (q_i - q_i^0)^2 \\ &\quad + \sum_{i>j} q_i J_{ij} q_j \end{aligned} \quad (3)$$

Next, we impose that the functions $J_{ij}(R_{ij})$ satisfy the condition:

$$\lim_{R_{ij} \rightarrow 0} J_{ij}(R_{ij}) = J_{ii} = \eta_i \quad (4)$$

employing the Ohno functional form.¹⁴ Then expansion of $(q_i - q_i^0)^2$ leads to the following expression for the energy:

$$\begin{aligned} \mathcal{E} &= \mathcal{E}_0 + \sum_i \chi_i (q_i - q_i^0) + \frac{1}{2} \sum_{ij} q_i J_{ij} q_j + \frac{1}{2} \sum_i \eta_i (q_i^0)^2 \\ &\quad - \sum_i \eta_i q_i q_i^0 \end{aligned} \quad (5)$$

\mathcal{E}_0 is just the sum of the energies ϵ_i^0 related to the q^0 charges and can be set to zero by choosing the proper reference for the energy. The functional in eq 5 can be viewed as a generalization of the previously implemented FQ method,¹³ which can be obtained in the limit of $q_i^0 \rightarrow 0$ for all atoms.

The set of FQ charges \mathbf{q} can be obtained by a constrained minimization of eq 5 with respect to the charges, imposing the constraint of the charge conservation on each fragment (molecule). In order to do that, it is convenient to split the i index running over the atoms ($i = 1, \dots, N$), into two indices, one (I) running over the fragments ($I = 1, \dots, N_{\text{frag}}$) and the other (l) running over the atoms within the I th fragment ($l = 1, \dots, N_l$),

$$i \rightarrow (I, l) \quad (6)$$

so that a generic summation over i becomes

$$\sum_i = \sum_I \sum_{l \in I} \quad (7)$$

The charge conservation constraint on each fragment can be easily written with this formalism as

$$\sum_{l \in I} q_{l,i} = Q_I, \quad \forall I \quad (8)$$

where Q_I is the total charge on the fragment I . One possibility is that the Q_I 's are assigned a priori at the beginning of the calculation and then kept fixed (e.g., by imposing that all the molecules stay neutral or have a fixed predefined charge), but, more generally, the fragment charges can be a function of the geometry of the system ($Q_I(\mathbf{R})$) as we will show in the next sections.

The Lagrangian \mathcal{L} associated with \mathcal{E} is then defined as

$$\begin{aligned} \mathcal{L} &= \sum_i \chi_i (q_i - q_i^0) + \frac{1}{2} \sum_{ij} q_i J_{ij} q_j + \frac{1}{2} \sum_i \eta_i (q_i^0)^2 \\ &\quad - \sum_i \eta_i q_i q_i^0 + \sum_I \lambda_I \left(\sum_{l \in I} q_{l,i} - Q_I \right) \end{aligned} \quad (9)$$

The derivatives of \mathcal{L} with respect to the variational parameters (charges and Lagrange multipliers) are then:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial q_k} &= \chi_k + \sum_i J_{ki} q_i - \eta_k q_k^0 + \lambda_K \\ \frac{\partial \mathcal{L}}{\partial \lambda_K} &= \sum_{\kappa \in K} q_{\kappa, \kappa} - Q_K \end{aligned} \quad (10)$$

where the index K runs over the fragments and κ over the atoms belonging to each fragment. Imposing the minimum condition for \mathcal{L} , the resulting set of equations is

$$\begin{aligned} \sum_i J_{ki} q_i + \lambda_K &= -\chi_k + \eta_k q_k^0 \\ \sum_{\kappa \in K} q_{\kappa, \kappa} &= Q_K \end{aligned} \quad (11)$$

This problem can be recast by using the formalism of ref 13, so that the charges can be obtained by solving

$$\begin{pmatrix} \mathbf{J} & \mathbf{1}_\lambda \\ \mathbf{1}_\lambda & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{q} \\ \lambda \end{pmatrix} = \begin{pmatrix} -\boldsymbol{\chi} + \boldsymbol{\eta} \circ \mathbf{q}^0 \\ \mathbf{Q} \end{pmatrix} \quad (12)$$

where in the right-hand side the atomic electronegativities ($\boldsymbol{\chi}$), the total charge on each fragment (\mathbf{Q}), and the vector obtained by the Hadamard product of the arrays containing the atomic hardness and reference charge ($\boldsymbol{\eta} \circ \mathbf{q}^0$) are included.

Whenever the FQ system interacts with a QM charge density, the electrostatic interaction between the charges and the QM charge density must be added to eq 5. In terms of \mathbf{D} , the one-particle QM density matrix in the atomic orbital (AO) representation, the electrostatic potential V_k at the coordinates \vec{R}_k of the k th atom is

$$V_k(\mathbf{D}) = \sum_{A \in \text{QM}} Z_A(A|k) - \sum_{\mu\nu\sigma \in \text{AO}} D_{\mu\nu\sigma}(\mu\nu\sigma|k) \quad (13)$$

where Z_A is the atomic number of the atom A of the QM subsystem, the μ, ν indices run over the atomic orbital basis functions $\{\chi_\mu(\vec{r}), \mu = 1, \dots, \text{AO}\}$, the σ index runs over the spin states (α and β), and the integrals are defined as

$$(\mu\nu\sigma|k) = \int d\vec{r} d\vec{r}' \frac{\chi_{\mu\sigma}(\vec{r}) \chi_{\nu\sigma}(\vec{r}') \delta(\vec{r}' - \vec{R}_k)}{|\vec{r} - \vec{r}'|} = V_{\mu\nu\sigma,k} \quad (14)$$

$$(A|k) = \int d\vec{r} d\vec{r}' \frac{\delta(\vec{r} - \vec{R}_A) \delta(\vec{r}' - \vec{R}_k)}{|\vec{r} - \vec{r}'|} \quad (15)$$

The FQ charges are obtained by solving

$$\begin{pmatrix} \mathbf{J} & \mathbf{1}_\lambda \\ \mathbf{1}_\lambda & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{q} \\ \lambda \end{pmatrix} = \begin{pmatrix} -\chi - \mathbf{V} + \eta \circ \mathbf{q}^0 \\ \mathbf{Q} \end{pmatrix} \quad (16)$$

A Strength Function for Hydrogen Bonds. Now, restricting the treatment to the MM subsystem only, we want to define a function of the geometry of the system which is able to count the intermolecular interactions occurring among the FQ fragments, and estimate the "strength" of such interactions.

Hence, indicating with $\{\vec{R}_k, k = 1, \dots, N\}$ the position vectors of the FQ atoms, for each pair of k and l atoms, the function $f(R_{kl}, \vartheta_{klm})$ can be defined, where $R_{kl} = |\vec{R}_k - \vec{R}_l|$ and ϑ_{klm} is the valence angle involving a third atom. As detailed in the original work,¹¹ such a function of two variables is defined as a product of two functions of one variable, and gives values in the interval $f: \mathbb{R}^3 \rightarrow [0, 1] \in \mathbb{R}$

$$f(R_{kl}, \vartheta_{klm}) = A(R_{kl}) \cdot B(\vartheta_{klm})$$

$$A(R_{kl}) = \begin{cases} \exp\left[-\frac{(R_{kl} - R_0)^2}{2\sigma_R^2}\right] & R_{kl} > R_0 \\ 1 & R_{kl} \leq R_0 \end{cases}$$

$$B(\vartheta_{klm}) = \begin{cases} \exp\left[-\frac{(\vartheta_{klm} - \vartheta_0)^2}{2\sigma_\vartheta^2}\right] & \vartheta_{klm} > \vartheta_0 \\ 1 & \vartheta_{klm} \leq \vartheta_0 \end{cases} \quad (17)$$

where the empirical parameters are $R_0 = 1.85 \text{ \AA}$, $\sigma_R = 0.20$, $\vartheta = 10^\circ$, and $\sigma = 10^\circ$.

The function provides a continuous estimate of the relative strength of hydrogen bonds, which vanishes, for a given pair of atoms, in the absence of any interaction and goes to the asymptotic value of 1 when the hydrogen bond reaches its maximum strength.

The dependence on ϑ can be better understood by looking at Figure 1, where a trimer of water molecules is shown.

Oxygen 1 is involved in two hydrogen bonds with atoms 5 and 9. For example, the strength of the first one depends on

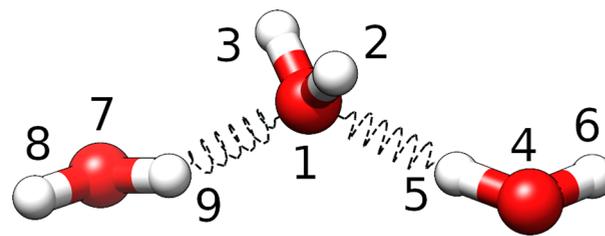


Figure 1. Three hydrogen bonded water molecules.

the distances R_{15} through $A(R_{15})$ and on the position of atom 4—the first bonded neighbor of atom 5—through the function $B(\vartheta_{514})$, so that the strength function is $f(R_{15}, \vartheta_{514})$.

The $f(R_{kl}, \vartheta_{klm})$ functions are symmetric with respect to the permutation of k and l indices, and once they are evaluated for all the intermolecular atomic pairs, the result is a very sparse lower-triangular matrix (\mathbf{F}) containing all the H-bond strengths. For the system in Figure 1, assuming that the two hydrogen bonds have strengths of, say, 0.8 and 0.7, the \mathbf{F} matrix has the following nonvanishing elements: $F(1,5) = 0.8$ and $F(1,9) = 0.7$.

A Model for Outer Boundary Effects from the Strength Function Analysis. As shown in a previous work,¹⁵ the FQ atoms located at the edge of the system suffer from truncated connectivity, resulting from nonphysical coordination numbers. Such a problem was solved by endowing the more external water molecules with fixed charges (FX) obtained from the TIP3P-FB model.¹⁶

While this approach was very effective to correctly reproduce the charge distribution of the FQ subsystem, it can become unpractical when it has to be applied to many snapshots, since it requires the manual identification of the atoms close to the outer boundary and the reassignment to them of the sought fixed charges. A more viable solution can be based on the observation that for systems whose primary intermolecular interactions are hydrogen bonds (e.g., aqueous solutions), the number of hydrogen bonds formed by each molecule can be used to define an effective and continuous coordination number and the strength function can be employed to automatically detect the atoms located at the edge of the molecular system which are affected by boundary effects. In particular, we defined a threshold value of 0.01 (1%) for the hydrogen bond function $f(R_{kl}, \vartheta_{klm})$ above which a (possibly weak) hydrogen bond does occur. Then, for an oxygen atom belonging to a water molecule in the bulk phase, the ideal coordination number is 2, since two lone pairs are available for two intermolecular bonds with the hydrogen atoms of other two water molecules. Analogously, for a hydrogen atom belonging to a bulk water molecule, the ideal coordination number is 1, since it can in principle bind one oxygen from another water molecule. Then, if the coordination number for a given water molecule is smaller than the above values, we assume that such a molecule is located near the outer boundary, so that it can interact with an artificially reduced number of other solvent molecules and the electrostatic potential experienced by such a molecule is wrong. Hence, a correction must be included in order to restore the correct electrostatics.

We decided to face such an issue by increasing the absolute value of the q_0 charge associated with the selected atoms as follows:

$$q_i^0 = \alpha NH_i \quad (18)$$

where NH is a signed adimensional integer array containing the number of hydrogen bonds formed by each solvent atom and α is an empirical parameter that provides the correct physical dimensions. Noted is that the sign of the correction depends on whether the atom is the donor or the acceptor in the hydrogen bond under examination. For example, in the molecular system shown in Figure 1, the integer array NH has dimension 9 and nonvanishing elements $NH(1) = 2$, $NH(5) = NH(9) = -1$, indicating that atom 1 is involved in 2 hydrogen bonds as a donor, atoms 5 and 9 are involved in 1 hydrogen bond each as acceptors, while all the other atoms are not involved in any hydrogen bonding. Thus, all the three water molecules suffer from nonphysical coordination due to boundary effects, and the q^0 charges correcting for such an effect are $q^0(2) = q^0(3) = q^0(6) = q^0(8) = -\alpha$ and $q^0(4) = q^0(7) = 2\alpha$, whereas $q^0(1) = q^0(5) = q^0(9) = 0.0$.

Thus, with such a model, the correction for the boundary effects can be automatically included in the calculations without any user modification of the input file, and this approach is particularly suitable when a large number of snapshots from a MD simulation need be analyzed.

A Charge Transfer Model from the Strength Function Analysis. We assume that, in a system ruled by intermolecular hydrogen bonds, the intermolecular charge transfer on one atom is proportional to the strength of all the hydrogen bonds occurring between that atom and the others. Assigning opposite signs when the atom behaves as donor or acceptor respectively, the total charge (Q_j) on the fragment J is proportional to the sum of all the atomic charge transfers coming from the atoms in that fragment ($l \in I$) and all the others.

The hydrogen bond function can be used to define effective charges over the fragments (\mathbf{Q}), to be included into the charge constraints of eq 16, thus providing an estimate of the charge transfer between the fragments. Since the original function $f(R_{kl}, \vartheta_{klm})$ takes values between 0 and 1, a simple approximation is to employ a single scaling factor to obtain fragment charges \mathbf{Q} :

$$Q_j = \beta \sum_{j \in J} s_j \sum_{I \neq J} \sum_{i \in I} f_{ij}(R_{ij}, \vartheta_{ijk}) \quad (19)$$

where s_j is a sign function with values +1 and -1 for donor and acceptor atoms, respectively, whereas a is the scaling factor, which allows us to go from the adimensional values of f to the charges, providing also the correct physical dimensions. For the molecular system of Figure 1, the \mathbf{Q} array is

$$\mathbf{Q} = \beta \begin{pmatrix} +1.5 \\ -0.8 \\ -0.7 \end{pmatrix} \quad (20)$$

suggesting that an escaped charge proportional to +1.5 is found on atom 1, which is compensated by charges of -0.8 on atoms 5 and -0.7 on atom 9. Of course, the convention adopted for the signs enforces the conservation of the total charge of the system.

It is worth noting that such a scheme is able to take into account the CT among the FQ fragments, but it cannot

account for any correction at the QM-FQ boundary, due to the intrinsic limitations of the QM/MM scheme.

QM/MM Energy. When the classical system is the low level part of a QM/MM calculation, the total energy is^{4,13,15,17–20}

$$E_{\text{tot}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{int}} \quad (21)$$

where the functional form of E_{QM} depends on the specific choice of the QM method employed for the treatment of the high level region, while E_{MM} contains all the intra- and intermolecular terms of the MM force field of the low-level region. When the QM and MM subsystems are not covalently bonded and charge transfer effects between the two regions are neglected, the interaction energy term E_{int} can be partitioned into electrostatic, polarization, repulsion and dispersion contributions.¹⁵ The first two contributions are treated by the approach sketched in the previous sections, whereas the latter two terms can be approximated by simple Lennard–Jones (LJ) functions, so that the QM/MM interaction energy can be written as

$$E_{\text{int}} = E_{\text{LJ}} + \mathbf{qV} \quad (22)$$

where E_{LJ} includes the dispersion-repulsion energy contributions between the QM and MM subsystems, while the second term is the electrostatic interaction between the QM and MM subsystems. In order to focus on the new developments of the FQ model, it is convenient to drop out from E_{MM} the electrostatic interaction between the charges of the MM subsystem, so that E_{tot} reads

$$E_{\text{tot}} = E_{\text{QM}} + E_{\text{MM}}^* + E_{\text{LJ}} + \chi(\mathbf{q} - \mathbf{q}^0) + \frac{1}{2} \mathbf{qVq} + \frac{1}{2} \mathbf{q}^0 \cdot (\boldsymbol{\eta} \circ \mathbf{q}^0) - \mathbf{q} \cdot (\boldsymbol{\eta} \circ \mathbf{q}^0) + \mathbf{qV(D)} \quad (23)$$

where E_{MM}^* includes all the bonded (stretching, bending, torsion...) and nonbonded interactions between the MM atoms, with the exception of the electrostatic interaction.

Since the q^0 charges do not enter the QM/MM electrostatic interaction energy, the Fock and coupling matrix elements related to eq 23 are analogous to the ones of the FQ model previously implemented,^{15,18} with the obvious modification that the FQ charges are obtained from eq 16.

Although not used in the present paper, we implemented also first derivatives with respect to nuclear coordinates. Since the q_0 charges are onsite corrections, they do not introduce any change in the gradient formulas, but this is not the case for the \mathbf{Q} vector. In fact, eq 9 is variational with respect to \mathbf{q} and λ but not with respect to the \mathbf{Q} . As long as the \mathbf{Q} are just fixed numbers, we can assume that

$$\frac{\partial \mathcal{L}}{\partial \mathbf{Q}} = 0 \quad (24)$$

is identically zero, since

$$\frac{\partial \mathbf{Q}}{\partial \mathbf{R}} = 0 \quad (25)$$

However, this is no longer true when $\mathbf{Q} = \mathbf{Q}(\mathbf{R})$ and a new term appears in the total gradient, which for the charge transfer contribution to the x-component of the force on a generic atom l , reads:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{Q}} \frac{\partial \mathbf{Q}}{\partial x_i} = \sum_j \lambda_j \frac{\partial Q_j}{\partial x_i} \quad (26)$$

The derivatives $\partial Q_j / \partial x_i$ are analytical and can be easily obtained from eqs 17 and 19 (see the Appendix).

Molecular Dynamics Simulations and Spectra Computation. The TIP3P-FB¹⁶ rigid water molecule was used in all the simulations, whereas the topology and initial GAFF2²¹ parameters of creatinine have been obtained with the PrimaDORAC²² web interface. The geometries of amine and imine tautomers were optimized at the B3LYP/6-31+G(d) level taking into account bulk solvent effects by means of the polarizable continuum model (PCM),²³ and point atomic charges were assigned performing a CMS population analysis²⁴ of the corresponding Kohn–Sham orbitals, following the protocol adopted with success for the study of several other systems.^{8,25–28} Molecular structures, atomic labels, and charges for the two tautomers are shown Figure 2.

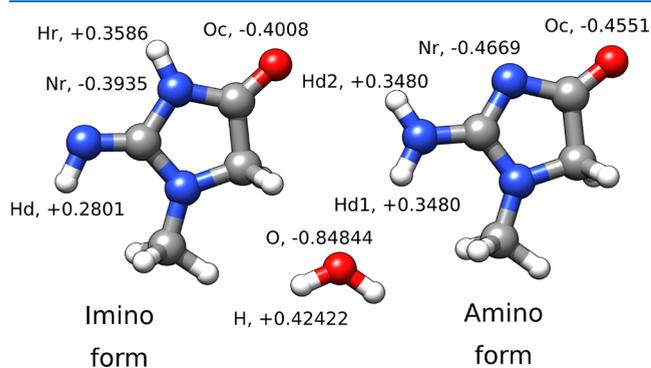


Figure 2. Structure of the imine and amine tautomers of creatinine with labels and atomic charges of atoms involved in hydrogen bonds. A TIP3P-FB water molecule with its corresponding atomic charges is also shown for the sake of completeness.

All the simulations were performed employing the Proxima library²⁹ to build spherical boxes filled with water molecules at the experimental density and, in the case of creatinine aqueous solution, placing a creatinine molecule at the center of the box and deleting all molecules within the sum of covalent radii from any creatinine atom. In order to avoid any convergence issue, MD simulations were performed using large cavities with radii of 20 and 21 Å (containing 1116 and 1170 water molecules) for pure water and creatinine solution, respectively. Then, QM/MM computations of charges and, possibly, spectra were performed for the four polarizable models taking the water molecules occupying the innermost 12 Å of the spherical cavities employed in the MD simulations.

Following an established protocol,^{30,31} we carried out MD simulations under spherical nonperiodic boundary conditions (NPBC) for both tautomers employing a locally modified version of the Gaussian³² suite of programs. The RVV1 integrator²⁶ was used in all simulations, with a convergence criterion $\epsilon = 10^{-9}$ for the calculation of quaternion derivatives. The “rough walls”²⁶ boundary condition and the Bussi–Donadio–Parrinello³³ thermostat were used in all the NVT/NPBC simulations. The equilibration of the system involved an initial minimization with the conjugate gradient method and a subsequent simulation for 1000 ps with a small integration step of 0.5 fs and temperature of 298.15 K. The production run was then initiated at 298.15 K and continued for 25 ns with an

integration time step of 4.0 fs. During the simulations the creatinine geometry was kept frozen. Snapshots were saved every 2 ps. Discarding an initial part of both simulations of 1 ns a total sampling of 9 ns was used for all the subsequent analyses.

The calculation of structural properties was carried out with standard procedures, taking into account proper normalization for NPBCs.

The selection of snapshots chosen in order to provide a well converged description of the trajectories was carried out with an in house implementation of the *density peaks* method.³⁴ To estimate the local density we used a Gaussian kernel with $\sigma = 0.02N$, where N is the number of points in the data set. As a feature space for the snapshot selection of the creatinine trajectories we used the F_{HB} hydrogen bond definition as a function of time for the atoms Hr, Oc, Hd, Nd (imine form) and Hd1, Hd2, Oc, Nr (amine form), i.e., a four dimensional feature space.

For the selected snapshots, TD-DFT computations were performed to simulate the UV–vis spectra employing the B3LYP density functional in conjunction with the 6-31+G(d) basis set. Environmental effects were taken into account by QM/MM computations employing four different polarizable methods, which starting from the standard fluctuating charge approach (FQ), include charge-transfer contributions (FQ-CT) or outer boundary corrections (FQ-BC) or both terms (FQ-All).

RESULTS AND DISCUSSION

The χ and η parameters entering eq 5 have been optimized in ref 18 in order to reproduce the dipole moment of an isolated water molecule and the TIP3P-FB charges¹⁶ for bulk water. The same results are obtained in all the FQ formulations for water molecules forming either 0 or 4 hydrogen bonds. The FQ-CT variant approaches the parent FQ model as β tends to zero restoring the original $Q = 0$ charge constraints (see eq 19), whereas it leads to different results for intermediate numbers of hydrogen bonds when $\beta \neq 0$, with $\beta = 0.1$ leading to charge transfers close to the reference CMS values for the six clusters shown in Figure 3. For instance, the charge transfer estimated from CMS charges on top of B3LYP/6-31+G(d) computations for the isolated water dimer is 0.09, which becomes 0.11 when including bulk solvent effects by means of PCM, with both values comparing well with the FQ-CT estimate of 0.1. Coming to the FQ-BC and FQ-All variants, $\alpha = 0.2$ (see eq 18) leads to charges close to the limiting TIP3P-FB values for solvent molecules near the outer boundary.

For purposes of illustration, the oxygen atomic charges of a representative configuration issued from a full MD simulation of pure water (see Figure 4) obtained employing the four different polarizable models (FQ, FQ-CT, FQ-BC, and FQ-All) and treating the central water molecule either at MM or QM levels are shown in Figure 5.

Noted is that MM charges are intrinsically larger than the QM counterparts due to the need of reproducing dipole moments by a three-site point charge model. The results displayed in Figure 5 show that, when present, the central QM molecule mainly perturbs the hydrogen bonded water molecules, while increasing the distance from the central oxygen atom (O1) the QM/MM and MM systems have almost the same atomic charge values.

From a quantitative point of view, the FQ-CT model provides charges slightly smaller, in absolute value, than the

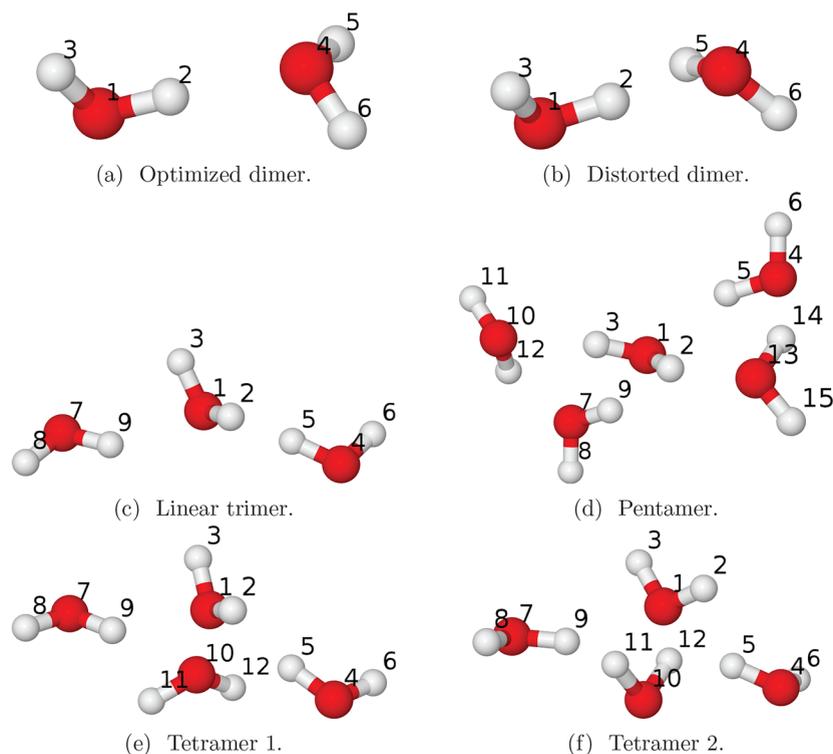


Figure 3. Cluster models with atom labeling.

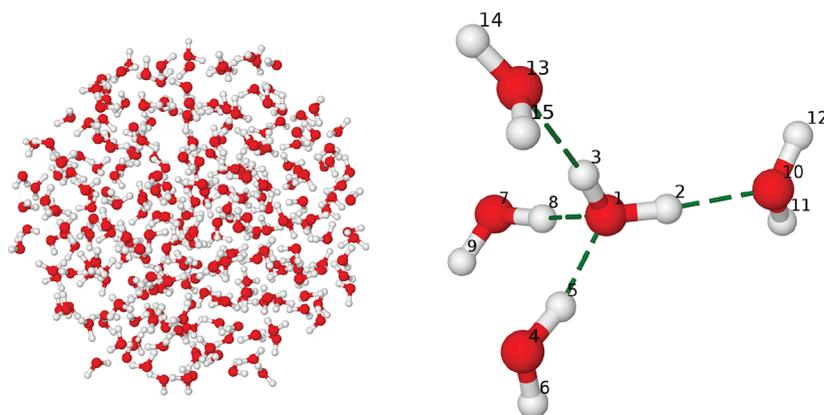


Figure 4. Snapshot of a water MD simulation (left) and atom labeling of the five central water molecules (right).

original FQ model without changing the spread between the smallest and largest value, whereas the FQ-BC model increases both the absolute values and the spread. Finally, the FQ-All model fulfills the sought requirement of decreasing the spread without reducing the smallest value of the FQ charges.

Analogous trends are apparent in Figure 6 for the whole MM droplet.

Creatinine in Aqueous Solution. Hydrogen Bond Structure and Dynamics. The trajectories obtained from molecular dynamics simulations have been analyzed to characterize the interactions of imine and amine tautomers of creatinine with the aqueous environment. The solute–solvent interactions are essentially ruled by the formation of intermolecular hydrogen bonds involving the creatinine atoms indicated in Figure 2.

The first structural information related to the hydrogen bond interaction is retrieved by computing the pair radial distribution functions between creatinine atoms involved in

hydrogen bonds with water molecules for both tautomers (see Figure 7). Although the average interaction between the creatinine tautomers with solvent is similar to that reported in the study by Léon et al.,³⁵ the analysis of molecular dynamics trajectories provides not only structural, but also dynamical information.

Comparison of the $g(r)$ related to the interaction between the carbonyl oxygen of both tautomers with the hydrogen atoms of water ($\text{O}=\text{C}\cdots\text{H}$ in Figure 7) shows that the amine tautomer has stronger interactions with neighboring solvent molecules and slightly higher coordination number ($g(r)$ integral) than the imine tautomer. Also the interactions involving the same hydrogen atom in both tautomers (Hd for the imine and Hd1 for the amine, respectively) are stronger for the amine tautomer. In particular, the Hd atom is the only one which forms a very weak intramolecular hydrogen bond, whereas all the other atoms considered in Figure 7 are involved in interactions with solvent molecules.

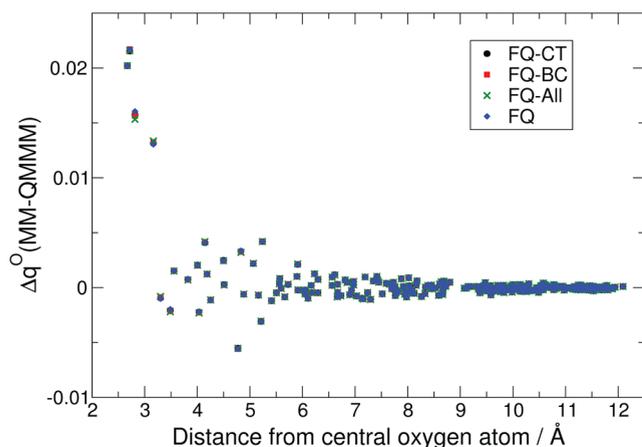


Figure 5. Difference (Δq°) between MM and QM/MM charges on water oxygen atoms.

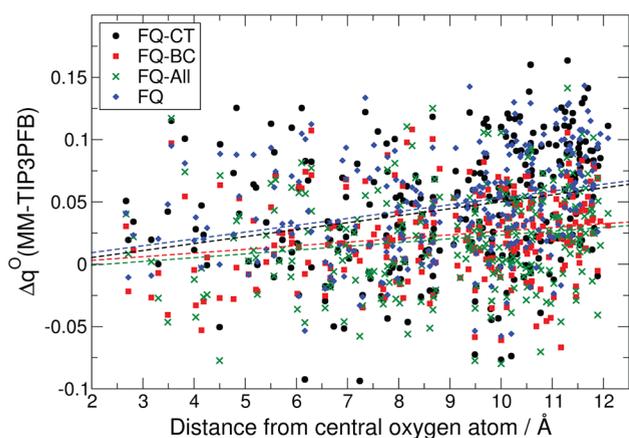


Figure 6. Difference (Δq°) between polarizable and TIP3P-FB charges on water oxygen atoms.

Further information on the strength and dynamics of the hydrogen bond interactions can be obtained by using the hydrogen bond function (F_{HB}) proposed by Pagliai and co-workers^{11,36} and adopted in the present work to determine the amount of charge transfer in both the FQ-CT and FQ-All models. The analysis has been performed separately for the two creatinine tautomers, considering the interactions involving the same atoms for which the radial distribution functions in Figure 7 have been computed.

Figure 8 shows the histogram of the F_{HB} function for the imine tautomer atoms involved in hydrogen bond interactions. The analysis provides additional information with respect to the pair radial distribution functions. In all the histograms, the bar related to $F_{\text{HB}} = 0$ has non-negligible values, indicating that the hydrogen bond does not involve the same molecules but processes of forming and breaking the interaction occur during the simulation. This picture is confirmed by the analysis of F_{HB} as a function of simulation time shown in Figure 9, where the low average value is essentially due to the Hd atom which is only seldom involved in hydrogen bond interactions with solvent molecules (see Figure 8).

The F_{HB} function has been analyzed also for the amine tautomer, showing a stronger overall solute–solvent interaction (see Figure 10). This result suggests a higher stability of the amine tautomer in aqueous solution, in agreement with previous results.³⁵ These conclusions are corroborated by the

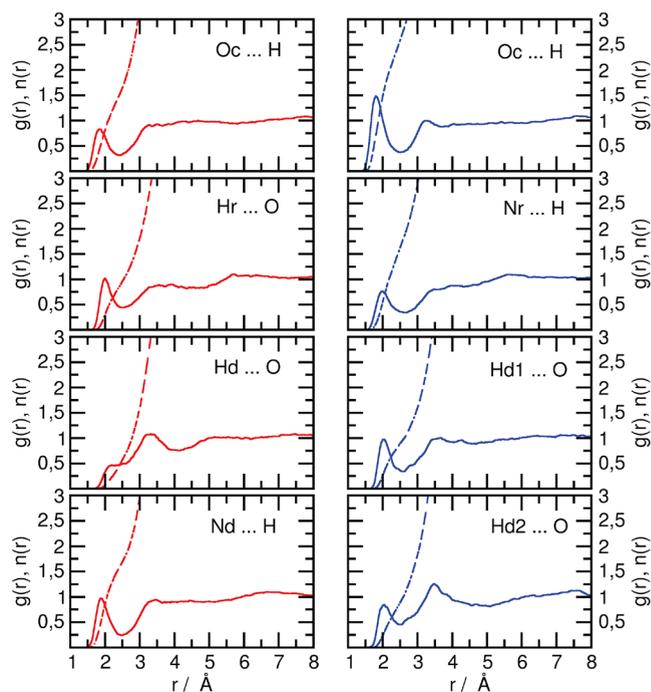


Figure 7. Radial distribution functions related to the creatinine/water interactions. Full and dashed lines refer to radial distribution function, $g(r)$, and coordination number, $n(r)$, respectively. In red and blue are the results for imine and amine tautomers, respectively.

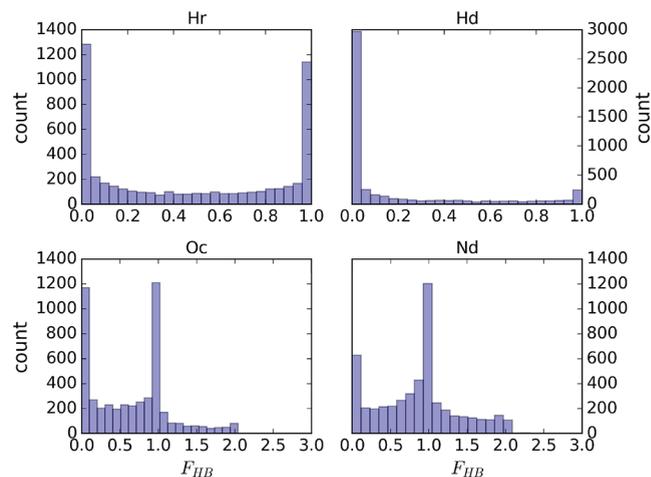


Figure 8. Histogram of the F_{HB} function values for the imine tautomer.

comparison of (i) the time evolution of the F_{HB} function and (ii) the total histograms of Figures 9 and 11. In particular, this behavior can be related to the interactions involving the Oc and Hd1 atoms, which form stronger hydrogen bonds with solvent molecules.

We used the hydrogen bond information also to select the snapshots to be used in the following spectra calculations. A preliminary principal component analysis (PCA)³⁷ of the F_{HB} values was run to check for the correlation between the various sites and to evaluate the possibility of limiting the clustering to a 2D problem. The fraction of total variance recovered by using the first two eigenvectors for the imine and amine tautomer was 67.06% and 66.08%, respectively, showing limited correlation between the four atoms involved in

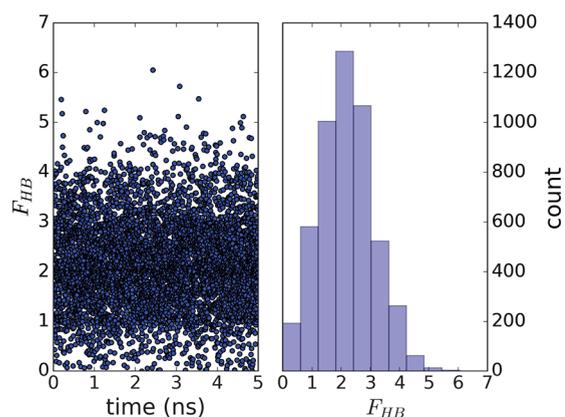


Figure 9. Simulation of the imine tautomer: value of the total F_{HB} function summed for all solute atoms and corresponding histogram. This histogram was used to select frames for the subsequent analysis.

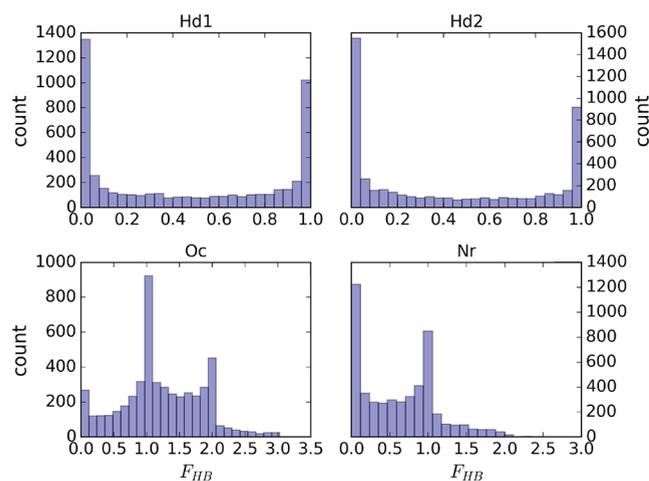


Figure 10. Histogram of the F_{HB} function values for the amine tautomer.

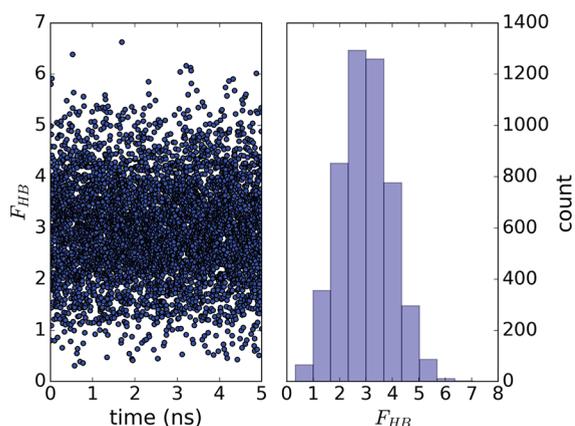


Figure 11. Simulation of the amine tautomer: value of the total F_{HB} function summed for all solute atoms and corresponding histogram. This histogram was used to select frames for the subsequent analysis.

hydrogen bonding in both trajectories. Given the limited amount of features, we discarded the idea of using a 2D space, determining, instead, the number of clusters and their centroids by inspection of a specialized plot (known as *decision graph*), which shows the local density of each point in the data set (ρ) on the x axis and the distance from the nearest high

density point (δ) on the y axis. Cluster centroids are determined as points with a high local density and far away from other high density points i. e. as density peaks. These are the outlier points highlighted in the right top of the graph.

The decision graphs for both tautomers shown in Figure 12 clearly indicate that four and three density peaks can be identified in the two trajectories; the sizes of the corresponding clusters (which can be used to estimate their statistical weight) are shown in Table 1. These clusters correspond to structures of the two tautomers with different values of F_{HB} (see Figures 8–11).

For the imine tautomer (Figure 13), it can be easily observed how the four clusters identify different coordination states for moieties such as the carbonyl oxygen or the Nd-Hd group that can form multiple hydrogen bonds and also the cooperative effect that favors the formation of single bonds for the carbonyl on the Nr-Hr side for all for clusters. An analogous effect can be observed for the amine tautomer in the formation of hydrogen bond networks with the atoms Hd2, Nr, and Ocb (see Figure 14).

Polarization Effects. Figure 15 shows the average charge of the oxygen atom of water molecules as a function of the distance from the creatinine center of mass for the different polarizable models. The analyzed system is made up by 1 QM solute (the imine tautomer of creatinine) and 250 TIP3P-FB water molecules, whose coordinates have been extracted from the molecular dynamics trajectory. Atomic charges have been subsequently obtained a posteriori through QM/MM calculations by modeling water molecules with FQ, FQ-BC, FQ-CT, or FQ-All approaches.

While FQ charges have been successfully applied to determine spectroscopic properties of QM solutes in aqueous solution, they have nonphysical behavior near the outer boundary region, as well evidenced by the significant depolarization occurring for distances longer than 10 Å (see Figure 15). This behavior does not occur by using the FQ-BC approach, which solves the problem of an incorrect number of hydrogen bond interactions for those water molecules located in the boundary region, thus leading to atomic charges similar to those of the TIP3P-FB model (see Figure 15).

Concerning the charge transfer model, Figure 15 shows that the water oxygen charges issuing from the FQ-CT and FQ-All models are quite similar to those provided by the FQ and FQ-BC counterparts, respectively. This is expected for the system under study, where the fragments are relatively weakly bonded (hydrogen bonds only) and each fragment behaves at the same time as donor and acceptor, leading to compensating charge flows in opposite directions.

As a matter of fact, the fragment charges are between -0.05 and 0.05 for more than half the water molecules and between -0.1 and 0.1 for nearly 90% of the water molecules (see Figure 16).

Electronic Absorption Spectra Simulation. The clustering procedure discussed in a previous section permits a strong reduction of the number of snapshots for which electronic spectra must be explicitly computed. Several tests showed that, together with cluster centroids, very few additional snapshots need be considered for obtaining results virtually indistinguishable from the reference results issued from 200 equally spaced snapshots. Although satisfactory results are obtained by adding just a single structure (the most distant from the centroid) for each cluster, we preferred to include about 10 snapshots for each cluster, with the precise number being proportional to its

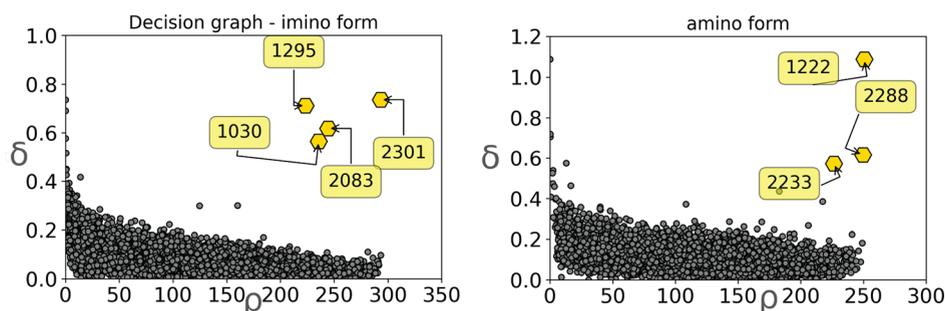


Figure 12. Decision graph for the imine (left) and amine (right) forms. Cluster centroids are indicated by a yellow hexagon and by their frame label.

Table 1. Centroids and Sizes of Clusters Determined for Imine and Amine Simulations

	cluster number			
	1	2	3	4
imine				
centroid at frame	1030	1295	2083	2301
number of frames in the cluster	1068	943	1459	1531
amine				
centroid at frame	1222	2233	2288	
number of frames in the cluster	1891	1443	1667	

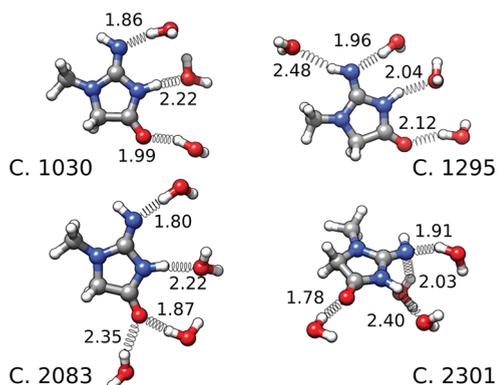


Figure 13. Cluster centroids for the imine tautomer. The structures corresponds to the simulation frames labeled in the decision graph (Figure 12, right panel). First neighbor water molecules of hydrogen bonding sites are shown with their donor (or acceptor) hydrogen distance in angstrom. The centroids are shown left to right following the order of Table 1.

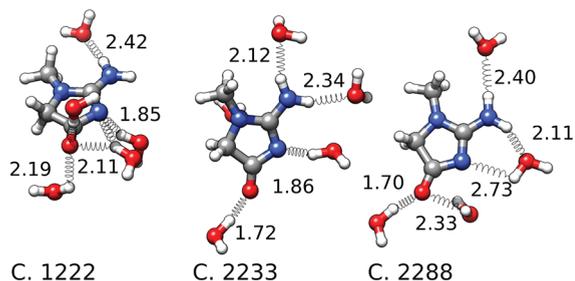


Figure 14. Cluster centroids for the amine tautomer. The structures corresponds to the simulation frames labeled in the decision graph (Figure 12, right panel). First neighbor water molecules of hydrogen bonding sites are shown with their donor (or acceptor) hydrogen distance in angstrom. The centroids are shown left to right following the order of Table 1.

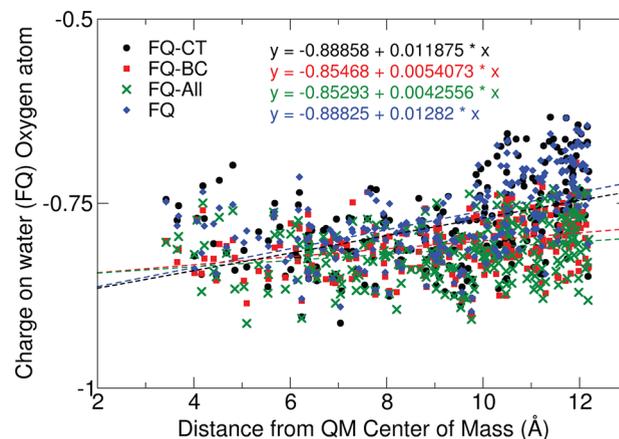


Figure 15. Atomic charges on water oxygen atoms issued from different polarizable models.

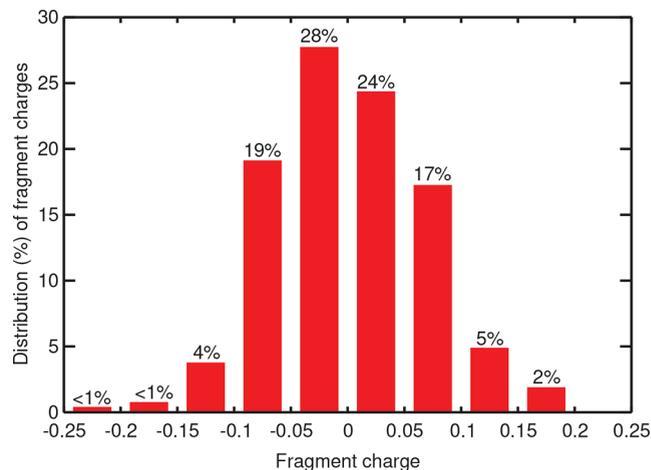


Figure 16. Histogram of total charge on water molecules issued from the FQ-All model.

size (given in Table 1). Fully converged results are obtained when employing 50 frames for each tautomer, and, in particular 11,9,15,15 frames for the 4 clusters of the imine tautomer and 19,14,17 frames for the 3 clusters of the amine tautomer.

The UV–vis spectra have been computed for all those frames (and then averaged) with QM/MM approaches involving fixed or different flavors of polarizable charges.

From the comparison of panels (a) and (b) of Figure 17, it is apparent that the simulated UV–vis spectra for the two tautomers show substantial differences. In particular, only the

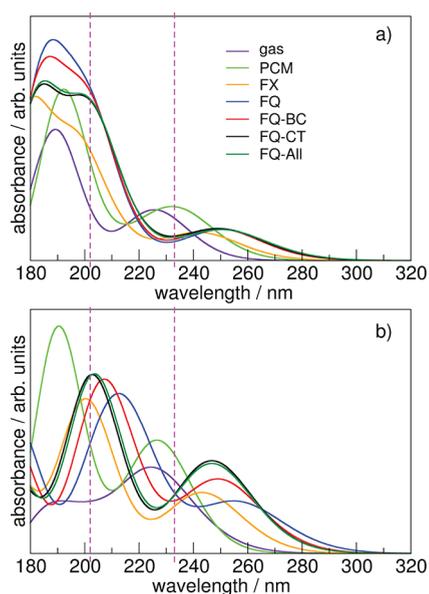


Figure 17. UV–vis spectra of creatinine tautomers: (a) imine and (b) amine tautomers. TD-DFT calculations have been carried out at the B3LYP/6-31+G(d) level for the QM part. Solvent molecules have been described with the FQ, FQ-CT, FQ-BC, and FQ-All polarizable models and with TIP3P-FB fixed charges (FX).

UV–vis spectrum of the amine tautomer is in good agreement with that observed experimentally,³⁸ confirming that this species is the one present in aqueous solution. A quantitative estimate of the agreement with the experimental data can be obtained by the comparison of the wavelengths experimentally observed³⁸ for the two bands that characterize the UV–vis spectrum of creatinine in aqueous solution with the counterparts computed with the different polarizable approaches.

The results collected in Table 2 show that the absorption maxima of amine and imine tautomers are similar in the gas phase and also employing the PCM to take into account bulk solvent effects. On the other hand, all the models based on fluctuating charges displace by more than 10 nm the red-side

Table 2. Wavelength (in nm) of Absorption Maxima and Relative Intensity of the Less Intense Band in the UV–Vis Spectra of Imine and Amine Tautomers of Creatinine^a

	imine tautomer	amine tautomer
gas B3	189, 226 (0.39)	192, 224 (0.64)
gas B2	187, 208	193, 216
PCM	192, 232 (0.31)	190, 227 (0.50)
FX	182, 242 (0.17)	200, 243 (0.39)
FQ	188, 251 (0.14)	213, 255 (0.33)
FQ-BC	187, 249 (0.16)	207, 249 (0.43)
FQ-CT	185, 249 (0.18)	203, 247 (0.52)
FQ-All	185, 250 (0.17)	204, 247 (0.50)
best	183, 232 (0.17)	205, 239 (0.50)
expt ³⁸		201, 233 (0.53)

^aThe TD-DFT calculations have been carried out at the B3LYP/6-31+G(d) (B3) level for the QM part, while the solvent has been described by the PCM, TIP3P-FB fixed charges (FX), and FQ, FQ-CT, FQ-BC, and FQ-All models. For the isolated molecule (gas), TD-DFT band maxima computed at the B2PLYP/jun-cc-pVTZ level (B2) are also reported and used, together with B3 relative intensities and QM/MM solvent shifts to obtain the best estimate values (best).

maximum of both tautomers and the blue-side maximum of the amine tautomer but have a negligible effect on the maximum near 190 nm of the imine tautomer.

Furthermore, the relative intensity of the red-side peak of the imine tautomer is much lower (both in the gas-phase and in aqueous solution) than that of the amine counterpart, with the latter being in remarkable agreement with experiment. Both inclusion of CT and correction for spurious outer boundary effects lead to results intermediate between those issued from fixed charges or bare fluctuating charges, with the role of BC contributions decreasing after introduction of CT. Although the computed solvatochromic shifts of the spectra are not strongly sensitive to small modifications of solvent charges, it is gratifying that the FQ-CT and FQ-All models improve the agreement with experiment, thus suggesting that charge-transfer between solvent molecules could play a non-negligible role in tuning the conventional FQ results, which possibly overcorrect the FX counterparts. The close similarity between the FQ-CT and FQ-All models is related to the marginal effect of water molecules near the outside boundary on the solute UV absorption.

Computation of the spectrum of the isolated molecule at a higher level of theory (B2PLYP in conjunction with the jun-cc-pVTZ basis set³⁹ at the CCSD(T)-F12/cc-pVDZ-F12 geometries reported in ref 35) has a negligible effect on the blue-side maximum but significantly improves the agreement between theory and experiment for the red-sideband. At this level, the difference between theory and experiment is comparable for both bands (4 and 6 nm, respectively) and also their relative heights are close to the experimental counterparts. It is noteworthy that the improvement related to more accurate gas-phase computations is comparable to that issued from more refined polarizable models of solvent effects, so that both contributions need to be taken into account for a balanced computation of spectroscopic properties in solution. In this connection, we point out that for broad bands like those characterizing the UV–vis spectrum of creatinine in aqueous solution, the role of vibronic contributions (which could be effectively taken into account by the methodology employed in previous studies^{8,9}) is negligible. In summary, a physically sound description of solvent effects rooted into the fluctuating charge model coupled to TD-DFT models employing hybrid and double-hybrid functionals provides remarkably accurate results with nonprohibitive computer times for the UV–vis spectrum of a medium-size chromophore in aqueous solution, allowing also for an unambiguous discrimination between different tautomeric forms.

CONCLUSIONS

In this paper, we have extended the fluctuating charge polarizable model to take into account charge transfer and to rectify the unbalanced polarization induced by outer boundary effects. The modifications with respect to the starting model are very simple and can be added straightforwardly to any code implementing the FQ model. A first parametrization for water has shown a remarkable robustness and accuracy for reproducing solvent characteristics. Next, the UV–vis spectra of two creatinine tautomers have been computed at the TD-DFT level by means of different polarizable QM/MM models. In particular, the FQ-All model, which takes into account charge transfer between solvent molecules and enforces correct coordination numbers at the outer boundary, leads to satisfactory results, thus paving the route toward its general

use in conjunction with nonperiodic boundary conditions for the study of spectroscopic phenomena in condensed phase. Work is in progress in our laboratory to extend the parametrization of the model to other widely used solvents, e.g., alcohols or CH₃CN.

APPENDIX

We have to calculate

$$\frac{\partial \mathcal{L}}{\partial \mathbf{Q}} \frac{\partial \mathbf{Q}}{\partial x_i} = \sum_J \lambda_J \frac{\partial Q_J}{\partial x_i} \quad (27)$$

where Q_J is the charge on a generic fragment J and x_i is the x -component of a generic atomic coordinate ($\vec{R}_l = (x_l, y_l, z_l)$). The Q_J 's are defined as in eq 19, with

$$f(R_{ij}, \vartheta_{ijk}) = A(R_{ij}) B(\vartheta_{ijk}) \quad (28)$$

where:

$$\begin{aligned} R_{ij} &= \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \\ \vartheta_{ijk} &= \arccos(X_{ijk}) \\ X_{ijk} &= \frac{(x_k - x_j)(x_k - x_i) + (y_k - y_j)(y_k - y_i) + (z_k - z_j)(z_k - z_i)}{R_{kj}R_{ki}} \end{aligned} \quad (29)$$

Then,

$$\begin{aligned} \frac{\partial Q_J}{\partial x_i} &= a \sum_{j \in J} s_j \sum_{I \neq J}^{N_{frag}} \sum_{i \in I}^{N_I} \left[\frac{\partial A(R_{ij})}{\partial R_{ij}} \frac{\partial R_{ij}}{\partial x_i} B(\vartheta_{ijk}) \right. \\ &\quad \left. + A(R_{ij}) \frac{\partial B(\vartheta_{ijk})}{\partial \vartheta_{ijk}} \frac{\partial \vartheta_{ijk}}{\partial x_i} \right] \end{aligned} \quad (30)$$

which for A and B being exponential functions of R_{ij} and ϑ_{ijk} , respectively, becomes

$$\begin{aligned} \frac{\partial Q_J}{\partial x_i} &= -\frac{a}{2} \sum_{j \in J} s_j \sum_{I \neq J}^{N_{frag}} \sum_{i \in I}^{N_I} f(R_{ij}, \vartheta_{ijk}) \left[\left(\frac{R_{ij} - R_0}{\sigma_R^2} \right) \frac{\partial R_{ij}}{\partial x_i} \right. \\ &\quad \left. + \left(\frac{\vartheta_{ijk} - \vartheta_0}{\sigma_\vartheta^2} \right) \frac{\partial \vartheta_{ijk}}{\partial x_i} \right] \end{aligned} \quad (31)$$

Then, we have only three nonzero cases.¹

Case $i = l$:

$$\begin{aligned} \frac{\partial R_{ij}}{\partial x_i} &= \frac{x_i - x_j}{R_{ij}} \\ \frac{\partial \vartheta_{ijk}}{\partial x_i} &= -\frac{1}{\sqrt{1 - X_{ijk}^2}} \left[-\frac{x_k - x_j}{R_{kj}R_{ki}} + \frac{X_{ijk}(x_k - x_i)}{R_{ki}^2} \right] \\ \frac{\partial Q_J}{\partial x_i} &= -\frac{a}{2} \sum_{j \in J} s_j f(R_{ij}, \vartheta_{ijk}) \left[\left(\frac{R_{ij} - R_0}{\sigma_R^2} \right) \frac{x_i - x_j}{R_{ij}} \right. \\ &\quad \left. + \left(\frac{\vartheta_{ijk} - \vartheta_0}{\sigma_\vartheta^2} \right) \frac{\partial \vartheta_{ijk}}{\partial x_i} \right] \end{aligned} \quad (32)$$

Case $j = l$ (please note that the indexes i and j are not equivalent since in the present formalism they are bound on the J and I fragments, respectively, where J is fixed while I is active):

$$\begin{aligned} \frac{\partial R_{ij}}{\partial x_j} &= -\frac{x_i - x_j}{R_{ij}} \\ \frac{\partial \vartheta_{ijk}}{\partial x_j} &= -\frac{1}{\sqrt{1 - X_{ijk}^2}} \left[-\frac{x_k - x_i}{R_{ki}R_{kj}} + \frac{X_{ijk}(x_k - x_j)}{R_{kj}^2} \right] \\ \frac{\partial Q_J}{\partial x_j} &= -\frac{a}{2} s_j \sum_{I \neq J}^{N_{frag}} \sum_{i \in I}^{N_I} f(R_{ij}, \vartheta_{ijk}) \left[-\left(\frac{R_{ij} - R_0}{\sigma_R^2} \right) \frac{x_i - x_j}{R_{ij}} \right. \\ &\quad \left. + \left(\frac{\vartheta_{ijk} - \vartheta_0}{\sigma_\vartheta^2} \right) \frac{\partial \vartheta_{ijk}}{\partial x_j} \right] \end{aligned} \quad (33)$$

Case $k = l$:

$$\begin{aligned} \frac{\partial R_{ij}}{\partial x_k} &= 0 \\ \frac{\partial \vartheta_{ijk}}{\partial x_k} &= -\frac{1}{\sqrt{1 - X_{ijk}^2}} \left[\frac{x_k - x_i}{R_{kj}R_{ki}} + \frac{x_k - x_j}{R_{kj}R_{ki}} - \frac{X_{ijk}(x_k - x_j)}{R_{kj}^2} \right. \\ &\quad \left. - \frac{X_{ijk}(x_k - x_i)}{R_{ki}^2} \right] \\ \frac{\partial Q_J}{\partial x_k} &= -\frac{a}{2} \sum_{j \in J} s_j \sum_{I \neq J}^{N_{frag}} \sum_{i \in I}^{N_I} f(R_{ij}, \vartheta_{ijk}) \left(\frac{\vartheta_{ijk} - \vartheta_0}{\sigma_\vartheta^2} \right) \frac{\partial \vartheta_{ijk}}{\partial x_k} \end{aligned} \quad (34)$$

AUTHOR INFORMATION

Corresponding Author

Vincenzo Barone – Scuola Normale Superiore, 56126 Pisa, Italy; orcid.org/0000-0001-6420-4107;
Email: vincenzo.barone@sns.it

Authors

Ivan Carnimeo – Scuola Internazionale Superiore di Studi Avanzati, 34136 Trieste, Italy
Giordano Mancini – Scuola Normale Superiore, 56126 Pisa, Italy; orcid.org/0000-0002-1327-7303
Marco Pagliai – Dipartimento di Chimica “Ugo Schiff”, Università degli Studi di Firenze, 50019 Sesto Fiorentino, Italy; orcid.org/0000-0003-0240-161X

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsomega.2c01110>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work has been supported by MIUR (Grant Number 2017A4XRCA) and by the Italian Space Agency (ASI; Life in Space project, No. 2019-3-U.0). The SMART@SNS Laboratory is acknowledged for providing high-performance computing facilities. We thank Gianni Cardini for useful discussions.

ADDITIONAL NOTE

$$\frac{1}{dx} \arccos x = -\frac{1}{\sqrt{1-x^2}}$$

REFERENCES

- (1) Crescenzi, O.; Pavone, M.; De Angelis, F.; Barone, V. Solvent effects on the UV ($n-\pi^*$) and NMR (C-13 and O-17) spectra of acetone in aqueous solution. An integrated Car-Parrinello and DFT/PCM approach. *J. Phys. Chem. B* **2005**, *109*, 445–459.
- (2) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.
- (3) Mennucci, B. Modeling environment effects on spectroscopies through QM/classical models. *Phys. Chem. Chem. Phys.* **2013**, *15*, 6583–6594.
- (4) Olsen, J. M. H.; Steinmann, C.; Ruud, K.; Kongsted, J. Polarizable density embedding: a new QM/QM/MM-based computational strategy. *J. Phys. Chem. A* **2015**, *119*, 5344–5355.
- (5) Boulanger, E.; Thiel, W. Solvent boundary potentials for hybrid QM/MM computations using classical Drude oscillators: a fully polarizable model. *J. Chem. Theory Comput.* **2012**, *8*, 4527–4538.
- (6) Lipparini, F.; Cappelli, C.; Barone, V. Linear response theory and electronic transition energies for a fully polarizable QM/classical hamiltonian. *J. Chem. Theory Comput.* **2012**, *8*, 4153–4165.
- (7) Mancini, G.; Brancato, G.; Chandramouli, B.; Barone, V. Organic solvents simulations under non periodic boundary conditions: a library of effective potentials for the GLOB model. *Chem. Phys. Lett.* **2015**, *625*, 186–192.
- (8) Del Galdo, S.; Chandramouli, B.; Mancini, G.; Barone, V. 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. *J. Chem. Theory Comput.* **2019**, *15*, 3170–3184.
- (9) Del Galdo, S.; Fusè, M.; Barone, V. The ONIOM/PMM model for effective yet accurate simulation of optical and chiroptical spectra in solution: camphorquinone in methanol as a case study. *J. Chem. Theory Comput.* **2020**, *16*, 3294–3306.
- (10) Mancini, G.; Fusè, M.; Lipparini, F.; Nottoli, M.; Scalmani, G.; Barone, V. Molecular Dynamics Simulations Enforcing Nonperiodic Boundary Conditions: New Developments and Application to Solvent Shifts of Nitroxide Magnetic Parameters. *J. Chem. Theory Comput.* **2022**, DOI: 10.1021/acs.jctc.2c00046.
- (11) Pagliai, M.; Muniz-Miranda, F.; Cardini, G.; Righini, R.; Schettino, V. Hydrogen bond dynamics of Methyl Acetate in Methanol. *J. Phys. Chem. Lett.* **2010**, *1*, 2951–2955.
- (12) Martin-Noble, G. C.; Reilley, D.; Rivas, L. M.; Smith, M. D.; Schrier, J. EQeq+C: An empirical bond-order-corrected extended charge equilibration method. *J. Chem. Theory Comput.* **2015**, *11*, 3364–3374.
- (13) Lipparini, F.; Barone, V. Polarizable Force Fields and Polarizable Continuum Model: A Fluctuating Charges/PCM Approach. 1. Theory and Implementation. *J. Chem. Theory Comput.* **2011**, *7*, 3711–3724.
- (14) Ohno, K. Some remarks on the Pariser-Parr-Pople method. *Theor. Chim. Acta* **1964**, *2*, 219–227.
- (15) Carnimeo, I.; Cappelli, C.; Barone, V. Analytical Gradients for MP2, Double Hybrid Functionals, and TD-DFT with Polarizable Embedding Described by Fluctuating Charges. *J. Comput. Chem.* **2015**, *36*, 2271–2290.
- (16) Wang, L.-P.; Martinez, T. J.; Pande, V. S. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 1885–1891.
- (17) Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A., Jr.; Morokuma, K.; Frisch, M. J. Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J. Chem. Theory Comput.* **2006**, *2*, 815–826.
- (18) Lipparini, F.; Cappelli, C.; Barone, V. Linear response theory and electronic transition energies for a fully polarizable QM/Classical Hamiltonian. *J. Chem. Theory Comput.* **2012**, *8*, 4153–4165.
- (19) Morzan, U. N.; Alonso de Armiño, D. J.; Foglia, N. O.; Ramirez, F.; González Lebrero, M. C.; Scherlis, D. A.; Estrin, D. A. Spectroscopy in Complex Environments from QMMM Simulations. *Chem. Rev.* **2018**, *118*, 4071–4113.
- (20) Cao, L.; Ryde, U. On the Difference Between Additive and Subtractive QM/MM Calculations. *Frontiers in Chemistry* **2018**, *6*, 89.
- (21) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (22) 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. Model.* **2017**, *57*, 1240–1245.
- (23) Mennucci, B. Polarizable continuum model. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 386–404.
- (24) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. 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.
- (25) Del Galdo, S.; Fusè, M.; Barone, V. CPL Spectra of Camphor Derivatives in Solution by an Integrated QM/MD Approach. *Front. Chem.* **2020**, *8*, 584.
- (26) Mancini, G.; Del Galdo, S.; Chandramouli, B.; Pagliai, M.; Barone, V. Computational Spectroscopy in Solution by Integration of Variational and Perturbative Approaches on Top of Clusterized Molecular Dynamics. *J. Chem. Theory Comput.* **2020**, *16*, 5747–5761.
- (27) Macchiagodena, M.; Mancini, G.; Pagliai, M.; Del Frate, G.; Barone, V. Fine-tuning of atomic point charges: Classical simulations of pyridine in different environments. *Chem. Phys. Lett.* **2017**, *677*, 120–126.
- (28) Macchiagodena, M.; Mancini, G.; Pagliai, M.; Cardini, G.; Barone, V. New atomistic model of pyrrole with improved liquid state properties and structure. *Int. J. Quantum Chem.* **2018**, *118*, No. e25554.
- (29) Lazzari, F.; Salvadori, A.; Mancini, G.; Barone, V. Molecular Perception for Visualization and Computation: The Proxima Library. *J. Chem. Inf. Model.* **2020**, *60*, 2668–2672.
- (30) Barone, V.; Puzzarini, C.; Mancini, G. Integration of Theory, Simulation, Artificial Intelligence and Virtual Reality: A Four-Pillar Approach for Reconciling Accuracy and Interpretability in Computational Spectroscopy. *Phys. Chem. Chem. Phys.* **2021**, *23*, 17079–17096.
- (31) Mancini, G.; Brancato, G.; Barone, V. Combining the Fluctuating Charge Method, Non-periodic Boundary Conditions and Meta-dynamics: Aqua Ions as Case Studies. *J. Chem. Theory Comput.* **2014**, *10*, 1150–1163.
- (32) Frisch, M. J. et al. *Gaussian 16*, rev. C.01; Gaussian Inc.: Wallingford, CT, 2016.
- (33) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, No. 014101.
- (34) Rodriguez, A.; Laio, A. Clustering by fast search and find of density peaks. *Science* **2014**, *344*, 1492–1496.
- (35) Léon, I.; Tasinato, N.; Spada, L.; Alonso, E. R.; Mata, S.; Balbi, A.; Puzzarini, C.; Alonso, J. L.; Barone, V. Looking for the Elusive Imine Tautomer of Creatinine: Different States of Aggregation Studied by Quantum Chemistry and Molecular Spectroscopy. *ChemPlusChem.* **2021**, *86*, 1374–1386.
- (36) 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.
- (37) Han, J.; Kamber, M. *Data mining: concepts and techniques*, 3rd ed.; Elsevier: Burlington, MA, 2011.
- (38) Berlet, H. H.; Völk, A. Die Ultraviolet-Absorption von Kreatinin und Glykocynamidin im Vergleich mit anderen Guanidin-

verbindungen und analogen Hydantoinen. *Z. Klin. Chem. Klin. Biochem.* **1975**, *13*, 53–59.

(39) Papajak, E.; Zheng, J.; Xu, X.; Leverentz, H. R.; Truhlar, D. G. Perspectives on Basis Sets Beautiful: Seasonal Plantings of Diffuse Basis Functions. *J. Chem. Theory Comput.* **2011**, *7*, 3027–3034.