Polarizable Multiscale Dynamics for probing solvent and complex environments

Daniele Loco, Jean-Philip Piquemal

Polarizable Multiscale Dynamics for probing solvent and complex environments

D. Loco[a*], J.-P. Piquemal[b*]

[a] Qubit Pharmaceuticals, Incubateur Paris Biotech Santé, 24 rue du Faubourg Saint Jacques 75014 Paris (France)., [b] Sorbonne Université, Laboratoire de Chimie Théorique, UMR 7616 CNRS, 4 Place Jussieu, 75005 Paris (France)

*corresponding email addresses: daniele@qubit-pharmaceuticals.com, jean-philip.piquemal@sorbonne-universite.fr

ABSTRACT/WEB SUMMARY

Computational modelling approaches have been developed in the last decades to account for environment effects, that can remarkably affect the spectroscopic features of organic dyes. With modern computing power, complex systems can be handled through QM/MM simulations, which are nowadays commonly used to assist experiments in many different

fields. We present here some of the latest advances made in our laboratory to develop and apply a pPolarizable Embedding QM/MM based molecular dynamics methodology, and discussing applications where it is found useful. Then we report on the spectroscopic study of a ligand-protein complex system, together with our perspective on the modelling of reactions in extreme conditions, illustrating our recent study on thermal Diels-Alder.

1.1 Introduction

A variety of photophysical and photochemical processes are common in Nature, and many different fields of research are dedicated to study them. Chromophores of various natures are exploited in different technologies[1], and some as chlorophylls are crucial in emblematic phenomena as the photosynthesis in plants and in different microorganisms. A crucial role is usually reserved to the environment, where photoactive species live. We define here as "environment" everything that is not directly, or "actively", involved in the processes induced by the interaction with light. From a purely quantitative point of view, e.g., the molecular composition in terms of atoms, the environment is overwhelmingly bigger than the photoactive species, *e.g.*, the p-coumaric acid embedded inside the photoactive yellow protein.

A simple but well known environment effect is the shift in the absorption and fluorescence signal[2,3] related to some electronic transition of a chromophore. As a side note, we remind here that the reference, often adopted in computational chemistry, to quantify the shifting effect, is the ideal condition of the chromophore *in vacuo*, *i.e.*, with no interactions in an "empty" surrounding. Modelling techniques are commonly adopted to aid in understanding this and many more processes, and different approaches have been developed to deal with the effects of a variety of environments, ranging from macro-biomolecules to nanomaterials[4].

In this contribution, we will report on some of our recent works, to develop a multiscale modelling approach, so to treat complex systems and, eventually, their interaction with light. We are going to present the basics of the method, mixing a quantum chemistry (QC) description of a limited portion of a complex molecular system (*e.g.*, a chromophore complexed with a protein), and a classical, force field (FF) based description of all the remaining. This modelling strategy, meant to deal with large systems with a high level of accuracy, is normally referred to as a QM/MM (Quantum Mechanics/Molecular Mechanics) approach.

The specific aspect of the QM/MM method we discuss here, is the explicit inclusion of the mutual polarization between the quantum and classical portions[5-8], adopting an advanced polarizable FF which allows such a treatment, namely the AMOEBA FF[9]. This is generally defined as a Polarizable Embedding (PE).

The AMOEBA classical model accounts for polarization effects assigning at each atom type an atomic polarizability. The polarizability allows each atom to respond to the presence of an applied electric field, as the one generated by the chemical environment where the atoms live. This results is an induced atomic dipole, a vectorial quantity having intensity and direction determined by the effect of the environment, so to minimize the interaction energy of all atoms between each other.

We especially focus our interest on the coupling of a PE QM/MM with the dynamics of the system, that enables to include the effects of polarization into the simulation of the molecular motion of a complex system.

The inclusion of polarization in the QM/MM philosophy dates back to the seminal paper by Warshel and Levitt, as it ensure a high degree of realism in the simulation[10]. We recently discussed our perspectives on the formulation and developments of a polarizable QM/MM based molecular dynamics (MD) methods, as a promising technology to study complex systems and complement experiments[5].

In the following sections, we first present the main methodological aspects of the QM/AMOEBA multiscale dynamics we worked on in the last years, so to illustrate later its usage in a photochemical application.

2.1 Modelling polarization in a QM/MM MD approach

At first place, we define how the polarization is included in the QM/AMOEBA model. To this aim, the first question to ask is how to define the polarization of an ensemble of atoms and/or molecules.

The polarization of a molecular system composed by N particles is determined by the mutual interaction between each of the components of the system. The polarization interactions are thus understood as purely many-body effects, for which each particle is affected by the simultaneous presence and interaction with all the other N-1 particles. From an algorithmic point of view, treating a polarization problem corresponds to solving a nonlinear system of equations , *i.e.*, the unknown polarization is also a variable of the equations to be solved, which can be treated numerically in a so called self-consistent field (SCF) procedure.

SCF procedures are commonly applied to numerically treat also QC models, widely used to compute the electronic structure properties of atoms and molecules, as the Hartree-Fock (HF) or Kohn-Sham (KS) density functional theory (DFT). For this reason the coupling of a QC method to a polarizable model, to include environment effects in a QM/MM fashion, appears natural.

2.1.1 Polarizable embedding QM/MM

We start from the definition of the QM/MM fundamental equations. These are needed to understand the basics of the approach and the physical quantities that can be computed: the ground state energy of the complex system, and its derivatives when needed to compute molecular properties.

Also nuclear gradients can be seen as a property, needed to simulate the dynamics of the system in a classical picture.

The QM/MM global energy of the whole embedded system (S) can be expressed as:

$$E^{QM/MM}(S) = E^{MM}(\mathcal{I}) + E^{QM}(\mathcal{A}) + E^{QM/MM}(\mathcal{A},\mathcal{I}) \quad (2.1)$$

$$E^{QM/MM}(\mathcal{A},\mathcal{I}) = E^{QM/MM}_{Bond} + E^{QM/MM}_{vdW} + E^{QM/MM}_{El} \quad (2.2)$$

$$E^{QM/MM}_{El} = E^{QM/MM}_{Fix\ El} + E^{QM/MM}_{Pol} \quad (2.3)$$

where E^{MM} is the classical FF energy for the environment (the inactive subsystem, I) and E^{QM} depends on the QC model used for the smaller active subsystem. $\mathrm{E}^{QM/MM}(\mathcal{A},\mathcal{I})$ represents the interaction energy between the two subsystems, generally expressed as a sum of bonding and non-bonding terms. The term $\mathrm{E}^{QM/MM}_{Bond}$ collects the firsts, as they are represented in the classical FF; this term could include also terms describing the bonding between QM and MM atoms[7].

The $\mathrm{E}^{QM/MM}_{vdW}$ are part of the non-bonding interactions, and represents short-range, dispersion-repulsion interactions, which are normally treated using Lennard-Jones-like potentials. Lastly, $\mathrm{E}^{QM/MM}_{El}$ accounts for the leading QM/MM coupling term, which is electrostatic in nature, and which also includes the self-consistent mutual polarization between the QM and the classical subsystem, as well as the polarization of the whole classical subsystem. Following the usual energy decomposition adopted by polarizable FFs,

the fixed-electrostatic energy $\mathrm{E}_{Fix\ El}^{QM/MM}$, which we define "fixed" as opposed to the polarization energy term, models the pure electrostatic interactions, while $\mathrm{E}_{Pol}^{QM/MM}$ accounts for the portion of the electrostatic QM/MM interactions which arises from the polarization between the two subsystems. In our formulation polarization is modelled through atomic induced-point dipole moments, generally denoted with the vector $\boldsymbol{\mu}$.

A double nested SCF method is needed to compute, at once, the QM and the polarization energy of the embedded system. To do so, the following eigenstate equation

$$\hat{\mathcal{H}}^{eff}(\mathbf{r}, \mathbf{R})\ \Psi_\alpha > = |\hat{\mathcal{H}}^0(\mathbf{r}) + \hat{\mathcal{H}}^{env}(\mathbf{R}) + \hat{\mathcal{H}}^{int}(\mathbf{r}, \mathbf{R})|\ \Psi_\alpha > = E_\alpha |\Psi_\alpha > \quad (2.4)$$

is formally stated, where the effective Hamiltonian operator includes the *in vacuo* QM Hamiltonian $\hat{\mathcal{H}}^0(\mathbf{r})$, the classical one for the classical environment $\hat{\mathcal{H}}^{env}(\mathbf{R})$, and the interaction term between the two $\hat{\mathcal{H}}^{int}(\mathbf{r}, \mathbf{R})$.

Once this set of equations is solved, the quantum state energy is formally given by the expectation value $E_\alpha = < \Psi_\alpha\ \hat{\mathcal{H}}^{eff} | \Psi_\alpha >$, for the α-th state, which is, in general, the electronic ground state of the QM subsystem.

In practical implementations, an approximate model is introduced so to solve for eq. 2.4, within a SCF approach. In the HF/KS-DFT, the QM subsystem wavefunction/electronic density is expanded in a finite, atomic orbitals (AOs) basis set, and a corresponding energy functional is formulated. The terms discussed in eq. 2.2 and 2.3 are also included, to properly account for the environment contributions. The system energy then reads

$$E(\mathbf{D}, \boldsymbol{\mu}) = tr\mathbf{h}\mathbf{D} + \frac{1}{2} tr\mathbf{D}\mathbf{G}(\mathbf{D}) + E^{xc} + E^{El}(\boldsymbol{\mu}, \mathbf{D}) + V^{FF} \quad (2.5)$$

with $\mathbf{D}$ the electronic one-body density matrix

$$\mathbf{D} = \sum_{K} \mathbf{C}_{K} \mathbf{C}_{K}^{\dagger} \quad (2.6)$$

Taking the derivative of eq. 2.5 with respect to the one-body density matrix $\mathbf{D}$, we can define the matrix elements of the HF/KS operator

$$F_{\mu\nu} = h_{\mu\nu} + G_{\mu\nu}(\mathbf{D}) + f_{\mu\nu}^{xc} + f_{\mu\nu}^{El}(\boldsymbol{\mu}) \quad (2.6)$$

We refer to the literature[4,5] for a detailed discussion of the terms of eq. 2.6, which includes, in order: 1- and 2-electrons integrals, the electronic exchange correlation integrals in terms of the AO basis, for which greek indexes are used, and the contribution due to the electrostatic of the environment.

In a compact matrix notation, the fundamental equation to solves is the Roothaan-Hall equation

$$\mathbf{F}(\boldsymbol{\mu})\mathbf{C} = \mathbf{F}(\boldsymbol{\mu})\mathbf{S}\epsilon \quad (2.7)$$

from which the density matrix $\mathbf{D}$, and the state energy ϵ, affected by the induced dipoles on each classical site, can be recomputed in the SCF procedure, until the convergence criterion, usually on the energy and density variations, is satisfied.

2.1.2 The QM/AMOEBA method for MD

The electrostatic interactions between the environment and the electronic density of the QM subsystem are at the core of model. In fact, the Coulomb interactions between the electrostatic model used to describe the classical environment, and the electrons and nuclei of the QM subsystem, are the leading terms of the QM/MM coupling, generally larger than other non-bonding interactions, as the dispersion-repulsion ones.

We can detail the form of this interactions expanding the electrostatic term in eq. (2.5) as

$$E^{El}(\boldsymbol{\mu}, \mathbf{D}) = E_{N}^{Fix-El} + E_{e}^{Fix-El}(\mathbf{D}) + E^{Pol}(\boldsymbol{\mu}, \mathbf{D}) \quad (2.8)$$

where

$$E_N^{Fix\ \ El} = \mathcal{Q}^{\dagger}\mathbf{V}_N^{Fix\ \ El} = \sum_i^{N_{MM}} \sum_{k=0}^{K_i} \frac{(-1)^k}{k} \mathcal{Q}_i^{(k)} \sum_N^{N_{QM}} T_{iN}^{(k)} Z_N \quad (2.9)$$

is the QM nuclear charges Z_N interaction with the multipoles $\mathcal{Q}_i$ of the AMOEBA atoms. In the AMOEBA electrostatic model, at each atom is assigned a set of multipoles from point charges to static quadrupolar moments. The terms $Q_i^{(k)}$ represent the multipoles of order k for the atom type i. The interaction fields $T_{iN}^{(k)}$ represent the Coulomb potential, the electric field and electric field gradient, respectively, for k from 0 to 2.

The second term of the equation

$$E_e^{Fix-El}(\mathbf{D}) = \mathcal{Q}^{\dagger}\mathbf{V}_e^{Fix-El}(\mathbf{D}) = \sum_i^{N_{MM}} \sum_{k=0}^{K_i} \frac{(-1)^k}{k} \mathcal{Q}_i^{(k)} \sum_{\mu\nu}^{N_b} D_{\mu\nu} t_{\mu\nu}^{(k)}(\vec{r}_i) \quad (2.10)$$

is the classical multipoles interaction with the electron density (*i.e.*, with the electron density matrix). The electrostatic interaction terms generated by the electron density at the classical sites coordinates $\vec{r}_i$ is expressed ad a sum of contributions in terms of the atomic basis set, represented through the greek letters μ and ν. More details can be found in our recent Account on the polarizable environments[5].

The eq. 2.9 and 2.10 only describe the fixed classical electrostatic interactions with the QM density, the polarization is still missing.

AMOEBA assumes an induced point dipole model to account for polarizabilities. For this reason, according to the definition of a polarizable model, the induced dipoles are computed at each nuclear configuration of the system solving the following set of equations

$$\mu(\mathbf{D}) = \mathbf{T}^{-1}(\mathcal{F}_n + \mathcal{F}_e(\mathbf{D}) + \mathcal{F}_{\mathcal{Q}}) = \mathbf{T}^{-1}\mathcal{F}_{ind}(\mathbf{D}) \quad (2.11)$$

Each atomic induced dipole μ is proportional to the electric fields $\mathcal{F}$ generated by the fixed electrostatic $\mathcal{Q}$ and the QM electron density $\mathbf{D}$ and nuclear charges. All together, those terms form the inducing electric field $\mathcal{F}_{ind}$. Eq. 2.11 is in a compact matrix form, and each electric field term can be expressed as

$$(\boldsymbol{\mathcal{F}}_n)_i = \sum_N^{N_{QM}} T_{iN}^{(1)} Z_N \qquad (2.12)$$

as for the QM nuclei electrostatic terms due to the QM nuclei in eq.2.10, while

$$(\boldsymbol{\mathcal{F}}_e(\mathbf{D}))_i = \sum_{\mu\nu}^{N_b} D_{\mu\nu} t_{\mu\nu}^{(1)}(\vec{r}_i) \qquad \forall i \in N_{MM} \qquad (2.13)$$

with N_{MM} the total number of classical atoms, is the corresponding term due to the electron density $\mathbf{D}$. The term $\mathcal{F}_{\mathcal{Q}}$ is an extension of eq. 2.12 up to the atomic electric quadrupole moment.

The induced dipoles in eq. 2.9 are proportional to the so called classical linear response matrix T, which accounts for the damped induced dipole-induced dipole interactions. It reads as

$$\mathbf{T} = \begin{pmatrix} \boldsymbol{\alpha}_1^{-1} & \mathbf{T}_{12}^{(2)} & \mathbf{T}_{1N_{Pol}}^{(2)} \\ \mathbf{T}_{21}^{(2)} & \boldsymbol{\alpha}_2^{-1} & \mathbf{T}_{2N_{Pol}}^{(2)} \\ \mathbf{T}_{N_{Pol}1}^{(2)} & \mathbf{T}_{N_{Pol}2}^{(2)} & \boldsymbol{\alpha}_{N_{Pol}}^{-1} \end{pmatrix} \qquad (2.14)$$

where diagonal elements are inverse atomic polarizabilities, which is a classical parameter in the AMOEBA FF, and the off-diagonal terms are the 3×3 dipole-dipole interaction tensors between atoms pairs. For convenience, we changed here the notation for the number of classical sites from N_{MM} to N_{Pol}, to enforce the fact that this term only arise from

polarization effects. Eventually the full classical system could be defined so that $N_{MM} \neq N_{Pol}$.

The system of equations 2.11 can be solved, *e.g.,* using an iterative approach. This is done at each new nuclear configuration during an MD simulation, and for each cycle of the SCF required to obtain the QM electron density.

Having solved the polarization problem, one can compute the polarization energy as[5]

$$E^{\mathrm{Pol}}(\boldsymbol{\mu}, \mathbf{D}) = \frac{1}{2}\boldsymbol{\mathcal{F}}^{\dagger}_{ind}(\mathbf{D})\mathbf{T}^{-1}\boldsymbol{\mathcal{F}}_{ind}(\mathbf{D}) = \frac{1}{2}\boldsymbol{\mu}^{\dagger}\boldsymbol{\mathcal{F}}_{ind}(\mathbf{D}) \quad (2.15)$$

which is the classical expression for the polarization energy of a dipole in an external inducing electric field.

Tacking the derivative of eq. 2.10, with respect to the QM electronic density matrix $\mathbf{D}$, one obtains the explicit expression for the modified Fock matrix, whose elements, in the atomic basis

$$F_{\mu\nu} = F^{SCF}_{\mu\nu} + \boldsymbol{\mathcal{Q}}^{\dagger}\mathbf{t}_{\mu\nu} - \frac{1}{2}\boldsymbol{\mu}^{\dagger}\mathbf{t}^{(1)}_{\mu\nu} \quad (2.16)$$

take explicitly into account the environment electrostatics.

Solving the coupled equations 2.10 and 2.16 also requires to solve for the induced dipoles in eq. 2.11. This amounts to completely solve the PE QM/MM nested SCF procedure.

The nuclear dynamics to perform a classical polarizable QM/MM MD is driven by the potential energy 2.10. An integration scheme is applied to solve for the time-discretized Newton's second law differential equations, using for example a velocity Verlet scheme.

The analytical expression of the forces is derived from eq. 2.10 with respect to the nuclear positions of each atom composing the system. Those forces are added to the full classical one, obtained from the classical FF, and the full quantum one, obtained from the electronic structure method used for the QM subsystem. We refer here to the available literature[5,7], for the readers interested in the expression of the QM/MM forces.

Formally this is a Born-Oppenheimer (BO) dynamics, meaning that the electronic and polarization degrees of freedom are completely relaxed at each step of the dynamics, so that the system is moving being always at the bottom of the PE QM/MM potential energy surfaces.

The model has been first implemented in a developing version of the Gaussian09 software[11] and coupled with the highly parallel molecular dynamics software Tinker-HP[12].

2.3 Excitation energy including environment effects

Electronic excitation energies of the QM subsystem can then be computed in a time-dependent DFT (TDDFT) scheme[6]. The set of equations read as

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{X}_K \\ \mathbf{Y}_K \end{pmatrix} = \omega_K \begin{pmatrix} 1 & \mathbf{0} \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \mathbf{X}_K \\ \mathbf{Y}_K \end{pmatrix}$$

where the electronic transition energies ω_K and densities X_K and Y_K are the solutions we are seeking. The $\mathbf{A}$ and $\mathbf{B}$ matrices incorporate the effects of the PE in two ways: first, the integrals from which they are built are generated within the effect of the environment, through molecular orbitals (MOs) determined in the QM/MM SCF computation. Moreover, they both explicitly include a term arising from the mutual QM/MM polarization.

Those matrices read:

$$A_{ai\,bj} = \delta_{ab}\delta_{ij}(\epsilon_a - \epsilon_i) + (ai|bj) - c_x(ab\,ij) + + c_l f^{xc}_{ai\,bj} + C^{Pol}_{ai\,bj} = A_{ai\,bj} + C^{Pol}_{ai\,bj}$$

$$B_{ai\,bj} = (ai|bj) \quad c_x(aj\,ib) + C^{Pol}_{ai\,bj} = B_{ai\,bj} + C^{Pol}_{ai\,bj}$$

with ϵ the orbital energies, (pq|rs) are two-electron integrals in Mulliken notation, and the coefficients c_x and c_l modulate the basic electronic structure theory assumed in the calculation (pure Hartree–Fock, pure DFT or a mix of those)[6].

The classical induced dipoles interaction with the electronic transition density, reads as

$$C^{Pol}_{ai\,bj} = \sum_k^{MM} \left(\int_{\mathbb{R}^3} \phi_a(\vec{r}) \frac{(\vec{r} - \vec{r}_p)}{\vec{r} - \vec{r}_p^{\,3}} \phi_i(\vec{r})\,\mathrm{d}\vec{r} \right) \times \boldsymbol{\mu}_k(\phi_b, \phi_j)$$

where a,b and i,j indices refer to occupied and virtual MOs, respectively, and the induced atomic dipoles are generated by the transition electronic density.[6,8]

3.1 Applications

A PE based QM/MM MD modelling can arguably help in understanding chemical and physical processes which require an adequate inclusion of environment effects. We would like here to highlight, between the many possible applications, two cases: i) the photophysics of an organic chromophore in biological matrices and ii) the chemical reactivity under extreme physical conditions.

We focus on this yet limited set of phenomena, to offer some examples of a computational protocol, exploiting accurate classical models to represent the environment and its effects.

For the first case, we briefly report on an example, where the electronic absorption of two different chemical forms of an organic dye is simulated in an enzyme, the cofactor-free (1H)-3-hydroxy-4-oxoquinaldine 2,4-dioxygenase (HOD). The two forms are showing a quite different experimental absorption spectra, which can be reproduced through the propose model, which can arguably be considered a good method to monitor the enzymatic mechanism.

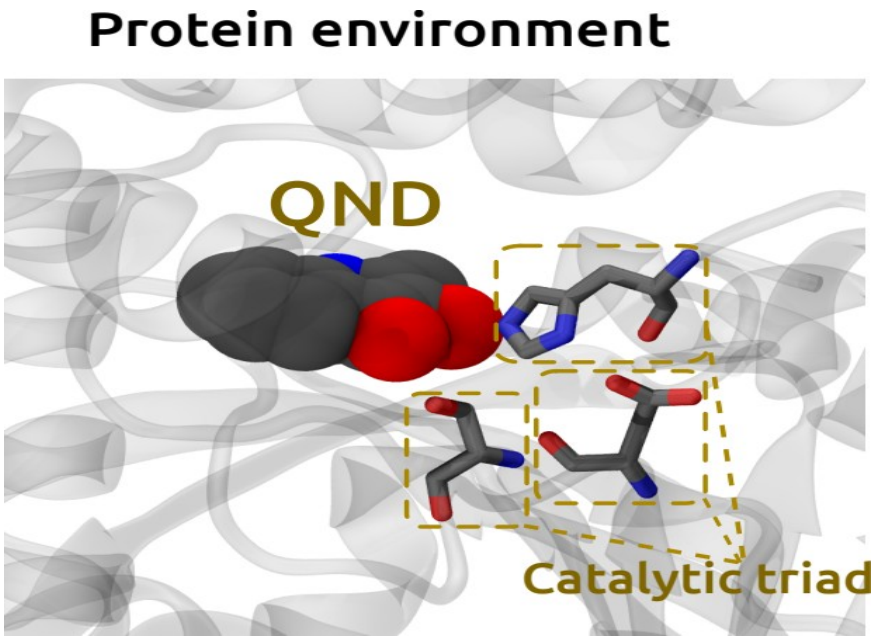

Figure 3.1: Zoom on the HOD binding pocket which accommodates the substrate 1-H-3-hydroxy-4-oxoquinaldine (QND). Hydrogen atoms are not shown for the sake of clarity. The protein residues highlighted are included, together with the QND substrate, in the QM subsystem in the QM/MM simulations.

For the second case, we propose our perspectives on the extension of the model to simulate the excited state reactivity of solvated organic species, under non-ordinary physical conditions, as e.g., high pressures or temperatures above 300 K, taking advantage of the accurate description of the environment effects. An example of thermal Diels-Alder is presented, as the basic scheme for a similar strategy which can be developed, including excited states dynamics.

3.1.1 Photophysics in biomolecules

The first application is dedicated to simulate the electronic absorption properties of the oxoquinaldine (QND) organic dye, which binds in the buried catalytic pocket of the cofactor-free (1H)-3-hydroxy-4-oxoquinaldine 2,4-dioxygenase (HOD).

The HOD protein is an oxidoreductase, whose mechanism of action in the oxidation process it catalyses is debated. In fact it involves a closed shell singlet organic species (QND) to interact the triplet oxygen molecular species. The spin-flip required is a largely unfavorable process[13,14]. Even though the full oxygenation reaction mechanism is still a source of debate, studies support the fact that the QND substrate is activated by an initial proton transfer. The proximal Histidine 251 residue (His251) could in fact extract a QND sufficiently acid hydrogen, so that the QND anionic form can react with the oxygen.[13,14]

The spectroscopic electronic absorption has been shown to be a valid technique to prove that.[13,14] The signature of the anionic species and of its protonated form, are in fact remarkably different. The advancement of the first step for the oxygen incorporation reaction, can be monitored observing a large shift toward the red, with a broad band, in the wild type QND-HOD complex in anaerobic conditions, where the absorption band extends from 340 nm, corresponding to the neutral QND absorption, to 420[13,14]. Since no oxygen is available, the anionic form of QND can be observed. Such a large shift is partially ascribable to the different electronic properties of the neutral and anionic QND, but also the environment is expected to play a role.

Simulations can greatly help towards understanding the processes involved in the different steps of the reaction mechanism, furnishing molecular level details.

As a first step, here we show the preliminary results obtained by applying a mixed, step-by-step strategy, based on classical polarizable dynamics and polarizable hybrid QM/MM simulations, so to reproduce the spectroscopic signatures of the neutral and anionic form of the QND dye in the protein environment.
With “mixed step-by-step strategy”, we refer to the following composite computational protocol:

1. Classical MD: the full system is treated at the AMOEBA FF level, allowing for long trajectory, in the order of 10 ns, so to sample different chromophore-protein configurations.

2. QM/AMOEBA MD: a higher accuracy is required to refine the chromophore structure and dynamics, taking into account the effects of the presence of the protein and the coupled QM and MM dynamics. Here the environment is still treated with the AMOEBA FF; such higher computational cost limits the length of single dynamics, which are run for ~1 ps, in parallel on several structures extracted from the classical MD.

3. Single point QM/AMOEBA TDDFT: a number of structures, in the order of 100, is extracted from the set of QM/AMOEBA MD, so to compute excitation energies for each chromophore-protein sampled structure. The excitation energies are then averaged.

The polarizable AMOEBA force field is consistently employed all along the application and is used to describe the environment and its effects on the QM subsystem.

Starting from the relevant wild type HOD and H251A variant, where the His251 is substituted by an alanine, so to avoid the formation of the QND anion (Protein Data Bank entries 2WJ4 and 4CFS, respectively), we perform an equilibration followed by a constrained dynamics, all at the full classical level. The first steps are used to keep the QND blocked in its crystallographic configuration, but within a protein structure equilibrated at room temperature and 1atm. The AMOEBA FF is used to model the protein scaffold and ~10000 water solvent molecules used to generate the MD simulation box. The simulation box is a cube of ~65 x 65 x 65 $Å^3$ size, at the centre of which the crystallographic structure is placed.

This stage of the procedure is meant both to equilibrate the protein structure in solution, at the required external conditions, and also to sample different configurations of the system, from which the excitation properties will be computed at the TDDFT level of theory.

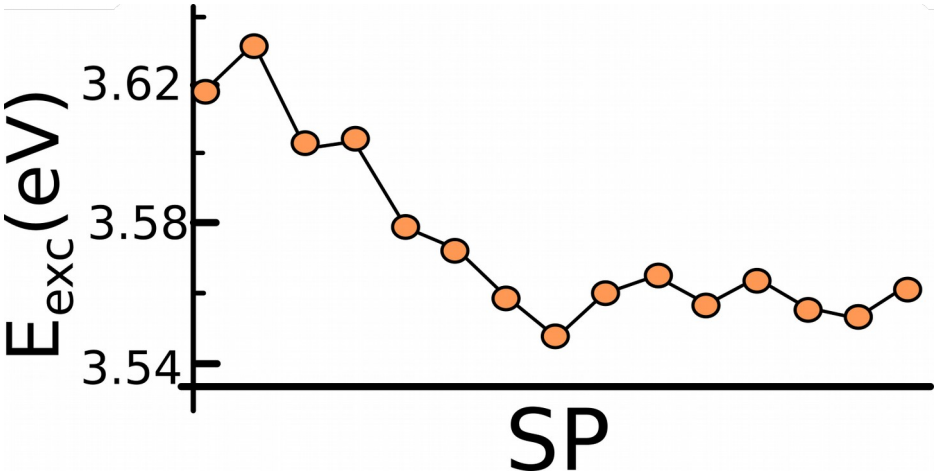

Figure 3.2: Convergence of the averaged excitation energy over an increasing number of single point calculations (SP), from 2 (first point of the graph) to 15 (last point of the graph). The excitation energies are reported in eV.

The configurations obtained from the classical MD are then used as new starting points for polarizable hybrid QM/AMOEBA dynamics. For this reason, the QND molecule is frozen in its initial crystal geometry, avoiding in this way the requirement of a coherent derivation of the AMOEBA parameters, in particular for the bonding terms. Moreover, the anionic QND, obtained by deprotonation of its hydroxyl group, transferring the hydrogen to the His251, would require a different set of FF parameters.

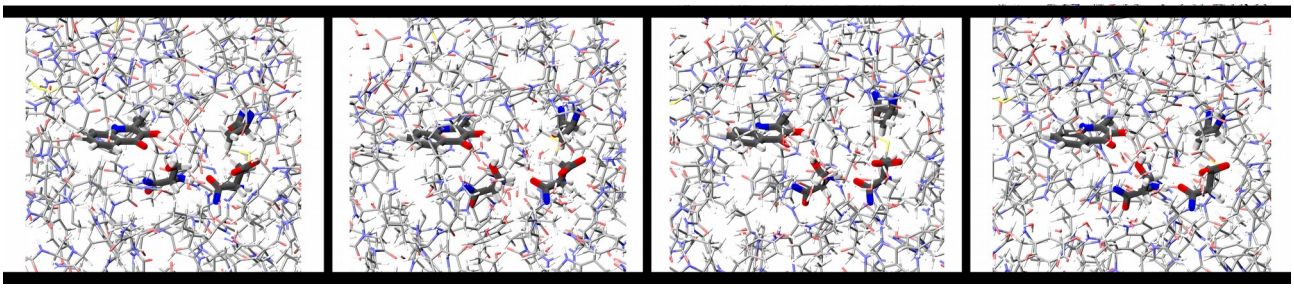

Figure 3.3: few configurations of the modified catalytic triad in the mutant protein, with the neutral QND chromophore; the His251 is substituted by an alanine[13,14].

For both the wild type HOD and the modified H251A, an initial minimization of the water-box is performed, keeping the protein fixed; this optimization is followed first by the minimization of the protein in a frozen solvent and after by the simultaneous minimization of all atoms positions at the same time. In all these steps the QND has been kept frozen. An initial thermalization is performed using the Berendsen thermostat, as implemented in the Tinker-HP package[12]. A series of 200 ps dynamics in the NVT ensemble is then performed, incrementing the temperature by 50K, from 0 to 300K.

Before the actual production, the system was further equilibrated in the NPT ensemble at 1 atm for 5 ns, so to equilibrate the solvation box size. The Monte Carlo barostat is used. Also in the following 30-ns classical production trajectory, the QND chromophore is kept frozen.

The initial classical trajectories for the two oxidation states of the chromophore are in fact used to sample few initial environment configurations, for the following Born-Oppenheimer (BO) QM/AMOEBA MD simulations, so that the QND nuclear/electron dynamics can be coupled to the environment dynamics.

From each environment configuration, NVE polarizable QM/AMOEBA MDs are run, using free (nonperiodic) boundary conditions. The QM subsystem defined in our hybrid approach is depicted in Figure 3.1, and is treated at the B3LYP/6-31G level, whereas the water

molecules and the protein scaffold are treated at the AMOEBA level, coupled to the QM subsystem through the PE scheme discussed in this contribution.

The classical equations of motion are integrated in a velocity Verlet scheme, using a time step δt = 0.5 fs for a total simulation time of 5 ps. From each QM/AMOEBA trajectory a number of configurations is extracted, using a uniform time step for the extraction. TDDFT excitation are obtained from each configuration, including the environment effects at the TPSSH/6-31+G(d)/QM/AMOEBA level. The final excitation energy is obtained by averaging over the single configurations excitations.

Table 1: average excitation energies and oscillator strengths for the first bright transition computed for QND embedded into the H251A protein variant, where His251 has been substituted by an alanine.

QM/AMOEBA MD label	$<E_{exc}>$ (eV)	<Osc. Strength> (arb. unit.)
500_1	3.57	0.11
500_2	3.62	0.10
2000_1	3.57	0.12
2000_2	3.58	0.12
6000_1	3.70	0.13
8000_1	3.60	0.12
Tot. Average	3.61±0.05	0.12±0.01

As shown in Fig. 3.2, the excitation energy averaged over 8 structures converges already toward an oscillating value, so we considered safe to use 16 structures to estimate the excitation energy from a single QM/AMOEBA MD.

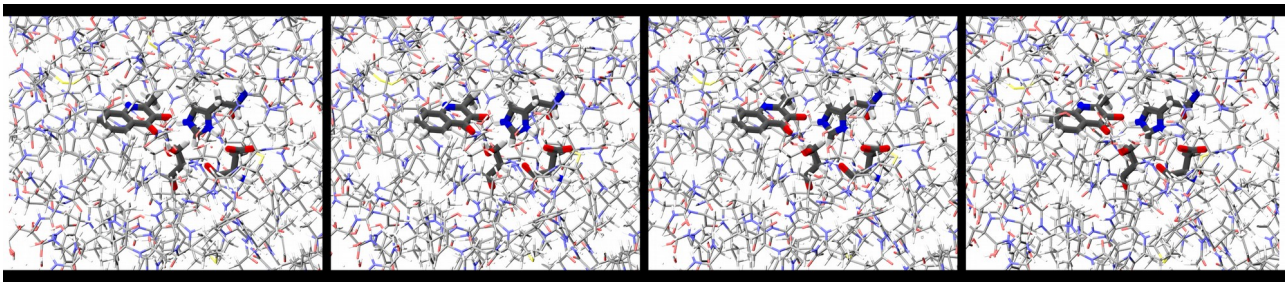

Figure 3.4: few configurations of the catalytic triad in the wild type protein, with the anionic form

The neutral QND absorption is then obtained from the H251A mutated protein, complexed with the QND chromophore. An example of configurations extracted for the QND-H251A electronic absorption simulation is reported in 3.3. The licorice style is used to represent the QM subsystem, while the rest is treated at the AMOEBA level. The simulated results are in good agreement with the experiments, giving an overall (computed from all 6 QM/AMOEBA trajectories) averaged excitation around 3.61 eV (343 nm), as shown in Tab. 1.

The same procedure is followed to obtain the excitation energy of the QND anion. This is achieved mutating the HOD protein so to remove the His-251, which is the first acceptor in the proton transfer reaction which activated the QND substrate, for the following incorporation of an oxygen molecule.[13,14]

In this case, the proton transferred on the proximal His-251, for some initial configurations and during the corresponding QM/AMOEBA MD, is observed to be transferred back to the QND, or form a strongly hydrogen bonded complex. In those cases no excitation could be computed, since the anionic QND does not survive. This is an interesting suggestion on the equilibrium of the process, which we can infer to be not clearly shifted toward the deprotonated QND. Also in the experimental spectrum, the anionic and neutral QND spectra coexist, indicating a not fully converted neutral QND.

Table 2: average excitation energies and oscillator strengths for the first bright transition computed for anion of QND, embedded into the wild type HOD protein.

QM/AMOEBA MD label	<E_{exc}> (eV)	<Osc. Strength> (arb. unit.)
500_1	2.62	0.0
1000_1	Back to neutral QND	-
3000_1	2.84	0.13
5000_2	2.72	0.10
5000_3	2.91	0.09
6000_1	Back to neutral QND	-
6000_2	3.11	0.10
Tot. Average	2.84±0.2	0.10±0.02

We reported the simulation results as average excitation properties in Tab. 2. As compared to the results in Tab 1, we can remark the shift of the QND band toward longer wavelength, where the absorption of the anionic, deprotoneted form of QND is expected.[13,14] Also in this case the simulations are in fair agreement with the experiment.

The final averaged excitation, computed over 5 series of trajectories and 75 structures, is shifted by about 15 nm toward the red, compared to the experiment (437 nm instead of the 420 nm in the experiment). A larger variance is also observed for the anionic form results. This effect could be cured by using more trajectories. We also remind here the the same DFT functional is used for the neutral and anionic QND, but this could also be crafted to improve the results.

3.1.2 Reactivity under extreme conditions

A vast literature is available on the effects of a wide range of temperatures and pressures on the reactivity of a variety of chemicals, so it is out of the scope of this Chapter to give an exhaustive review on the subject. We would like just to point out few references that we find interesting for our discussion here. What we mean with “extreme conditions”, can be

reflected in the examples offered by the photopolimerization of carbon monoxide at high pressures[15], or high temperatures reactions in photoreactors[16].

Another interesting example is offered by the prebiotic chemistry and exobiology, concerned, *e. g.*, by the mechanism of formation of amino acids, sugars and nucleobasese in the interstellar medium. Such compounds have been observed on meteorites surfaces and in simple interstellar molecular clouds, due to the irradiation with ultraviolet light, followed by thermal processes.[17,18] The possibility of direct formation in the interstellar medium of such chemicals opens the hypothesis of the emergence of prebiotic chemistry, outside the early Earth's surface.

In this context, ice-like surfaces represent a possible environment, whose properties and fundamental characteristics could be explored computationally, extending and adapting the many available polarizable MM water models developed and implemented in our research group, to be consistently coupled within a QM/MM scheme developed for these specific problems.

In addition to the environment modelling, an explicit excited state dynamics is required. A scheme to perform such kind of simulations, in the context of the QM/AMOEBA MD and accounting for the nonadiabatic coupling between the different electronic states involved in the dynamics, is under study in our laboratory.

Recently, we have been interested in understanding another kind of external condition, namely the effects of high pressures (HPs), and their favourable effects on the kinetics of Diels-Alder cycloadditions[19], a chemical transformation which can still provides key solutions for the synthesis of (poly)cyclic molecules.

Even though the reaction we investigated does not involve any light-matter interaction, we propose a short review of the main results we have found. It is in fact interesting to mention the effect of the environment as a main actor in interpreting the effect of HPs, as well as

describing the computational procedure followed to model it. The computational protocol is expected to be still valid also for photocemical process, where excited state dynamics are required, with minor adjustments.

HP conditions can be used, in fact, as a catalytic strategy, capable of activating reactions unfeasible in conventional conditions, often increasing chemo-, regio- and stereo-selectivities. [19]

Concerning Diels-Alder reactions, it is well known how the heating of the reaction mixture is typically used to activate loosely reactive species, but it can also shift the equilibrium towards the reverse reaction. Also the use of catalysts can be problematic.

Working under pressure, is often an efficient alternative, but specific set-ups are required in the research laboratories, which, at the best of our knowledge, are not yet standard.[19]

To get some insights into the role of the environment in the catalytic action of high pressure conditions, we took as an example the prototypical Diels-Alder reaction between cyclopentadiene and acrylonitrile. In particular, we showed how the HP conditions affect the environment, which in turn affects the diastereoselectivity of the reaction, which can in principle either follow an *endo* or *exo* mechanism (see Fig. 3.1, Top).

A similar step-by-step protocol described in the previous application, is also applied here. In this case no TDDFT single point calculations are required, since the whole reaction takes place in the electronic ground state of the involved chemicals.

A set of polarizable MDs are performed for each molecular complex reported in Fig. 3.1, (Top), to equilibrate the solvent with frozen optimized species. The reacting species (RS) were solvated by 603 CH_2Cl_2 molecules, running about 2 ns of polarizable MD simulations, using the Bussi–Parrinello algorithm to keep an NPT ensemble.

The solvent equilibration is then followed by full QM/AMOEBA MDs, in order to let the RS nuclear configuration to relax in the presence of the classical environment, to improve the

description of the solute-solvent interactions. The RS representing the main steps of the reaction (see Fig.3.1, Top) are treated at M06-2X/3-21G level, and during the polarizable QM/AMOEBA dynamics they mainly vibrate around their optimized geometries, mutually interacting with the CH_2Cl_2 solvent, described at te the AMOEBA level, according to the variational embedding scheme discussed in Section 2.1.

The effect of the pressure on the reaction kinetics is estimated by performing two sets of MD simulations: in the first set the pressure is set to 1 atm, while in the second at 10 Katm. The temperature is 300 K for all production dynamics[19]. Reaction barriers are estimated in both cases by averaging the energies extracted from an ensemble of indipendent trajectories run for each RS[19]. The energy barriers are computed as a difference between the TS energies and the reactant energies. Those two sets of simulations are performed for both *endo* and *exo* forms.

In Fig. 3.1 (Bottom) we report the different energy diagrams for the 1 atm and 10 Katm reactions. The Tinker-HP[12]/ Gaussian09[11] coupled suite of software was used for the QM/AMOEBA MDs, with a 0.5 fs time step in the NVE ensemble.[19]

Tinker-HP is used also for the purely classical dynamics. We remark here that nowadays, also the GPU version of Tinker-HP is available.[20]

At 10 Katm, TSs are found to be lower in energy than reactants, but with pre-TSs (*vide infra*) still more stable than the corresponding TS (see Figure 3.1, Bottom). This is not the case at 1 atm, where the TSs are higher in energy respect to the corresponding reactants. Here we define as pre-TS additional structures between the reactants and the TS. Such pre-TSs have been determined from the intrinsic reaction coordinate (IRC) connecting reactants with TS.

For determining the reaction barriers for the reactions at 10 Katm, we then assume the pre-TS as the structures at the lowest energy.[19]

The presence of such a “submerged” TS only at 10 Katm, is mainly due to the increase of favourable solute-solvent interactions in the denser molecular packing of the solvent box under the HP conditions, mostly due to solvent polarization.[19]

As a result, the reaction is kinetically favoured by high-pressure conditions: the barrier for the *exo* reaction decreases from being of about 18 kcal/mol at 1 atm, to about 9 kcal/mol at 10 Katm. In the case of the the *endo* reaction, the reaction barrier drops from about 11~kcal/mol at 1 atm to almost 0 kcal/mol at 10 Katm, and the kinetics can be then considered diffusion controlled.

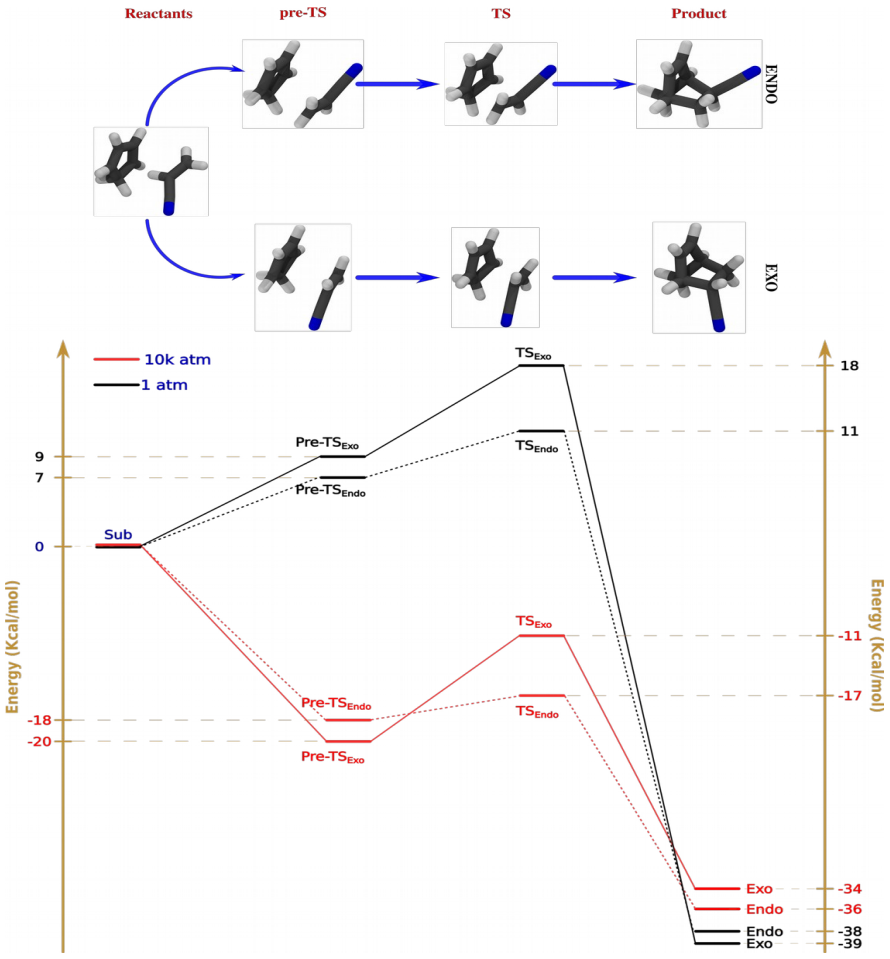

Figure 3.5: Top) Structures of the different reacting molecular species involved in the different stages of the reaction: the reactants, the molecular complex formed by the reactants preceding the transition state (pre-TS), the TS and products of the reactions,

represented in licorice style. Bottom) Energy diagram as from QM/MM simulations. Reproduced from Ref. 19 with permission from RSC, Copyright 2020.

In conclusion, an enhanced diastereoselectivity towards the *endo* product is predicted under HP conditions, due to both a balance between a differential solvation, between *endo* and *exo* pre-TS, and the inverted stability of their respective TS (see Fig. 3.1, Bottom).

4.1 Conclusions

In this contribution, we presented the QM/AMOEBA MD based strategy to simulate complex systems, where chemical transformations and light-matter interactions take place. At first We recalled the main equations and principles of the basics of model, then presenting a new study on the absorption spectra of an organic chromophore in a proteic matrix, and in the end we recalled a reactivity study performed under high pressure conditions. This last example can offer a perspective on the near future application of the model on light-induced reactions in extreme external conditions.

New methods, based on modern models[21] and technologies[20], have been recently developed in our group, from which we can expect a remarkable impact in improved versions of the multiscale approach discussed in this contribution.

ABBREVIATIONS

QC: Quantum Chemistry

FF: Force Field

PE: Polarizable Embedding

MD: Molecular Dynamics

QM/MM MD: Quantum Mechanics/Molecular Mechanics Molecular Dynamics

BO: Born-Oppenheimer

QND: oxoquinaldine

HOD: (1H)-3-hydroxy-4-oxoquinaldine 2,4-dioxygenase

HF: Hartree-Fock

KS: Kohn-Sham

DFT: density functional theory

TDDFT: time-dependent DFT

SCF: self-consistent field

ACKNOWLEDGEMENTS

D.L thanks Dr A. Mezzetti and Prof. A. Steiner for the many very useful discussions on the QND's photophysics.

References

1. D. Bialas, E. Kirchner, M. I. S. Röhr, F. Würthner**,** *J. Am. Chem. Soc.* 2021, 143, 12, 4500–4518.
2. C. Reichardt, Chem. Rev. 1994, 94, 2319–2358
3. D. Loco, N. Gelfand, S. Jurinovich, S. Protti, A. Mezzetti, B. Mennucci**,** J. Phys. Chem. A 2018, 122, 1, 390–397
4. B. Mennucci, S. Corni, *Nat. Rev. Chem.* 2019, 3, 315–330
5. D. Loco, L. Lagardère, O. Adjoua, J.-P. Piquemal, *Acc. Chem. Res.* 2021, 54, 13, 2812–2822.
6. D. Loco, É. Polack, S. Caprasecca, L. Lagardère, F. Lipparini, J.-P. Piquemal, B. Menucci, *J. Chem. Theory Comput.* 2016, 12, 8, 3654–3661

7. D. Loco, L. Lagardère, S. Caprasecca, F. Lipparini, J.-P. Piquemal, *J. Chem. Theory Comput.* 2017, 13, 9, 4025–4033

8. D. Loco, L. Lagardère, G. A. Cisneros, G. Scalmani, M. Frisch, F. Lipparini, B. Mennucci, J.-P. Piquemal, Chem. Sci., 2019, 10, 7200

9. P. Ren, J. Ponder, *J. Phys. Chem. B* 2003, 107, 24, 5933–5947

10. A. Warshel, M. Levitt, J. Mol. Biol. 1976, 103, 227–249.

11. M. J. Frisch et al, Gaussian Development Version, Revision H.36, Gaussian Inc. Wallingford CT 2010.

12. L. Lagardère, L.-H. Jolly, F. Lipparini, F. Aviat, B. Stamm, Z. F. Jing, M. Harger, H. Torabifard, G. A. Cisneros, M. J. Schnieders, N. Gresh, Y. Maday, P. Y. Ren, J. W. Ponder and J.-P. Piquemal, Chem. Sci., 2018, 9, 956–972.

13. S. Bui, R. A. Steiner, Current Opinion in Structural Biology, 2016, 41:109–118

14. A. Hernandez-Ortega , M. G. Quesne , S. Bui , D. P. H. M. Heuts, R. A. Steiner, D. J. Heyes , S. P. de Visser, N. S. Scrutton1, J. Biol. Chem, 2014; 289(12):8620-32.

15. V. Schettinoab, R. Biniab, *Phys. Chem. Chem. Phys.*, 2003,5, 1951-1965.

16. T. D. Svejstrup, A. Chatterjee, D. Schekin, T. Wagner, J. Zach, M. J. Johansson, G. Bergonzini, B. König, *ChemPhotoChem* 2021, *5*, 808.

17. A. Ruf, J. Lange, B. Eddhif, C. Geffroy, L. Le Sergeant d'Hendecourt, P. Poinot, G. Danger, *Astrophys. J.*, 887:L31, 2019.

18. Y. Oba, Y. Takano, H. Naraoka, N. Watanabe, A. Kouchi, *Nat. Com.*, 10:4413, 2019.

19. D. Loco, R. Spezia, F. Cartier, I. Chataigner, J.-P. Piquemal, *Chem. Comm.*,**56**, 6632-6635, 2020.

20. O. Adjoua, L. Lagardère, L.-H. Jolly, A. Durocher, T. Very, I. Dupays, Z. Wang,T. J. Inizan, F. Célerse, P. Renet al.,J. Chem. Theory. Comput., 2021,17, 2034–2053.

21. P. P. Poier, T. Jaffrelot Inizan, O. Adjoua, L. Lagardère, J.-P Piquemal, *J. Phys. Chem. Lett.* 2022, 13, 19, 4381–4388