Geometry optimization in polarizable QM/MM models: the induced dipole formulation

Stefano Caprasecca,[*,†] Sandro Jurinovich,[†] Lucas Viani,[†] Carles E. Curutchet,[‡] and Benedetta Mennucci[*,†]

Dipartimento di Chimica e Chimica Industriale, University of Pisa, Via Risorgimento 35, 56126 Pisa, Italy, and Departament de Fisicoquímica Facultat de Farmàcia, Universitat de Barcelona Av. Joan XXIII s/n, 08028 Barcelona, Spain

E-mail: stefano.caprasecca@for.unipi.it; benedetta.mennucci@unipi.it

Abstract

We present the mathematical derivation and the computational implementation of the analytical geometry derivatives for a polarizable QM/MM model (QM/MMPol). In the adopted QM/MMPol model, the focussed part is treated at QM level of theory, while the remaining part (the environment) is described classically as a set of fixed charges and induced dipoles. The implementation is performed within the ONIOM procedure, resulting in a polarizable embedding scheme which can be applied to solvated and embedded systems and combined with different polarizable force fields available in the literature. Two test cases characterized by strong hydrogen-bond and dipole–dipole interactions, respectively, are used to validate the method with respect to the non-polarizable one. Finally, an application to geometry optimization of the chromophore of Rhodopsin is presented to investigate the impact of including mutual polarization between the QM and the classical parts in conjugated systems.

*To whom correspondence should be addressed
†Dipartimento di Chimica e Chimica Industriale, University of Pisa, Via Risorgimento 35, 56126 Pisa, Italy
‡Departament de Fisicoquímica Facultat de Farmàcia, Universitat de Barcelona Av. Joan XXIII s/n, 08028 Barcelona, Spain

1 Introduction

Nowadays multiscale approaches represent one of the most effective computational strategies to study complex systems made of many interacting moieties. Among the multiscale approaches, one largely used combines a quantum-mechanical description of a specific part of the whole system, whose properties are the object of the research, and a classical approach for the remainder, which is generally indicated as the "environment". This hybrid QM/classical approach can be formulated in many different ways, depending on the selected classical formulation and the interactions between the two parts that are included. Generally speaking, we can identify two different classical formulations, namely that introducing a continuum description in terms of macroscopic properties of the environment, and that treating it as classical particles interacting through MM force fields. The resulting QM/continuum[1–7] and QM/MM[8–15] methods represent very popular approaches in different research fields, ranging from chemistry, to physics, material science and biology, just to quote the main ones. In particular, QM/classical formulations are largely adopted to describe properties and processes which cannot be treated with fully classical models, such as chemical and biochemical reactivity, electronic processes, as well as spectroscopic properties. Their success in these applications is mostly related to the accuracy with which the interactions between the QM and the classical parts are treated, as well as to the availability of analytical derivatives of such interactions with respect to various internal and external perturbations. The derivatives, for instance, are necessary to determine reactant, product and transition state structures and the corresponding (free) energies for chemical reactivity, to calculate the response properties required for simulating spectra, as well as to determine excited-state structures and properties.

Analytical derivatives have been available for both QM/continuum and QM/MM models for many years; however, in the latter case they are usually limited to non-polarizable MM descriptions, although derivatives have been formulated for polarizable QM/MM schemes adopting fluctuating-charges[16,17] or Drude oscillators.[18]

This alternative and very effective way to introduce polarization effects using MM force fields adopts a strategy based on fixed classical charges (or multipolar expansions) and induced dipoles.[19–25]

The latter are generally obtained in terms of atomic isotropic polarizabilities even if more refined descriptions using the full tensorial character of the polarizability have been proposed.[24] While the charges (or the multipoles) are fixed in magnitude, the polarizabilities allow the classical part to respond to the electric field of the solute, generating induced dipole moments, which in turn affect the solute electron density. Such a response represents the mutual polarization between the environment and the solute, and is the key difference between polarizable and non-polarizable discrete models.

As far as we know, no analytical derivatives have been presented so far for polarizable QM/MM approaches based on the induced dipole model (from now on indicated as QM/MMPol approaches).

Here in particular, we shall focus on the QM/MMPol model which has been recently proposed within the framework of the Gaussian code[23] to investigate both ground and excited state properties, and processes including electronic energy transfer (EET) between chromophoric moieties,[26,27] and can also be used together with the Polarizable Continuum Model[28] for a two-layer, fully polarizable, discrete/continuum description of the environment.

This work is concerned with the extension of the QM/MMPol model to evaluate analytical derivatives of the energy with respect to nuclear and MM coordinates, in order to allow for geometry optimizations of embedded systems. The mathematical derivation is presented in Section 2, while the computational implementation into Gaussian[29] is introduced in Section 3. Calculations on test systems are reported in the same Section to verify the performances of the polarizable QM/MMPol treatment when applied to strongly interacting systems, such as hydrogen-bonded and dipole–dipole dimers. Finally (in Section 4), the model is applied to the Rhodopsin system. In that case geometry optimizations at QM/MMpol level are presented and discussed in comparison with crystal data and non polarizable results. All the presented applications show the applicability and the stability of the implemented derivatives as well as their impact in the determination of geometries of systems of increasing complexity.

2 The mathematical derivation

In a very general framework, the energy derivative of a QM/MM system with respect to a generic geometrical parameter λ can be obtained by means of the chain rule, exactly as for a full QM system, namely:

$$E_{\text{QM/MM}}^{\lambda}(\boldsymbol{P}) = \frac{\partial}{\partial \lambda} E_{\text{QM/MM}} + \frac{\partial E_{\text{QM/MM}}}{\partial \boldsymbol{P}} \frac{\partial \boldsymbol{P}}{\partial \lambda} \tag{1}$$

where $\boldsymbol{P}$ refers to the QM density matrix, whose elements are $P_{\mu\nu}$, being μ and ν functions of the basis set used. The term involving the derivatives of the density matrix gives rise to the usual energy-weighted density matrix contribution which is non zero only if λ represents the position of a QM-described nucleus.

The other term, $\frac{\partial}{\partial \lambda} E_{\text{QM/MM}}$, which is the partial derivative of the energy with respect to the position of either a QM or a MM site, can be formulated as:

$$\frac{\partial}{\partial \lambda} E_{\text{QM/MM}} = \frac{\partial}{\partial \lambda} E_{\text{QM}} + \frac{\partial}{\partial \lambda} E_{\text{MM}} + \frac{\partial}{\partial \lambda} E_{\text{int}} \tag{2}$$

where the first two terms indicate the energy of the QM and of the MM part of the system, whereas the last term is the interaction between the two parts. In the following we shall not consider the QM dervatives, $\frac{\partial}{\partial \lambda} E_{\text{QM}}$, as they are completely identical to those of a full QM system. We note instead, that in non-polarizable MM approaches $\frac{\partial}{\partial \lambda} E_{\text{MM}}$ are completely decoupled from the QM part and their derivatives are exactly the same involved in a full MM system. However, this is not the case for a polarizable MM, as we shall show in the following sections. For such a reason the last two terms will be considered together and will be indicated as derivatives of an effective interaction energy. In particular, from now on, the coordinates of nuclei, MM sites and electrons will be referred to as $\boldsymbol{r}$, $\boldsymbol{R}$ and $\boldsymbol{\rho}$, respectively, while indices n and k will be used to label nuclei and MM sites. When dealing separately with charged and polarizable MM sites, the indices c and p will be used instead of the generic k. The coordinate component will be generically indicated as $\zeta = \{x, y, z\}$. Derivatives will be either written explicitly, or as a superscript with the derivation

variables in brackets.

2.1 The effective interaction energy

The effective interaction energy $E_{MM} + E_{int}$ is composed of an electrostatic term, E^{es}, and a polarization term, E^{pol}:

$$E^{es} = U_q^{es} + U_n^{es} + U_e^{es} \tag{3}$$

$$E^{pol} = \frac{1}{2}\left[U_{en}^{pol} + U_{ne}^{pol} + U_{ee}^{pol} + U_{nn}^{pol}\right] \tag{4}$$

The electrostatic terms account for the interaction between MM charges and the potential induced by the other MM charges (U_q^{es}), the nuclei (U_n^{es}), and the electron density (U_e^{es}):

$$U_q^{es} = \frac{1}{2}\sum_c\sum_{c'}^{*}\frac{q_c q_{c'}}{|\mathbf{R}_c - \mathbf{R}_{c'}|} \tag{5}$$

$$U_n^{es} = \sum_c\sum_n\frac{q_c Z_n}{|\mathbf{R}_c - \mathbf{r}_n|} \tag{6}$$

$$U_e^{es} = \mathrm{tr}(\mathbf{P}\mathbf{h}^{es}) = \sum_{\mu\nu}P_{\mu\nu}\sum_c q_c V_c^{\mu\nu} = \sum_c q_c\sum_{\mu\nu}P_{\mu\nu}\langle\mu|\frac{-1}{|\mathbf{R}_c - \boldsymbol{\rho}|}|\nu\rangle \tag{7}$$

where $\{q_c\}$ are the MM charges, $\{Z_n\}$ the nuclear ones, μ and ν indices labelling the basis functions, and $P_{\mu\nu}$, and $V_c^{\mu\nu}$ are the (μ,ν) elements of the electronic density and of the electrostatic potential on the charged site c.

The star symbol over the summation in Eq. (12) indicates that the index c' does not run over all MM charges, but only over those that are allowed to interact with the charge c. The star symbol will be used throughout the following text, for both charges and induced dipoles, with the same meaning. Indeed, not all MM sites (charged or polarizable) are in general allowed to interact with each other. This is because polarizable models can be divided into additive and nonadditive, also called interactive, models.[22] In additive models, only intermolecular polarization interactions are explicitly accounted for. In nonadditive models, however, all charges and dipoles are allowed to interact intra- and intermolecularly. Because the true diffuse charge distribution is described by

point polarizabilities, nonadditive models can lead to the so-called polarization catastrophe at short distances. This can be avoided, for instance, turning off polarization interactions among 1–2 and 1–3 neighbours, a scheme adopted in Amber polarizable force fields. Another possibility is to use Thole's smeared dipole tensor, which prevents the polarization catastrophe at short distances by introducing appropriate screening functions. We have implemented several polarization models, both based on additive and nonadditive schemes, Table 1 reports the different treatments among different MMPol models, for what concerns the charge–charge, charge–dipole and dipole–dipole interactions.

The polarization terms of Eq. (4) account for the interaction between the induced dipoles and the electric fields generated by charges, nuclei and electrons; the dipoles induced by these terms are here labelled $\boldsymbol{\mu}^{\text{chg}}$, $\boldsymbol{\mu}^{\text{nuc}}$ and $\boldsymbol{\mu}^{\text{ele}}$. The interaction between induced dipoles is explicitly accounted for during the calculation of the dipoles themselves: since they mutually polarize, they are calculated self-consistently. The four polarization terms are:

$$U_{\text{en}}^{\text{pol}} = \text{tr}(\boldsymbol{P}\boldsymbol{j}^{\text{pol}}) = -\sum_{\mu\nu} P_{\mu\nu} \sum_{p\zeta} \mu_{p\zeta}^{\text{nuc+chg}} E_{p\zeta}^{\mu\nu} \tag{8}$$

$$U_{\text{ne}}^{\text{pol}} = \text{tr}(\boldsymbol{P}\boldsymbol{y}^{\text{pol}}) = -\sum_{\mu\nu} P_{\mu\nu} \sum_{p\zeta} \mu_{p\zeta}^{\mu\nu} E_{p\zeta}^{\text{nuc+chg}} \tag{9}$$

$$U_{\text{ee}}^{\text{pol}} = \text{tr}(\boldsymbol{P}\boldsymbol{X}^{\text{pol}}(\boldsymbol{P})) = -\sum_{\mu\nu} P_{\mu\nu} \sum_{p\zeta} \mu_{p\zeta}^{\text{ele}} E_{p\zeta}^{\mu\nu} \tag{10}$$

$$U_{\text{nn}}^{\text{pol}} = -\sum_{p\zeta} \mu_{p\zeta}^{\text{nuc+chg}} E_{p\zeta}^{\text{nuc+chg}} \tag{11}$$

In the previous equations, the electric fields calculated at the positions of the polarizable sites, $\boldsymbol{R}_p$, are:

$$E_{p\zeta}^{\text{nuc+chg}} = E_{p\zeta}^{\text{nuc}} + E_{p\zeta}^{\text{chg}} = \sum_n \frac{Z_n(\zeta_p - \zeta_n)}{\left|\boldsymbol{R}_p - \boldsymbol{r}_n\right|^3} + \sum_{c'}^* \frac{q_{c'}(\zeta_p - \zeta_{c'})}{\left|\boldsymbol{R}_p - \boldsymbol{R}_{c'}\right|^3} \tag{12}$$

$$E_{p\zeta}^{\mu\nu} = \langle\mu| \frac{(-)(\zeta_p - \zeta)}{\left|\boldsymbol{R}_p - \boldsymbol{\rho}\right|^3} |\nu\rangle \tag{13}$$

and the corresponding induced dipoles on the polarizable sites are obtained through the MMPol

matrix $\boldsymbol{D}$:

$$\mu_{p\zeta}^{\mathrm{x}} = (\boldsymbol{D}\boldsymbol{E}^{\mathrm{x}})_{p\zeta}, \ \mathrm{x} = \{\mathrm{chg}, \mathrm{nuc}, \mathrm{ele}\}. \tag{14}$$

Table 1: Polarization–polarization, polarization–charge and charge–charge interactions allowed in the MMPol models used. (a): Using Thole smeared dipole tensor, linear version: AL Model, Table 3, in Wang et al.;[30] screening factor $a = 2.5874$. (b): Interaction scaled by $\frac{1}{1.2}$. (c): Using standard dipole interaction tensor. (d): Using Thole smeared dipole tensor, linear version: Eq. 9c in van Duijnen and Swart;[31] screening factor $a = 1.7278$. (e): These Pol–Chg interactions can also be included in a Thole scheme if the dipole-induced self-polarization of the molecule is accounted for in the charge fitting.[32]

Model	Atom connectivity	Pol–Pol	Pol–Chg	Chg–Chg
Amber12	1–2 neighbours:	Excluded	Excluded	Excluded
	1–3 neighbours:	Excluded	Excluded	Excluded
	1–4 neighbours:	Included[(a)]	Included	Included
	Others:	Included[(a)]	Included	Included
Amber	1–2 neighbours:	Excluded	Excluded	Excluded
	1–3 neighbours:	Excluded	Excluded	Excluded
	1–4 neighbours:	Included[(c)]	Included	Included[(b)]
	Others:	Included[(c)]	Included	Included
Thole	Within same group:	Included[(d)]	Excluded[(e)]	Included
	In different groups:	Included[(d)]	Included	Included
Groups	Within same group:	Excluded	Excluded	Included
	In different groups:	Included[(c)]	Included	Included

2.2   First derivatives with respect to the MM coordinates

In this section the electrostatic and polarization terms of Eq. (5) to Eq. (7) and Eq. (8) to Eq. (11) are derived with respect to the coordinates of the MM sites, $\boldsymbol{R}_k$, in order to obtain analytical expressions for the forces acting upon them.

We first derive the **electrostatic terms** of Eq. (5), Eq. (6) and Eq. (7); as shown, the resulting

force on each of the MM sites is zero if the site is not charged ($q_k = 0$):

$$\frac{\partial}{\partial R_{k\zeta}} U_{\mathrm{q}}^{\mathrm{es}} = -\sum_{c}^{*} \frac{q_k q_c}{|\boldsymbol{R}_k - \boldsymbol{R}_c|^3}(\zeta_k - \zeta_c) = -q_k E_{k\zeta}^{\mathrm{chg}} \tag{15}$$

$$\frac{\partial}{\partial R_{k\zeta}} U_{\mathrm{n}}^{\mathrm{es}} = -q_k \sum_{n} \frac{Z_n}{|\boldsymbol{R}_k - \boldsymbol{R}_n|^2}(\zeta_k - \zeta_n) = -q_k E_{k\zeta}^{\mathrm{nuc}} \tag{16}$$

$$\frac{\partial}{\partial R_{k\zeta}} U_{\mathrm{e}}^{\mathrm{es}} = q_k \sum_{\mu\nu} P_{\mu\nu} \langle\mu| \frac{(-)(\zeta - \zeta_k)}{|\boldsymbol{\rho} - \boldsymbol{R}_k|^3} |\nu\rangle = -q_k E_{k\zeta}^{\mathrm{ele}} \tag{17}$$

We now calculate the derivatives of the polarization terms (Eq. (8) to Eq. (11)). These terms contain a product between the induced dipole moments and the electric field generated by nuclei, charges and electron density at the polarizable sites' positions. The dipoles themselves are induced by nuclei, electron density and MM charges, according to the relation in Eq. (14); their derivatives are:

$$\frac{\partial}{\partial R_{k\zeta}} \mu_{p'\zeta'}^{\mathrm{x}} \equiv \mu_{p'\zeta'}^{\mathrm{x}(k\zeta)} = \frac{\partial}{\partial R_{k\zeta}}(\boldsymbol{D}\boldsymbol{E}^{\mathrm{x}})_{p'\zeta'} = \left(\boldsymbol{D}^{(k\zeta)}\boldsymbol{E}^{\mathrm{x}} + \boldsymbol{D}\boldsymbol{E}^{\mathrm{x}(k\zeta)}\right)_{p'\zeta'} \tag{18}$$

where x = {nuc, chg, ele}, $\boldsymbol{D}^{(k\zeta)}$ indicates the derivative of the MMPol matrix $\boldsymbol{D}$ with respect to the MM sites' positions (calculated later), and $\boldsymbol{E}^{\mathrm{x}(k\zeta)}$ is the shorthand for the derivatives of the electric fields due to x.

The derivatives of the **polarization terms** are:

$$\frac{\partial}{\partial R_{k\zeta}} U_{\mathrm{en}}^{\mathrm{pol}} = -\sum_{\mu\nu} P_{\mu\nu} \sum_{p'\zeta'} \left(\mu_{p'\zeta'}^{\mathrm{nuc+chg}(k\zeta)} E_{p'\zeta'}^{\mu\nu} + \mu_{p'\zeta'}^{\mathrm{nuc+chg}} E_{p'\zeta'}^{\mu\nu(k\zeta)}\right) =$$

$$= -\left(\boldsymbol{E}^{\dagger\mathrm{nuc+chg}}\boldsymbol{D}^{\dagger(k\zeta)}\boldsymbol{E}^{\mathrm{ele}} + \boldsymbol{E}^{\dagger\mathrm{nuc+chg}(k\zeta)}\boldsymbol{D}^{\dagger}\boldsymbol{E}^{\mathrm{ele}} + \boldsymbol{E}^{\dagger\mathrm{nuc+chg}}\boldsymbol{D}^{\dagger}\boldsymbol{E}^{\mathrm{ele}(k\zeta)}\right) \tag{19}$$

$$\frac{\partial}{\partial R_{k\zeta}} U_{\mathrm{ne}}^{\mathrm{pol}} = -\left(\boldsymbol{E}^{\dagger\mathrm{ele}}\boldsymbol{D}^{\dagger(k\zeta)}\boldsymbol{E}^{\mathrm{nuc+chg}} + \boldsymbol{E}^{\dagger\mathrm{ele}(k\zeta)}\boldsymbol{D}^{\dagger}\boldsymbol{E}^{\mathrm{nuc+chg}} + \boldsymbol{E}^{\dagger\mathrm{ele}}\boldsymbol{D}^{\dagger}\boldsymbol{E}^{\mathrm{nuc+chg}(k\zeta)}\right) \tag{20}$$

$$\frac{\partial}{\partial R_{k\zeta}} U_{\mathrm{ee}}^{\mathrm{pol}} = -\left(\boldsymbol{E}^{\dagger\mathrm{ele}}\boldsymbol{D}^{\dagger(k\zeta)}\boldsymbol{E}^{\mathrm{ele}} + \boldsymbol{E}^{\dagger\mathrm{ele}(k\zeta)}\boldsymbol{D}^{\dagger}\boldsymbol{E}^{\mathrm{ele}} + \boldsymbol{E}^{\dagger\mathrm{ele}}\boldsymbol{D}^{\dagger}\boldsymbol{E}^{\mathrm{ele}(k\zeta)}\right) \tag{21}$$

$$\frac{\partial}{\partial R_{k\zeta}} U_{\mathrm{nn}}^{\mathrm{pol}} = -\left(\boldsymbol{E}^{\dagger\mathrm{nuc+chg}}\boldsymbol{D}^{\dagger(k\zeta)}\boldsymbol{E}^{\mathrm{nuc+chg}} + \boldsymbol{E}^{\dagger\mathrm{nuc+chg}(k\zeta)}\boldsymbol{D}^{\dagger}\boldsymbol{E}^{\mathrm{nuc+chg}} + \boldsymbol{E}^{\dagger\mathrm{nuc+chg}}\boldsymbol{D}^{\dagger}\boldsymbol{E}^{\mathrm{nuc+chg}(k\zeta)}\right)$$

$$\tag{22}$$

and the derivatives of the **electric fields**, $E^{\mathrm{x}(k\zeta)}$, are calculated from Eq. (12) and Eq. (13) as

8

follows:

$$\frac{\partial}{\partial R_{k\zeta}}E^{\text{nuc}}_{p'\zeta'} \equiv E^{\text{nuc}(k\zeta)}_{p'\zeta'} = \sum_n Z_n \delta_{kp'}\left[\frac{\delta_{\zeta\zeta'}}{\left|\boldsymbol{R}_{p'}-\boldsymbol{r}_n\right|^3} - \frac{3(\zeta'_{p'}-\zeta'_n)(\zeta_{p'}-\zeta_n)}{\left|\boldsymbol{R}_{p'}-\boldsymbol{r}_n\right|^5}\right] \tag{23}$$

$$\frac{\partial}{\partial R_{k\zeta}}E^{\text{chg}}_{p'\zeta'} \equiv E^{\text{chg}(k\zeta)}_{p'\zeta'} = \sum_{c''}^* q_{c''}(\delta_{kp'}-\delta_{kc''})\left[\frac{\delta_{\zeta\zeta'}}{\left|\boldsymbol{R}_{p'}-\boldsymbol{R}_{c''}\right|^3} - \frac{3(\zeta'_{p'}-\zeta'_{c''})(\zeta_{p'}-\zeta_{c''})}{\left|\boldsymbol{R}_{p'}-\boldsymbol{R}_{c''}\right|^5}\right] \tag{24}$$

$$\frac{\partial}{\partial R_{k\zeta}}E^{\text{ele}}_{p'\zeta'} \equiv E^{\text{ele}(k\zeta)}_{p'\zeta'} = -\sum_{\mu\nu} P_{\mu\nu}\left\langle \mu\right|\delta_{kp'}\left[\frac{\delta_{\zeta\zeta'}}{\left|\boldsymbol{R}_{p'}-\boldsymbol{\rho}\right|^3} - \frac{3(\zeta'_{p'}-\zeta')(\zeta_{p'}-\zeta)}{\left|\boldsymbol{R}_{p'}-\boldsymbol{\rho}\right|^5}\right]\left|\nu\right\rangle \tag{25}$$

In the MMPol implementation, the MMPol matrix $\boldsymbol{D}$ is obtained by inverting a matrix, which we will call $\tilde{\boldsymbol{D}}$, that only depends on the polarizability, geometry and connectivity of the MM sites, and whose expression in terms of such data is known. Therefore, in order to calculate the derivative of $\boldsymbol{D}$, we make use of the relation $\frac{\partial}{\partial R}\boldsymbol{A}^{-1} = -\boldsymbol{A}^{-1}\frac{\partial \boldsymbol{A}}{\partial R}\boldsymbol{A}^{-1}$, so that it is possible to write:

$$\frac{\partial \boldsymbol{D}}{\partial R_{p\zeta}} \equiv \boldsymbol{D}^{(p\zeta)} = -\boldsymbol{D}\frac{\partial \tilde{\boldsymbol{D}}}{\partial R_{p\zeta}}\boldsymbol{D} = -\boldsymbol{D}\tilde{\boldsymbol{D}}^{(p\zeta)}\boldsymbol{D} \tag{26}$$

Matrix $\tilde{\boldsymbol{D}}$ is defined in terms of the dipole tensor; different MMPol models (Amber12, Amber, Thole, Groups) use slightly different definitions for $\tilde{\boldsymbol{D}}$. In general, $\tilde{D}_{p'\zeta',p'\zeta'} = \alpha_{p'}^{-1}$, where $\alpha_{p'}$ is the isotropic polarizability of the p'-th MM site; $\tilde{D}_{p'\zeta',p''\zeta''} = 0$ ($p' = p''$, $\zeta' \neq \zeta''$); and

$$\tilde{D}_{p'\zeta',p''\zeta''} = S^{(3)}_{p'p''}\frac{\delta_{\zeta'\zeta''}}{r^3_{p'p''}} - S^{(5)}_{p'p''}\frac{3(\zeta'_{p'}-\zeta'_{p''})(\zeta''_{p'}-\zeta''_{p''})}{r^5_{p'p''}} \tag{27}$$

when $p' \neq p''$. In the last equation, $S^{(3)}$ and $S^{(5)}$ are screening terms, introduced in the linear version of Thole smeared dipole tensor (for those models employing it — see Table 1) and are defined in terms of a parameter a (also tabulated) and distance parameters $s_{p'p''} = a\sqrt[6]{\alpha_{p'}\alpha_{p''}}$ and $v_{p'p''} = \frac{r_{p'p''}}{s_{p'p''}}$. Refer to Table SI.1 for an expression of $S^{(3)}_{p'p''}$ and $S^{(5)}_{p'p''}$.

The derivative of Eq. (27) with respect to the coordinate of the MM site p — $\tilde{D}^{(p\zeta)}_{p'\zeta',p''\zeta''}$ — is detailed in the Supporting Information. Here we only note that the derived matrix is non-zero only when $p' \neq p''$, and, most importantly, only along the p-th row and the p-th column, since it contains a $(\delta_{pp'} - \delta_{pp''})$ term. This makes the calculation and handling of the matrix quite

9

inexpensive computationally.

Summing all the electrostatic and polarization terms, we obtain an expression for the **total force on each MM site**:

$$F_{k\zeta} = \frac{\partial E}{\partial R_{k\zeta}} = \frac{\partial}{\partial R_{k\zeta}} \left(U^{\text{es}} + U^{\text{pol}} \right) =$$
$$= -q_k E_{k\zeta} - \frac{1}{2} \left(\boldsymbol{E}^{\dagger(k\zeta)} \boldsymbol{D}^\dagger \boldsymbol{E} + \boldsymbol{E}^\dagger \boldsymbol{D}^{\dagger(k\zeta)} \boldsymbol{E} + \boldsymbol{E}^\dagger \boldsymbol{D}^\dagger \boldsymbol{E}^{(k\zeta)} \right) \tag{28}$$

where $\boldsymbol{E} = \boldsymbol{E}^{\text{chg}} + \boldsymbol{E}^{\text{nuc}} + \boldsymbol{E}^{\text{ele}}$.

Employing the definition of induced dipole $\boldsymbol{\mu} = \boldsymbol{D}\boldsymbol{E}$, the symmetry property of the MMPol matrix ($\boldsymbol{D}^\dagger = \boldsymbol{D}$) and Eq. (26), this expression can be simplified as follows:

$$F_{k\zeta} = -q_k E_{k\zeta} - \frac{1}{2} \left(\boldsymbol{E}^{\dagger(k\zeta)} \boldsymbol{D}\boldsymbol{E} + \boldsymbol{E}^\dagger (-\boldsymbol{D}^\dagger \tilde{\boldsymbol{D}}^{\dagger(k\zeta)} \boldsymbol{D}^\dagger) \boldsymbol{E} + \boldsymbol{E}^\dagger \boldsymbol{D}^\dagger \boldsymbol{E}^{(k\zeta)} \right) =$$
$$= -q_k E_{k\zeta} - \boldsymbol{\mu}^\dagger \boldsymbol{E}^{(k\zeta)} + \frac{1}{2} \boldsymbol{\mu}^\dagger \tilde{\boldsymbol{D}}^{\dagger(k\zeta)} \boldsymbol{\mu} \tag{29}$$

2.3 First derivatives with respect to the nuclear coordinates

In this section the derivatives are carried out with respect to the nuclear coordinates, in order to obtain additional forces due to the presence of the classical polarizable environment.

Firstly, we calculate the derivatives of the **electrostatic terms**. Note that the charge-charge interaction term, U_{q}^{es}, does not depend on $\boldsymbol{r}$:

$$\frac{\partial}{\partial r_{n\zeta}} U_{\text{q}}^{\text{es}} = 0 \tag{30}$$

$$\frac{\partial}{\partial r_{n\zeta}} U_{\text{n}}^{\text{es}} = \sum_c q_c Z_n \frac{(\zeta_c - \zeta_n)}{|\boldsymbol{R}_c - \boldsymbol{r}_n|^3} = \sum_c q_c V_c^{\text{nuc}(n\zeta)} \tag{31}$$

$$\frac{\partial}{\partial r_{n\zeta}} U_{\text{e}}^{\text{es}} = \sum_c q_c \sum_{\mu\nu} P_{\mu\nu} \left[\left\langle \frac{\partial \mu}{\partial r_{n\zeta}} \left| \frac{-1}{|\boldsymbol{\rho} - \boldsymbol{R}_c|} \right| \nu \right\rangle + \left\langle \mu \left| \frac{-1}{|\boldsymbol{\rho} - \boldsymbol{R}_c|} \right| \frac{\partial \nu}{\partial r_{n\zeta}} \right\rangle \right] = \sum_c q_c V_c^{\text{ele}(n\zeta)} \tag{32}$$

Secondly, we recall that the MMPol matrix $\boldsymbol{D}$ does not depend on the nuclear coordinates $\boldsymbol{r}$,

and therefore:

$$\frac{\partial}{\partial r_{n\zeta}}\mu^{\mathrm{x}}_{p'\zeta'} \equiv \mu^{\mathrm{x}(n\zeta)}_{p'\zeta'} = \left(\boldsymbol{D}\boldsymbol{E}^{\mathrm{x}(n\zeta)}\right)_{p'\zeta'} \tag{33}$$

It is now possible to derive the four **polarization terms**:

$$\frac{\partial}{\partial r_{n\zeta}}U^{\mathrm{pol}}_{\mathrm{en}} = -\sum_{\mu\nu}P_{\mu\nu}\sum_{p'\zeta'}\left(\mu^{\mathrm{nuc+chg}(n,\zeta)}_{p'\zeta'}E^{\mu\nu}_{p'\zeta'} + \mu^{\mathrm{nuc+chg}}_{p'\zeta'}E^{\mu\nu(n,\zeta)}_{p'\zeta'}\right) =$$

$$= -\left(\boldsymbol{E}^{\dagger\mathrm{nuc}(n\zeta)}\boldsymbol{D}^{\dagger}\boldsymbol{E}^{\mathrm{ele}} + \boldsymbol{E}^{\dagger\mathrm{nuc+chg}}\boldsymbol{D}^{\dagger}\boldsymbol{E}^{\mathrm{ele}(n\zeta)}\right) \tag{34}$$

$$\frac{\partial}{\partial r_{n\zeta}}U^{\mathrm{pol}}_{\mathrm{ne}} = -\left(\boldsymbol{E}^{\dagger\mathrm{ele}(n\zeta)}\boldsymbol{D}^{\dagger}\boldsymbol{E}^{\mathrm{nuc+chg}} + \boldsymbol{E}^{\dagger\mathrm{ele}}\boldsymbol{D}^{\dagger}\boldsymbol{E}^{\mathrm{nuc}(n\zeta)}\right) \tag{35}$$

$$\frac{\partial}{\partial r_{n\zeta}}U^{\mathrm{pol}}_{\mathrm{ee}} = -\left(\boldsymbol{E}^{\dagger\mathrm{ele}(n\zeta)}\boldsymbol{D}^{\dagger}\boldsymbol{E}^{\mathrm{ele}} + \boldsymbol{E}^{\dagger\mathrm{ele}}\boldsymbol{D}^{\dagger}\boldsymbol{E}^{\mathrm{ele}(n\zeta)}\right) \tag{36}$$

$$\frac{\partial}{\partial r_{n\zeta}}U^{\mathrm{pol}}_{\mathrm{en}} = -\left(\boldsymbol{E}^{\dagger\mathrm{nuc}(n\zeta)}\boldsymbol{D}^{\dagger}\boldsymbol{E}^{\mathrm{nuc+chg}} + \boldsymbol{E}^{\dagger\mathrm{nuc+chg}}\boldsymbol{D}^{\dagger}\boldsymbol{E}^{\mathrm{nuc}(n\zeta)}\right) \tag{37}$$

where the derivatives of the electric fields due to x (MM charges, nuclei and electron density), with respect to the nuclear coordinates, have been indicated with $E^{\mathrm{x}(n\zeta)}$. Note that the charge-induced electric field does not depend on the nuclear coordinates $\boldsymbol{r}$. Moreover, the derivatives of the electronic electric field, E^{ele}, do not involve the electric field operator as before, since the latter does not depend on $\boldsymbol{r}$ (compare with Eq. (25), but act on the basis functions instead:

$$\frac{\partial}{\partial r_{n\zeta}}E^{\mathrm{nuc}}_{p'\zeta'} = -Z_n\left[\frac{\delta_{\zeta\zeta'}}{\left|\boldsymbol{R}_{p'} - \boldsymbol{r}_n\right|^3} - \frac{3(\zeta'_{p'} - \zeta'_n)(\zeta_{p'} - \zeta_n)}{\left|\boldsymbol{R}_{p'} - \boldsymbol{r}_n\right|^5}\right] \equiv E^{\mathrm{nuc}(n\zeta)}_{p'\zeta'} \tag{38}$$

$$\frac{\partial}{\partial r_{n\zeta}}E^{\mathrm{chg}}_{p'\zeta'} \equiv E^{\mathrm{chg}(n\zeta)}_{p'\zeta'} = 0 \tag{39}$$

$$\frac{\partial}{\partial r_{n\zeta}}E^{\mathrm{ele}}_{p'\zeta'} = \sum_{\mu\nu}P_{\mu\nu}\left[\left\langle\frac{\partial\mu}{\partial r_{n\zeta}}\left|\frac{(-)(\zeta'_{p'} - \zeta')}{\left|\boldsymbol{R}_{p'} - \boldsymbol{\rho}\right|^3}\right|\nu\right\rangle + \left\langle\mu\left|\frac{(-)(\zeta'_{p'} - \zeta')}{\left|\boldsymbol{R}_{p'} - \boldsymbol{\rho}\right|^3}\right|\frac{\partial\nu}{\partial r_{n\zeta}}\right\rangle\right] \equiv E^{\mathrm{ele}(n\zeta)}_{p'\zeta'} \tag{40}$$

Summing all the contributions from the electrostatic and polarization terms, we obtain the **total**

11

forces acting on the nuclei due to the polarizable environment:

$$F_{n\zeta} = \frac{\partial E}{\partial r_{n\zeta}} = \boldsymbol{q}^\dagger \boldsymbol{V}^{(n\zeta)} - \frac{1}{2} \left(\boldsymbol{E}^{\dagger(n\zeta)} \boldsymbol{D}^\dagger \boldsymbol{E} + \boldsymbol{E}^\dagger \boldsymbol{D}^\dagger \boldsymbol{E}^{(n\zeta)} \right)$$

$$= \boldsymbol{q}^\dagger \boldsymbol{V}^{(n\zeta)} - \boldsymbol{\mu}^\dagger \boldsymbol{E}^{(n\zeta)} \tag{41}$$

3 Computational implementation

3.1 Derivatives within the ONIOM scheme

The calculation of analytical derivatives in QM/MMPol has been implemented in a locally modified version of Gaussian.[29] Such implementation makes use of the ONIOM scheme.[13,14,33–35] The advantage of the implementation of QM/MM geometry optimizations in a polarizable environment within the ONIOM scheme relies on the fact that all non-electrostatic contributions to the forces (bonded and non-bonded ones) are available, and the additional force terms generated by the presence of the MM electrostatic and polarization interactions can be added independently.

Here, we quickly review the ONIOM scheme, in order to justify our approach. According to such a scheme, the total system, which in the ONIOM jargon is indicated as *real* system, is partitioned into an inner region, generally indicated as *model* system, and an outer or *embedding* region. The energy of the total system is calculated following a subtractive scheme:

$$E^{\text{QM/MM}} = E_{\text{real}}^{\text{MM}} + E_{\text{model}}^{\text{QM}} - E_{\text{model}}^{\text{MM}} \tag{42}$$

where the superscripts QM or MM refer to the level of accuracy used in the three different calculations. This general expression further differentiates depending on the type of interaction present between inner and outer regions, namely which embedding scheme is used. The so-called *mechanical embedding* schemes account for such interaction only at MM level. In the *electronic embedding*, the electrostatic interaction between the QM region and the MM charges enters into the QM description, thus polarizing the electron density. Our approach extends the model to a

polarizable embedding scheme, whereby the QM region is polarized both by the MM charges and by the MM induced dipoles; moreover, the dipoles themselves respond to the QM electron density, and they mutually polarize.

In our implementation, such electrostatic and polarization interactions between the two regions have been added to the standard interactions computed within a mechanical embedding ONIOM. Since atomic charges and polarizabilities must be defined and treated consistently, according to the various MMPol models available (see Table 1), we have preferred not to use the standard electrostatic embedding of ONIOM, but to take care of both electrostatic and polarization terms directly in the QM step of the calculations.

Following such a scheme, the total energy of the system is still obtained in terms of three separate calculations but it needs to be reformulated as:

$$E^{\mathrm{QM/MMPol}} = E(I) + E(II) - E(III) = E_{\mathrm{real}}^{\mathrm{MM/noele}} + \left[E_{\mathrm{model}}^{\mathrm{QM}} + E_{\mathrm{real}}^{\mathrm{es}} + E_{\mathrm{real}}^{\mathrm{pol}} \right] - E_{\mathrm{model}}^{\mathrm{MM/noele}} \quad (43)$$

where in both the first step on the real and the third one on the model MM system, the electrostatic interactions are switched off as they are calculated for the full system in the second QM/model step.

The calculation of energy derivatives is also partitioned into the three steps. Here, the symbols $\boldsymbol{X}^{\mathrm{m}}$ and $\boldsymbol{X}^{\mathrm{emb}}$ are used to indicate the generic inner (model) and outer (embedding) region

coordinates, respectively:

$$\frac{\partial E(I)}{\partial \boldsymbol{X}^{\text{emb}}} = \frac{\partial (E_{\text{real}}^{\text{bond}} + E_{\text{real}}^{\text{vdW}})}{\partial \boldsymbol{X}^{\text{emb}}} \tag{44}$$

$$\frac{\partial E(II)}{\partial \boldsymbol{X}^{\text{emb}}} = \frac{\partial (E_{\text{real}}^{\text{es}} + E_{\text{real}}^{\text{pol}})}{\partial \boldsymbol{X}^{\text{emb}}} \tag{45}$$

$$\frac{\partial E(III)}{\partial \boldsymbol{X}^{\text{emb}}} = 0 \tag{46}$$

$$\frac{\partial E(I)}{\partial \boldsymbol{X}^{\text{m}}} = \frac{\partial (E_{\text{real}}^{\text{bond}} + E_{\text{real}}^{\text{vdW}})}{\partial \boldsymbol{X}^{\text{m}}} \tag{47}$$

$$\frac{\partial E(II)}{\partial \boldsymbol{X}^{\text{m}}} = \frac{\partial E_{\text{model}}^{\text{QM}}}{\partial \boldsymbol{X}^{\text{m}}} + \frac{\partial (E_{\text{real}}^{\text{es}} + E_{\text{real}}^{\text{pol}})}{\partial \boldsymbol{X}^{\text{m}}} \tag{48}$$

$$\frac{\partial E(III)}{\partial \boldsymbol{X}^{\text{m}}} = \frac{\partial (E_{\text{model}}^{\text{bond}} + E_{\text{model}}^{\text{vdW}})}{\partial \boldsymbol{X}^{\text{m}}} \tag{49}$$

The procedure followed for the actual implementation of geometry optimization is first to run a mixed ONIOM/MMPol calculation where a polarizable embedding is performed by the MMPol routines; and then to calculate the MMPol force terms of Eq. (29) and Eq. (41) and add them to the ONIOM forces. In those cases where the inner and outer regions are bonded, the atom connecting them is defined *link atom*,[14] and is treated differently in the three ONIOM steps. The QM/MMPol optimization has been implemented consistently. Link atoms are assigned neither charge nor polarizability. The analytical gradients have been checked against numerical finite–difference results and found dicrepancies below the numerical accuracy. A further analysis has been done on computational times: as expected, we observe a general increase in the time needed to compute the forces, which scales linearly with the number of induced dipoles. Of this increase, the most expensive step is the computation of the dipole-induced force on the nuclei, which is responsible of $\sim$65% of the total (the full analysis is reported in the Supporting Information).

3.2 Preliminary tests

In this Section we present the QM/MM geometry optimization of two simple molecular systems, to test our approach before applying it to larger systems. The test systems have been selected so to

represent two prototypical cases of two of the main intermolecular interactions, namely hydrogen-bonding (here represented by a Guanine–Cytosine pair), and dipole-dipole interactions (here represented by a Merocyanine dimer). The aim of these test studies is to compare the QM/MMPol optimizations to those obtained from a full-QM and a well established non-polarizable QM/MM, in order to check whether a polarizable QM/MM approach is appropriate for the determination of geometries and provides reasonable results.

Before applying the polarizable QM/MM scheme, it is necessary to define the charge model used in the MM part of the system. In AMBER polarizable force fields, parameterization of the permanent point charges depends on the polarization model.[30,32,36] The standard procedure for deriving point charges for non-polarizable force fields is to perform an ab initio calculation, evaluate the electrostatic potential on a grid placed around the molecule, and derive the ESP charges with a (possibly restrained) fitting procedure so that they reproduce the QM potential.[37,38] When a polarizable model is used, a set of atomic polarizabilities is included in addition to the set of charges, and the sum of the potential induced by both sets on the grid points must now reproduce the calculated QM potential. We followed the method developed by Cieplak et al.,[32] where the dipole-induced self-polarization of the molecule is taken into account by iteratively fitting the atomic point charges to the difference of the QM potential and that generated by the induced dipoles.

The Guanine–Cytosine system is characterized by three strong hydrogen bonds between the two bases. Their lengths, labelled r_1, r_2 and r_3, and shown in Figure 1, will be used as parameters to compare the optimized structures.
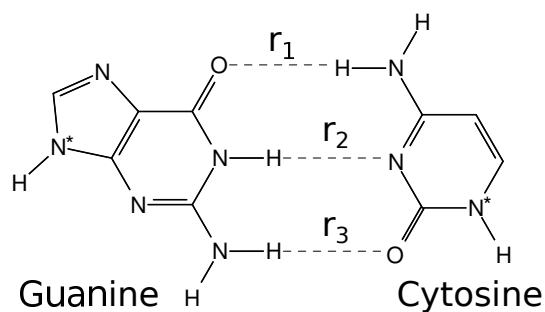

Figure 1: Geometry of the Guanine–Cytosine model pair used in the calculations. The three hydrogen bonds analysed in the text are labelled r_1, r_2 and r_3.

The parameters used for the system are the standard ones of Amber94 force field[39] for the non-polarizable MM, while we employed the atomic polarizabilities and polarization-consistent charges of Amber ff02 force field[32] for the MMPol. In both cases, the charges were reported for the nucleosides, including the sugar, which we do not include in our system (see Figure 1). We proceed by removing the sugar, saturating the N^* atom, and assigning a charge to the saturation hydrogen so that the whole system is neutral.

The reference geometrical parameters (indicated as *Full-QM* in Table 2) were taken from Szalay et al.,[40] who obtained them at MP2/aug-cc-pVDZ level imposing a planarity constraint. Then, two models were used for the QM/MM optimization calculations, differing by which base was treated quantum-mechanically and which classically. These models are named (G^{QM}, C^{MM}) and (C^{QM}, G^{MM}). Furthermore, for each model two different schemes were adopted: one, named *Frozen-MM scheme*, where the internal coordinates of the MM subsystem were not optimized, but were kept fixed at the values obtained from the monomer QM optimization; and another, *Free-MM scheme*, where no contraints were imposed on the coordinates.

Finally, for each model and each scheme, the MM subsystem was either optimized using a standard non-polarizable method (labelled QM/MM in Table 2), or following the polarizable method under testing (QM/MMPol). The results are reported in Table 2.

Table 2: Full-QM, QM/MM and QM/MMPol results on the Guanine–Cytosine pair. Distances are reported in Å.

Model	Parameter	Full-QM scheme	Frozen-MM scheme		Free-MM scheme	
			MM	MMPol	MM	MMPol
G^{QM}, C^{MM}	r_1	1.712	1.788	1.785	1.805	1.800
	r_2	1.869	1.865	1.847	1.874	1.854
	r_3	1.873	1.783	1.741	1.769	1.727
C^{QM}, G^{MM}	r_1	1.712	1.761	1.730	1.766	1.737
	r_2	1.869	1.855	1.840	1.868	1.850
	r_3	1.873	1.823	1.829	1.833	1.844

As it can be seen from the results reported in Table 2, the behaviour of QM/MMPol is very similar to that of the non-polarizable QM/MM and they both satisfactorily reproduce the ab initio intermolecular distances. The addition of the induced dipoles seems to lead to a generalized slight decrease of the H-bond distances with respect to the parallel approach with charges only, but in all

cases the differences are below 5%. Even when we allow for a complete relaxation of the classical unit (Free-MM scheme) a similar agreement is found between polarizable and non-polarizable MM on one side, and between QM/MM and full-QM approaches on the other side.

The second test system, a dimeric H-aggregate of Merocyanine dyes, is characterized by a strong dipole–dipole interaction between the monomers.[41] The monomer, shown in Figure 2(a), is roughly planar. In the dimeric structure, two monomers are stacked face to face, lying on parallel planes. The dimer structural parameters, α, θ and R, defined consistently with Rösch et al.,[41] are illustrated in Figure 2(b).
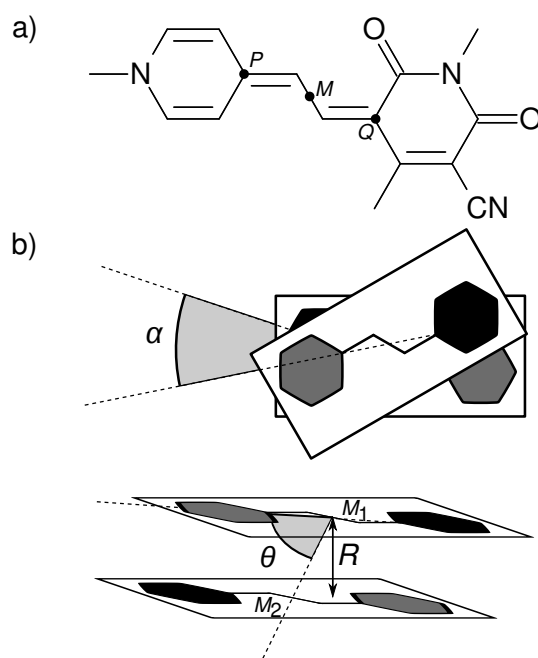
Figure 2: (a) Geometrical structure of the Merocyanine monomer. Points M, P and Q have been arbitrarily chosen for the calculation of the structural parameters. P and Q define the molecular axis, while M is half-way between the two carbon atoms. (b) structural parameters α, θ and R of the Merocyanine dimer. The *rotational angle* α is that between the monomer axes, the *slipping angle* θ is that between the M_1–M_2 line and the axis of the QM monomer, and R the distance between M and the plane of the other monomer. Two additional angles, $\theta_\parallel$ and $\theta_\perp$, named *parallel* and *perpendicular slipping angles*, not shown in the figure, have been defined similarly to θ, but considering the parallel and perpendicular components of M_1–M_2 with respect to the QM molecular axis.

The parameters for the MM subsystem were taken from the General Amber Force Field.[42] The atomic charges were calculated using the standard RESP procedure for the non-polarizable

MM calculation at HF/6-31G* level (consistently with the GAFF procedure), while those for the polarizable MM calculations were derived consistently with the ff02 force field, using the iterative approach presented above[32] (at B3LYP/cc-pVTZ level).

In the Full-QM optimization scheme, the whole dimeric system was optimized at MP2/6-31G(d,p) level, imposing the C_i symmetry, consistently with the study of Rösc et al. The optimized geometry was then used as a starting point for QM/MM and QM/MMPol optimizations, both in the Frozen-MM scheme, where the internal coordinates of the MM region are not modified. The resulting structural parameters are reported in Table 3 and the three geometries are shown in Figure 3.

As for the H-bonded dimer, also in this case the QM/MMPol and the QM/MM schemes give very similar results and both of them are in agreement with the full ab initio structure even if both QM/MM schemes give a larger interplane distance, a smaller shift along the main axis (i.e. a smaller $\theta_\parallel$) and a larger shift perpendicular to that (i.e. a larger $\theta_\perp$). This different packing is graphically shown in Figure 3, where the three structures optimized at the three different levels of description are superimposed. These results can be explained by noting that, in the full-QM calculation, possible dispersion effects are taken into account at a quantum-mechanical level, while they are included in the QM/MM approaches through the empirical Lennard-Jones terms. Apparently the addition of polarization does not significantly change this picture.

These two simple systems clearly show that the QM/MMPol derivatives give reasonable results if compared with the standard QM/MM scheme, showing exactly the same pros and cons. We now move to a more challenging test, i.e., the geometry optimization of Rhodopsin.

Table 3: Geometrical parameters of the Merocyanine dimers studied. Parameters α, θ and R are defined consistently with Rösc et al.[41] Distances are reported in Å and angles in degrees. The full-QM optimization has been performed at C_i symmetry.

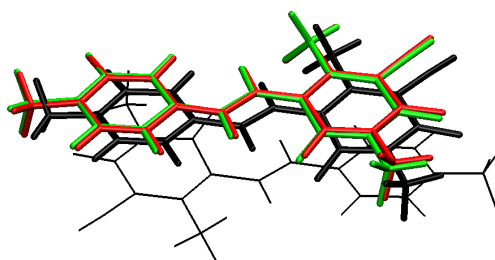| Parameter	Full-QM scheme	Frozen-MM scheme	
		QM/MM	QM/MMPol
α	0.000	1.6	0.2
$\theta_\parallel$	73.1	67.7	66.5
$\theta_\perp$	68.5	80.6	82.2
θ	63.5	66.2	65.5
R	3.134	3.477	3.474

Figure 3: Graphical representation of the three optimized structures of the Merocyanine dimer: full-QM (black), QM/MM (red), QM/MMPol (green).

4 Geometry optimization of Rhodopsin

Rhodopsin is the retinal-protein system involved in the primary events of the complex machinery of vision in the vertebrates. It is able to absorb photons hitting the eye retina and it uses such energy to activate a cascade of chemical reactions responsible for signal transmission.[43,44] The chromophore of Rhodopsin is the protonated Schiff base 11-cis Retinal (PSB11), that is covalently bound to the Lys-296 Nitrogen of the protein. The electronic excitation of the PSB11 promotes the photochemical isomerization converting the 11-cis isomer into the full-trans isomer (PSBT).[45–47] The structure of the 11-cis retinal Schiff base is reported in Figure 4.
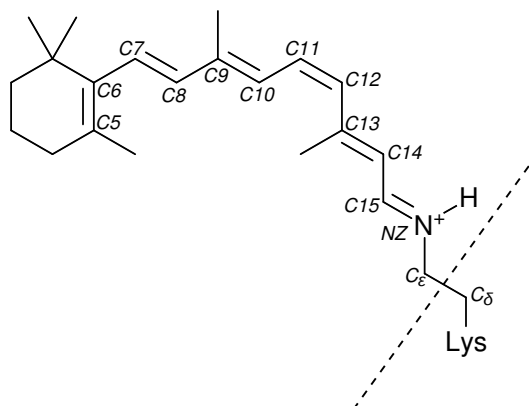

Figure 4: Structure of PSB11. The atoms of the conjugate skeleton involved in the BLA calculations are labelled and numbered; the C_ε Carbon, linked to the NZ Nitrogen, is the link atom between high-level/low-level ONIOM layers. The dotted line separates the atoms allowed to move during the geometry optimization (left) to the fixed ones (right).

19

The retinal is embedded in a pocket located at the centre of the protein barrel and surrounded by a hydrogen-bond network, consisting of functional groups and water molecules. The protein environment modulates the absorption maximum of the chromophore because of its high electronic polarizability and structural flexibility.[48] It is possible to recognize two different effects of the protein environment on the absorption energies: i) a structural effect related to the conformation of the chromophore which is constrained in the protein binding pocket; ii) an electrostatic effect due to the interaction between the chromophore electronic density and the protein electric field.

The spectral tuning of the retinal can be properly explored using hybrid QM/MM approaches. They are capable of including the protein effects both on the structures and on the excited state properties, by combining a quantum-mechanical description of the chromophore with a classical description of the environment.[49–57]

In this section we aim to explore the protein effects on the structural properties of the retinal chromophore, using our hybrid approach that also includes polarization effects. We first present the computational strategy adopted, and then we compare our geometry results with the most recent non-polarizable QM/MM ones.

4.1   Computational details

The initial structure of Rhodopsin was taken from the X-ray resolved structure available for the cis conformer (PDB code: 1U19, resolution = 2.2 Å).[58] The input structures were prepared from the crystallographic data selecting the A chain of the protein dimer. All the residues are considered in their standard protonation state except the Glu-122 and the Asp-83, that are considered neutral, and the three Histidines His-65, His-100 and His-278, that are considered protonated. Such changes were carried out according to the recently reported Rh181(-) structure of Garavelli and coworkers.[55] The total charge of the system is -1. Hydrogen atoms were added and their positions were optimized using AMBER suite of programs.

The low-level calculations, on both the *real system* and the *model subsystem*, follow the subtractive scheme of Eq. (42). The model subsystem includes the chromophore plus the N-H and C_{ε}

20

with its hydrogens of Lys-296, C_δ being the link atom between the inner and outer regions. The atoms in the inner subsystem are described at QM level of theory, using DFT with B3LYP and cam-B3LYP functionals and the 6-31G(d) basis set. During the geometry optimization, the protein environment and the water molecules were kept frozen to limit the computational requirements.

In the non-polarizable MM calculations a standard ONIOM scheme with electrostatic embedding was employed. The MM parameters used for standard residues were those of the Amber94 force field.[39] For the (non-standard) model part the charges were derived with a RESP calculation on the whole PSB11 plus Lys-296, at HF/6-31G* level. The charge of the C_δ link atom was set to zero and the remaining charge that completes the overall charge of the model system to -1 was distributed over the C_ε atom and its hydrogens with a ratio of 1:2 for H:C, following the same strategy of Altun et al.[50] For what concerns the Lys-296 atoms not included in the model system, the standard charges were adjusted in a similar way by setting the C_δ charge to zero and distributing the remaining charge over the C_δ hydrogens, and the C_γ carbon and attached hydrogens. The missing Amber bond parameters for PSB11 and its connection to Lys-296 were derived from the General Amber Force Field.[42]

For the polarizable calculation (following a polarizable embedding scheme as detailed in the previous Sections), we used the same parameters, with the exception of charges, which were derived consistently with the polarization scheme, as described above. The atomic isotropic polarizabilities (model AL) of Wang[30] were used, which are based on Thole's linear model. The environment was partitioned into a polarizable region around the chromophore and a non-polarizable region further. All the environment atoms within 15 Å from any of the chromophore atoms were assigned a charge and a polarizability, while the rest of the environment atoms were only assigned a charge.

4.2   Results and discussion

The optimized PSB11 structure was analysed taking into account three geometrical parameters: two dihedral angles (γ, determining the orientation of the β-ionone ring with respect to the retinal
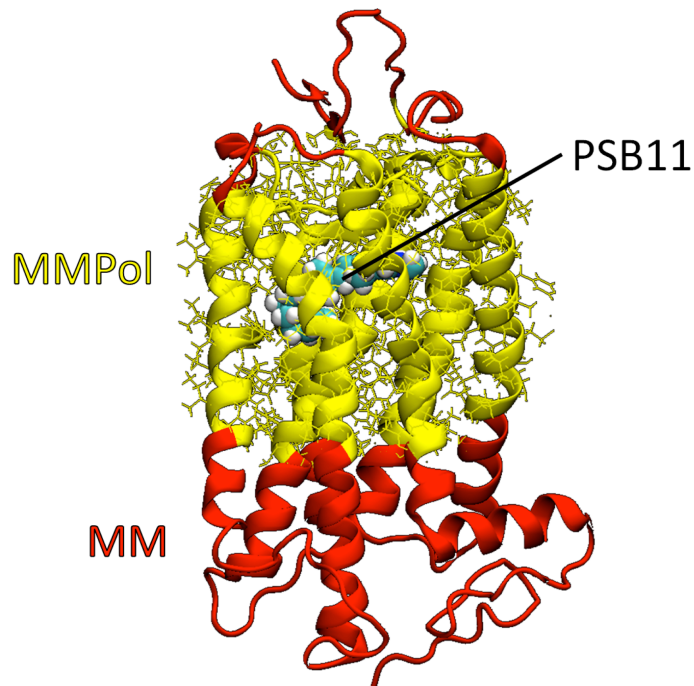
21

Figure 5: Chain A of Rhodopsin, as used for QM/MMPol calculations. The PSB11 atoms are represented as spheres; the polarizable outer region (within 15 Å from the PSB11) is coloured yellow, while the non-polarizable one is coloured red.

chain, and ϕ, related to the torsional orientation of the 11-*cis* double bond), and the Bond Length Alternation (BLA), which is computed as follows:

$$\text{BLA} = \frac{1}{N_\text{s}} \sum_{s=1}^{N_\text{s}} r_s - \frac{1}{N_\text{d}} \sum_{d=1}^{N_\text{d}} r_d$$

where N_s and N_d are the number of single and double bonds considered and r_s and r_d their lengths.

The results are reported in Table 4 for the non-polarizable QM/MM and polarizable QM/MMPol structures using both B3LYP and cam-B3LYP functionals. In Figure 6 a comparison with the two available crystal structures[58,59] is also reported.

As shown in Figure 6, the QM/MMPol optimized structures present the typical "zig-zag" bond length variation along the conjugate chain, in qualitative agreement with both crystal structures.[58,59] A more detailed comparison with the measured bond lengths is however not possible because, in one case, the crystal resolution is too low (2.2 Å), and, in the other case, both the

Table 4: Structural parameters for PSB11 computed at different levels of theory. Dihedral angles are reported in degrees and bond lengths and BLA in Å. The atom numbering is that reported in Figure 4.

Bond	QM/MM	QM/MMPol	QM/MM	QM/MMPol
	B3LYP		cam-B3LYP	
C5–C6	1.359	1.359	1.348	1.348
C6–C7	1.470	1.472	1.477	1.477
C7–C8	1.362	1.360	1.347	1.347
C8–C9	1.449	1.450	1.460	1.459
C9–C10	1.384	1.381	1.366	1.363
C10–C11	1.423	1.429	1.433	1.440
C11–C12	1.384	1.382	1.368	1.365
C12–C13	1.426	1.434	1.433	1.443
C13–C14	1.405	1.401	1.392	1.386
C14–C15	1.395	1.407	1.396	1.410
C15–NZ	1.333	1.326	1.322	1.314
BLA	0.062	0.070	0.082	0.092
$\gamma_{C5-C6-C7-C8}$	−16.7	−14.7	−13.5	−12.7
$\phi_{C10-C11-C12-C13}$	−43.3	−46.0	−46.9	−48.0

conformation and the environment are different, being the data obtained for the *trans* conformer in retinylidene iminium salt. It is therefore more interesting to focus on the differences between computational models. In particular, the B3LYP/MM structure is characterized by a BLA of 0.062 Å, similar to the value obtained by Morokuma and coworkers at the same level of theory (0.064 Å), although they allowed all the protein atoms to move during the optimization.[60] Such values of BLA correspond to a more delocalized structure than that of the two crystals (0.098 and 0.109 Å).

When the QM/MMPol scheme is employed, the structure becomes more localized resulting in an higher BLA value of 0.070 Å (B3LYP/MMPol). This trend is maintained when cam-B3LYP was employed, as a comparable increase in the BLA (from 0.082 to 0.092 Å) was observed moving from cam-B3LYP/MM to cam-B3LYP/MMPol. The last results are in good agreement with the BLA obtained by Tomasello et al.[55] using CASPT2//CASSCF (BLA = 0.110 Å). As expected, the absolute values of the BLA are strongly dependent on the DFT functional and the cam-B3LYP structures are more localized than the corresponding B3LYP ones. However, it is important to stress that a consistent increase in the BLA is observed when moving from MM to MMPol with
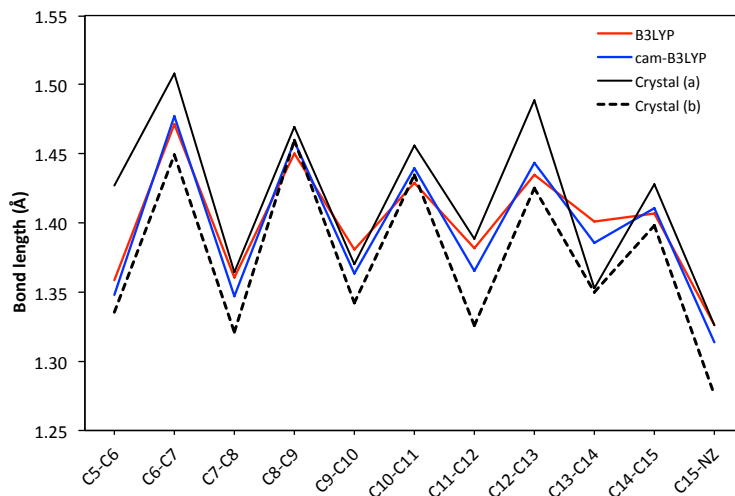
Figure 6: Bond lengths in Å along the conjugate chain of PSB11, resulting from QM/MMPol optimizations at different levels of theory. Crystal structures refer to: (a) PSB11 in Rhodopsin, 2.2 Å resolution;[58] (b) PSBT in retinylidene iminium salt, mÅ resolution.[59] The atom numbering is that reported in Figure 4

both functionals. This shows that the effect of polarization is almost independent on the level of calculation.

Moving to dihedral angles, the optimized structures differ from the crystal data available for the same conformer, in particular for what concerns the orientation of the β-ionone ring, but no significant differences were found between the optimized structures at different levels of theory.

The polarization effects can be further investigated by analysing the interaction energy between the QM system and the induced dipoles (E^{pol} of Eq. (4)). Such quantity is partitioned into contributions due to individual aminoacidic residues surrounding the Retinal. The results are displayed in Figure 7 using different colours to indicate different interaction energies. E^{pol} ranges from about -2 to 2 Kcal/mol; negative values indicate that the QM energy is stabilized by the polarization of the residue, whereas positive values indicate that the opposite is true.

The highest absolute values are observed for those residues that are closer to the NH^+ group of PSB11; residues located on the opposite side of the β-ionone ring have a stabilization contribution, whereas closer residues show positive values of E^{pol}.

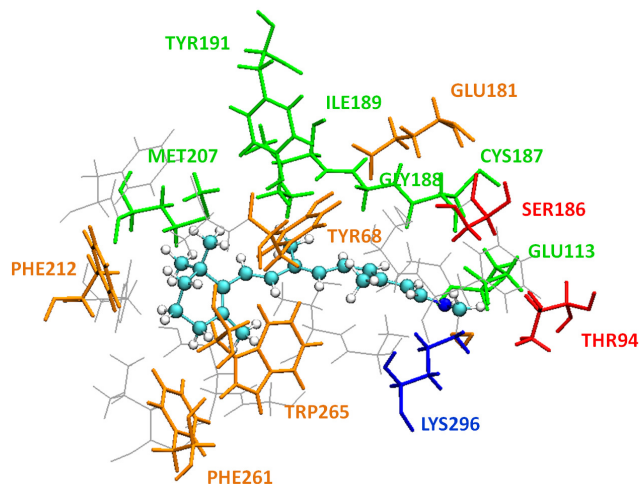To conclude, it appears that using a polarizable MM environment induces small but not negli-

24

Figure 7: Graphical representation of the E^{pol} interaction energy partitioned on the residues surrounding PSB11 in Rhodopsine computed at B3LYP/MMPol level. Residues in grey contribute little; Lys296 is characterised by a stabilization energy of -1.9 Kcal/mol; residues in green and orange belong to the $[-0.6, -0.3]$ and $[0.3, 0.6]$ Kcal/mol ranges, respectively; Thr94 and Ser186 show large 1.9 and 1.2 Kcal/mol contributions. Similar values are obtained at cam-B3LYP/MMPol level.

gible effects, particularly on the bond lengths along the conjugated chain. This is indeed a delicate structural aspect that can largely affect the position and the nature of excited states in lon and highly conjugated systems: even small differences in the BLA values can modify the relative energies of the lowest excited states and induce important mixing in their character.[61]

5   Conclusion

We have formulated analytical derivatives for the polarizable QM/MMPol model based on induced dipoles and implemented them into Gaussian to allow for geometry optimization. The implementation was performed within the ONIOM procedure, resulting in a new ONIOM-polarizable embedding scheme. The method is applicable to many polarizable force fields adopting different schemes for the interactions between 1-2 and 1-3 neighbours.

The results on two test molecules characterized by strong hydrogen-bond and dipole–dipole interactions show a general good agreement between the QM/MMPol and the non-polarizable

QM/MM models, revealing, for such systems, a very small effect induced by the inclusion of polarization in the picture. On the other hand, the comparison of QM/MMPol and QM/MM optimized parameters with the crystal structure for the Retinal molecule in Rhodopsin suggests that a small but quite marked effect is played by the polarizability, resulting in a more "localized" structure of single and double bonds along the conjugation skeleton. This example shows that structural issues of embedded systems require careful handling as they can be affected by the environment in various ways, including through mutual polarization effects. As a matter of fact, we expect that these effects can become even more important when moving to electronically excited states. It is in fact well known that polarizable force fields have a strong influence on the electronic spectra of embedded chromophores (including Retinal[56]) due to the less "compact" electronic density characterizing the excited states: we can therefore expect that the presence of a polarizable environment will be significantly reflected in the relaxed geometries as well as in emission properties and excited-state processes.

Acknowledgement

Supporting Information

Derivative of the MMPol matrix with respect to the coordinate of the MM site. Analysis of computational times. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Rivail, J. L.; Rinaldi, D. *Chem. Phys.* **1976**, *18*, 233–242.

(2) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117 – 129.

(3) Mikkelsen, K.; Agren, H.; Jensen, H.; Helgaker, T. *J. Chem. Phys.* **1988**, *89*, 3086–3095.

(4) Tomasi, J.; Persico, M. *Chem Rev* **1994**, *94*, 2027–2094.

(5) Cramer, C.; Truhlar, D. *Chem. Rev.* **1999**, *99*, 2161–2200.

(6) Orozco, M.; Luque, F. *Chem. Rev.* **2000**, *100*, 4187–4225.

(7) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3093.

(8) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.

(9) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700–733.

(10) Gao, J. *Accounts Chem. Res.* **1996**, *29*, 298–305.

(11) Cui, Q.; Karplus, M. *J Chem Phys* **2000**, *112*, 1133–1149.

(12) Friesner, R. A.; Guallar, V. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389–427.

(13) Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185–199.

(14) Senn, H. M.; Thiel, W. *Angew. Chem. Int. Edit.* **2009**, *48*, 1198–229.

(15) Olsen, J. M. H.; Kongsted, J. In *Chapter 3 - Molecular Properties through Polarizable Embedding*; Sabin, J. R., BrÃd'ndas, E., Eds.; Advances in Quantum Chemistry; Academic Press, 2011; Vol. 61; pp 107 – 143.

(16) Bryce, R. A.; Buesnel, R.; Hillier, I. H.; Burton, N. A. *Chem. Phys. Lett.* **1997**, *279*, 367–371.

(17) Lipparini, F.; Cappelli, C.; Scalmani, G.; De Mitri, N.; Barone, V. *J. Chem. Theory Comput.* **2012**, *8*, 4270–4278.

(18) Boulanger, E.; Thiel, W. *J. Chem. Theory Comput.* **2012**, *8*, 4527–4538.

(19) Thompson, M. A.; Schenter, G. K. *J. Phys. Chem.* **1995**, *99*, 6374–6386.

(20) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580–10594.

(21) Illingworth, C. J. R.; Gooding, S. R.; Winn, P. J.; Jones, G. A.; Ferenczy, G. G.; Reynolds, C. A. *J. Phys. Chem. A* **2006**, *110*, 6487–6497.

(22) Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J. *J. Phys. Condens. Mat.* **2009**, *21*, 333102.

(23) Curutchet, C.; Muñoz Losa, A.; Monti, S.; Kongsted, J.; Scholes, G. D.; Mennucci, B. *J. Chem. Theory Comput.* **2009**, *5*, 1838–1848.

(24) Olsen, J. M.; Aidas, K.; Kongsted, J. *J. Chem. Theory Comput.* **2010**, *6*, 3721–3734.

(25) 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., Jr.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.

(26) Mennucci, B.; Curutchet, C. *Phys. Chem. Chem. Phys.* **2011**, *13*, 11538–11550.

(27) Caprasecca, S.; Curutchet, C.; Mennucci, B. *J. Chem. Theory Comput.* **2012**, *8*, 4462–4473.

(28) Caprasecca, S.; Curutchet, C.; Mennucci, B. *J. Phys. Chem. C* **2013**, *117*, 12423–12431.

(29) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, Maurizio; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J., Gaussian 09 Revision A.1, Gaussian, Inc., Wallingford CT.

(30) Wang, J.; Cieplak, P.; Li, J.; Hou, T.; Luo, R.; Duan, Y. *J. Phys. Chem. B* **2011**, *115*, 3091–9.

(31) van Duijnen, P. T.; Swart, M. *J. Phys. Chem. A* **1998**, *102*, 2399–2407.

(32) Cieplak, P.; Caldwell, J.; Kollman, P. *J. Comp. Chem.* **2001**, *22*, 1048–1057.

(33) Vreven, T.; Morokuma, K.; Farkas, O.; Schlegel, H. B.; Frisch, M. J. *J. Comp. Chem.* **2003**, *24*, 760–9.

(34) Vreven, T.; Frisch, M. J.; Kudin, K. N.; Schlegel, H. B.; Morokuma, K. *Mol. Phys.* **2006**, *104*, 701–714.

(35) Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Montgomery, J. a.; Morokuma, K.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2*, 815–826.

(36) Wang, Z.-X.; Zhang, W.; Wu, C.; Lei, H.; Cieplak, P.; Duan, Y. *J. Comp. Chem.* **2006**, *27*, 781–90.

(37) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, 10269–10280.

(38) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollmann, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620–9631.

(39) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(40) Szalay, P. G.; Watson, T.; Perera, A.; Lotrich, V. F.; Bartlett, R. J. *J. Phys. Chem. A* **2012**, *116*, 6702–10.

(41) Rösch, U.; Yao, S.; Wortmann, R.; Würthner, F. *Angew. Chem. Int. Edit.* **2006**, *45*, 7026–30.

(42) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. a.; Case, D. a. *J. Comp. Chem.* **2004**, *25*, 1157–74.

(43) Kandori, H.; Schichida, Y.; Yoshisawa, T. *Biochemistry (Mosc)* **2001**, *66*, 1197–1209.

(44) Mathies, R. A.; Lugtenburg, J. In *Handbook of Biological Physics*; Stavenga, D. G., de Grip, W. J., Pugh, E. N., Eds.; North Holland: Amsterdam, 2000; Vol. 3; pp 55–90.

(45) Hubbard, R.; Kropf, A. *Proc. Nat. Acad. Sciences* **1958**, *44*, 130–139.

(46) Green, B. H.; Monger, T. G.; Alfano, R. R.; Aton, B.; Callender, R. H. *Nature* **1977**, *269*, 179–180.

(47) Wang, Q.; Schoenlein, R. W.; Peteanu, L. A.; Mathies, R. A.; Shank, C. V. *Science* **1994**, *266*, 422–424.

(48) Wanko, M.; Hoffmann, M.; Strodel, P.; Koslowski, A.; Thiel, W.; Neese, F.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2005**, *109*, 3606–3615.

(49) Hoffmann, M.; Wanko, M.; Strodel, P.; König, P. H.; Frauenheim, T.; Schulten, K.; Thiel, W.; Tajkhorshid, E.; Elstner, M. *J. Am. Chem. Soc.* **2006**, *128*, 10808–10818.

(50) Altun, A.; Yokoyama, S.; Morokuma, K. *J. Phys. Chem. B* **2008**, *112*, 6814–6827.

(51) Huntress, M. M.; Gozem, S.; Malley, K. R.; Jailaubekov, A. E.; Vasileiou, C.; Vengris, M.; Geiger, J. H.; Borhan, B.; Schapiro, I.; Larsen, D. S.; Olivucci, M. *J. Phys. Chem. B* **2013**, *117*, 10053–10070.

(52) Valsson, O.; Campomanes, P.; Tavernelli, I.; Rothlisberger, U.; Filippi, C. *J. Chem. Theory Comput.* **2013**, *9*, 2441–2454.

(53) Sekharan, S.; Morokuma, K. *J. Am. Chem. Soc.* **2011**, *133*, 19052–19055.

(54) Rajamani, R.; Lin, Y.-L.; Gao, J. *J. Comp. Chem.* **2011**, *32*, 854–865.

(55) Tomasello, G.; Olaso-González, G.; Altoè, P.; Stenta, M.; Serrano-Andrés, L.; Merchán, M.; Orlandi, G.; Bottoni, A.; Garavelli, M. *J. Am. Chem. Soc.* **2009**, *131*, 5172–5186.

(56) Soederhjelm, P.; Husberg, C.; Strambi, A.; Olivucci, M.; Ryde, U. *J. Chem. Theory Comput.* **2009**, *5*, 649–658.

(57) Hayashi, S.; Taikhorshid, E.; Schulten, K. *Biophys. J.* **2009**, *96*, 403–416.

(58) Okada, T.; Sugihara, M.; Bondar, A.-N.; Elstner, M.; Entel, P.; Buss, V. *J. Mol. Biol.* **2004**, *342*, 571–583.

(59) Elia, G. R.; Childs, R. F.; Britten, J. F.; Yang, D. S.; Santarsiero, B. D. *Canadian J. Chem.* **1996**, *74*, 591–601.

(60) Altun, A.; Yokoyama, S.; Morokuma, K. *Photochem Photobiol* **2008**, *84*, 845–854.

(61) Knecht, S.; Marian, C. M.; Kongsted, J.; Mennucci, B. *J. Phys. Chem. B* **2013**, *117*, 13808–13815.

Graphic Table of Content