Article Analytic energy, gradient, and hessian of electrostatic embedding QM/MM based on electrostatic potential tted atomic charges scaling linearly with the MM subsystem size

Miquel Huix-Rotllant, Nicolas Ferré

Analytic energy, gradient, and hessian of electrostatic embedding QM/MM based on electrostatic potential fitted atomic charges scaling linearly with the MM subsystem size

Miquel Huix-Rotllant[a)] and Nicolas Ferré

Aix-Marseille Univ, CNRS, ICR, Marseille, France.

Electrostatic potential fitting method (ESPF) is a powerful way of defining atomic charges derived from quantum density matrices fitted to reproduce a quantum mechanical charge distribution in the presence of an external electrostatic potential. These can be used in the Hamiltonian to define a robust and efficient electrostatic embedding QM/MM method. The original formulation of ESPF QM/MM contained two main approximations, namely, the neglect of grid derivatives and the non-conservation of the total QM charge. Here, we present a new ESPF atomic charge operator which solves these drawbacks at virtually no extra computational cost. The new charge operators employ atom-centered grids and conserve the total charge when traced with the density matrix. We present an efficient and easy-to-implement analytic form for the energy, gradient, and hessian that scale linearly with the MM subsystem size. We show that grid derivatives and charge conservation are fundamental to preserve the translational invariance properties of energies and its derivatives and exact conditions to be satisfied by the atomic charge derivatives. As proof of concept, we compute the transition state that leads to the formation of hydrogen peroxide during cryptochrome's reoxidation reaction. Last, we show that the construction of the full QM/MM hessian scales linearly with the MM subsystem size.

I. INTRODUCTION

Quantum mechanics/molecular mechanics (QM/MM) methods are now a routine way to account for the electrostatic interaction between a quantum system and an environment that can be described as an electrostatic scalar external potential.[?][?][?][?][?] Three main strategies can account for this interaction: mechanical embedding, electrostatic embedding, and polarizable embedding.[?][?] In mechanical embedding, the interacting multipoles are fixed both for the quantum and classical subsystems. In electrostatic embedding, the interacting multipoles are varying for the quantum subsystem and fixed for the classical subsystem. In polarizable embedding, both the quantum and classical multipoles are allowed to vary, usually self-consistently.

The ElectroStatic Potential Fitting (ESPF) method is a general method for the accounting of the QM and MM electrostatic interaction that can be formulated in any of the embedding strategies, from mechanical to polarizable.[?][?][?] Indeed, ESPF is based on a multipolar expansion of the quantum density of charges. At the lowest order, it can be used to define either classical atomic partial charges that interact with the classical charges (mechanical embedding)[?] or atomic charge operators[?][?][?] that can be added in the one-electron operators used to solve the QM problem (electrostatic embedding). Being straightforwardly generalized to higher-order multipoles, the ESPF formalism is formally compatible with any type of MM force field and systematically improvable.

The main advantage of the ESPF method over other QM/MM strategies is that the scaling of QM equations is independent of the MM subsystem size.[?] This is possible by defining atomic multipole operators in the QM subsystem based on the fitting of electrostatic integrals computed on a QM-centered regular grid. This implies that the QM subsystem only "knows" about the MM environment by means of its electrostatic potential. Consequently, the QM equations avoid the scaling with the MM atom number for the energy and its derivatives. Indeed, the derivative of the ESPF operators due to a perturbation of the MM environment only has a contribution from the electron density polarization. We recently showed a strategy to efficiently obtain this contribution[?], opening the way to the first complete and efficient calculation of the full QM/MM energy, gradient and hessian in a QM/MM electrostatic embedding framework.[?][?]

The original ESPF formulation in Ref. ? contained two main approximations: (i) the charge of the QM subsystem was not exactly conserved, inherent to the finite size of the grid on which electrostatic integrals are calculated, and (ii) the grid derivative contributions were missing for the energy derivatives. The computation of grid-based atomic charges offer not only an obvious computational advantage for the calculation of electrostatic integrals, but also the fact that atomic charges only depend on the coordinates of the QM subsystem. The latter property has been used several times to derive charges fitted to the electrostatic potential calculated on a grid due to the QM distribution of charges.[?][?][?][?][?][?][?] However, depending how the grid is constructed and whether grid derivatives (with respect to QM atom coordinates) are included, it can have severe consequences for sum-rules of atomic charges and the translational and rotational invariance of atomic charges.[?] In electrostatic embedding QM/MM, this symmetry breaking is introduced in the

————————
[a)] Electronic mail: miquel.huixrotllant@univ-amu.fr

Hamiltonian and the resulting energy derivatives, similar to what it was observed in density-functional theory (DFT) exchange-correlation functional integration on numerical grids.[?][?]

Here, we present a new ESPF formulation based on atom-centered Lebedev grids with no approximations to a given order of multipolar expansion. To the lowest order of this expansion, we define a new atomic charge operator that conserve the total charge of the QM subsystem without employing Lagrange multipliers, and includes grid derivative contributions for its derivatives. In all formulas for constructing the gradient and hessian, we avoid the derivative of the electrostatic kernel pseudoinverse. This boosts the efficiency of the ESPF computations, allowing routine energy, gradient and hessian calculations at the electrostatic embedding QM/MM level.

II. METHODOLOGY

A. Electrostatic Potential fitting method

The main equations of the electrostatic embedding ESPF method were presented elsewhere.[?][?][?][?][?] For completeness, we summarize the main equations for the energy expression, first and second derivatives. In the ESPF method, the electrostatic kernel ($\mathbf{T}$) plays a central role. To the lowest order, it is defined as a Coulomb interaction kernel between a quantum atom B and a grid point k around atom B,

$$T_{k;B} = \frac{1}{|\mathbf{r}_k - \mathbf{r}_B|} \, . \qquad (1)$$

This allows a simple definition of a charge operator based on the on-grid electrostatic integrals to the same order $V_{k;\mu\nu} = \langle \chi_\mu | |r - r_k|^{-1} | \chi_\nu \rangle$, in which χ_μ is an atomic orbital. The ESPF atomic charge operator is thus obtained by solving the linear set of equations,[?]

$$\sum_k^{N_{grid}} \sum_A^{N_{QM}} T_{B;k} w_k T_{k;A} \hat{Q}_{A,\mu\nu} = \sum_k^{N_{grid}} T_{B;k} w_k V_{k;\mu\nu} \, . \qquad (2)$$

Here $\mathbf{w}$ are the weights of the grid, N_{grid} are the number of grid points and N_{QM} are the number of atoms treated at the quantum level of theory. Here, $\hat{Q}$ correspond to operator leading to the electronic net population (for simplicity, hereafter it will be referred as atomic charge operator). The electrostatic kernel forms a rectangular matrix, that can be inverted using the Moore-Penrose inverse technique,[?] which allows us to write the charge operator as,[?]

$$\hat{Q}_{A,\mu\nu} = \sum_k^{N_{grid}} \sum_B^{N_{QM}} (T_{A;k} w_k T_{k;B})^{-1} \sum_{k'}^{N_{grid}} T_{B;k'} w_{k'} V_{k';\mu\nu}$$
$$:= \sum_k^{N_{grid}} [T_w^+]_{A;k} V_{k;\mu\nu} \, . \qquad (3)$$

Hereafter, we use the common notation for pseudoinverse matrices $\mathbf{T}_\mathbf{w}^+ = (\mathbf{T}^\dagger \mathbf{w} \mathbf{T})^{-1} \mathbf{T}^\dagger \mathbf{w}$. Using this definition for the charge operator, we can easily define electronic net populations centred on the quantum atoms A, by tracing the charge operator with the density matrix ($\mathbf{P}$),

$$Q_A = \sum_{\mu\nu}^{N_{AO}} P_{\mu\nu} \hat{Q}_{A;\mu\nu} = \mathrm{Tr}_{AO} \left[\mathbf{P} \hat{\mathbf{Q}}_A \right] \, . \qquad (4)$$

These charges can interact with the classically defined external potential ϕ_A, giving a total energy contribution

$$\Delta E = \sum_A^{N_{QM}} (Z_A - Q_A) \phi_A = \mathrm{Tr}_{QM}[\mathbf{q}\boldsymbol{\phi}] \, , \qquad (5)$$

in which Z_A is the nuclear charge and $q_A = Z_A - Q_A$ is the atomic partial charge of atom A. This energy can be used to define a mechanical embedding approach directly. Alternatively, for an electrostatic embedding scheme in which the density matrix is polarized by the electrostatic environment, one can define an ESPF operator

$$h_{\mu\nu} = \sum_A^{N_{QM}} \phi_A \hat{Q}_{A,\mu\nu} \, , \qquad (6)$$

that can be added in the QM Fock operator.

a. Energy gradient: The first derivative of the ESPF interaction energy[?][?][?] given in eq. 5 with respect to an atom C can be easily written as

$$\Delta E^{x_C} = -\mathrm{Tr}_{QM}[\mathbf{Q}^{x_C}\boldsymbol{\phi}] + \mathrm{Tr}_{QM}[\mathbf{q}\boldsymbol{\phi}^{x_C}] \, . \qquad (7)$$

Hereafter we use the derivative notation $\Delta E^{x_C} \equiv \partial \Delta E / \partial x_C$. Equation 7 has a simple physical interpretation. The first term involves the charge fluctuations due to the fact that the quantum density is polarizable. The second term correspond to the fluctuations of the external field at fixed atomic charges, and has a trivial expression since the external field has a classical analytic form. Here, we concentrate on the first term, involving the derivative of the ESPF charge operator. Deriving eq. 4, we obtain

$$\mathbf{Q}_A^{x_C} = \mathrm{Tr}_{AO}[\mathbf{P}^{x_C}\hat{\mathbf{Q}}_A] + \mathrm{Tr}_{AO}[\mathbf{P}\hat{\mathbf{Q}}_A^{x_C}] \, . \qquad (8)$$

The first term corresponds to the dipole induced by the fluctuations of the density matrix, while the second corresponds to the induced dipole operator at fixed density. The first term requires the solution of the coupled perturbed self-consistent field equations (CPSCF). In general, this has to be solved for all MM and QM atomic motions, although this can be avoided for the computation of gradients.[?] The second term involves the first-order derivative of the ESPF charge operator, which is given

$$\hat{\mathbf{Q}}^{x_C} = \mathbf{T}_w^+ [\mathbf{V}^{x_C} - \mathbf{T}^{x_C} \mathbf{Q}]$$
$$+ (\mathbf{T}^\dagger \mathbf{w} \mathbf{T})^{-1} [\mathbf{T}^\dagger \mathbf{w}]^{x_C} [\mathbf{V} - \mathbf{T} \mathbf{Q}] \, . \qquad (9)$$

Thus, the derivative of the charge operator requires the derivative of the on-grid electrostatic integrals and the electrostatic kernel. On the one hand, the integral derivative have an easy expression

$$V_{k;\mu\nu}^{x_C} = \langle \chi_\mu^{x_C} | \frac{1}{|\mathbf{r} - \mathbf{r}_k|} | \chi_\nu \rangle + \langle \chi_\mu | \frac{1}{|\mathbf{r} - \mathbf{r}_k|} | \chi_\nu^{x_C} \rangle , \quad (10)$$

in which the derivative of the electrostatic kernel have been neglected. Here, χ_μ symbolizes an atomic orbital, and the bra-ket notation for integral should be interpreted as $\langle \chi_\mu || \mathbf{r} - \mathbf{r}_k |^{-1} | \chi_\nu \rangle = \int d\mathbf{r} \chi_\mu^*(\mathbf{r}) |\mathbf{r} - \mathbf{r}_k|^{-1} \chi_\nu(\mathbf{r})$. On the other hand, the derivative of the electrostatic kernel is given by

$$T_{k;B}^{x_C} = \delta_{BC} \frac{x_k - x_B}{|\mathbf{r}_k - \mathbf{r}_B|^3} . \quad (11)$$

Here δ symbolizes a Kronecker's delta. Again, the grid derivatives have been neglected.

    b. Energy hessian: The second derivative of the ESPF interaction energy[?][?] given in eq. 5 with respect to an atom C and an atom D can be written as

$$\Delta E^{x_C y_D} = -\text{Tr}_{\text{QM}}[\mathbf{Q}^{x_C y_D} \boldsymbol{\phi}] + \text{Tr}_{\text{QM}}[\mathbf{q} \boldsymbol{\phi}^{x_C y_D}] \\ + \text{Tr}_{\text{QM}}[\mathbf{Q}^{y_D} \boldsymbol{\phi}^{x_C}] - \text{Tr}_{\text{QM}}[\mathbf{Q}^{x_C} \boldsymbol{\phi}^{y_D}] . \quad (12)$$

The first term corresponds to an induced quadrupole at fixed external potential, the second term corresponds to the second order variations of the external field at fixed atomic charges, and the last two terms correspond to the induced dipoles interacting with a first-order variation of the external potential. The second term has an easy analytic expression, while the third and fourth terms are constructed with information from the energy gradient. Therefore, here we concentrate on the expression of the induced quadrupole, which can be expressed as

$$\mathbf{Q}_A^{x_C y_D} = \text{Tr}_{\text{AO}}[\mathbf{P}^{x_C y_D} \hat{\mathbf{Q}}_A] + \text{Tr}_{\text{AO}}[\mathbf{P} \hat{\mathbf{Q}}_A^{x_C y_D}] \\ + \text{Tr}_{\text{AO}}[\mathbf{P}^{y_D} \hat{\mathbf{Q}}_A^{x_C}] + \text{Tr}_{\text{AO}}[\mathbf{P}^{x_C} \hat{\mathbf{Q}}_A^{y_D}] . \quad (13)$$

For constructing it, the first and second derivatives of the density matrix and the charge operator are required. Similar to the energy gradient, one can avoid the computation of the second derivative of the density matrix, but the first derivative density matrix is required, for which a set of CPSCF equations have to be solved to compute it. For the ESPF hamiltonian, this can be efficiently done for the QM and MM perturbations, avoiding the solution of CPSCF for each MM atom (for further details see Ref. [?]). The expression for the second derivative of the atomic charge operator is given by

$$\hat{\mathbf{Q}}^{x_C y_D} = \mathbf{T}_w^+ [\mathbf{V}^{x_C y_D} - \mathbf{T}^{x_C} \mathbf{Q}^{y_D} - \mathbf{T}^{y_D} \mathbf{Q}^{x_C} - \mathbf{T}^{x_C y_D} \mathbf{Q}] \\ + (\mathbf{T}^\dagger \mathbf{w} \mathbf{T})^{-1} [\mathbf{T}^\dagger \mathbf{w}]^{x_C y_D} [\mathbf{V} - \mathbf{T} \mathbf{Q}] \\ + (\mathbf{T}^\dagger \mathbf{w} \mathbf{T})^{-1} [\mathbf{T}^\dagger \mathbf{w}]^{x_C} [\mathbf{V} - \mathbf{T} \mathbf{Q}]^{y_D} \\ + (\mathbf{T}^\dagger \mathbf{w} \mathbf{T})^{-1} [\mathbf{T}^\dagger \mathbf{w}]^{y_D} [\mathbf{V} - \mathbf{T} \mathbf{Q}]^{x_C} , \quad (14)$$

Similar to the gradient, the second derivative of the atomic charge operator requires the information of the

first-derivative of the operator in addition to the second-derivative of the electrostatic kernel pseudoinverse and the on-grid integrals. On the one hand, the expression for the second-derivative integrals is given by

$$V_{k;\mu\nu}^{x_C y_D} = \langle \chi_\mu^{x_C y_D} | \frac{1}{|\mathbf{r} - \mathbf{r}_k|} | \chi_\nu \rangle + \langle \chi_\mu | \frac{1}{|\mathbf{r} - \mathbf{r}_k|} | \chi_\nu^{x_C y_D} \rangle \\ + \langle \chi_\mu^{x_C} | \frac{1}{|\mathbf{r} - \mathbf{r}_k|} | \chi_\nu^{y_D} \rangle + \langle \chi_\mu^{y_D} | \frac{1}{|\mathbf{r} - \mathbf{r}_k|} | \chi_\nu^{x_C} \rangle , \quad (15)$$

in which the grid derivatives have been neglected. On the other hand, the second derivative of the electrostatic kernel, which is given by

$$T_{k;B}^{x_C y_D} = \delta_{CB} \delta_{DB} \frac{(x_k - x_B)(y_k - y_B) - \delta_{xy} |\mathbf{r}_k - \mathbf{r}_B|^2}{|\mathbf{r}_k - \mathbf{r}_B|^5} . \quad (16)$$

Again, the grid derivatives have been neglected.

B.   ESPF based on atom-centered grids

    Here, we reformulate the original ESPF method by using atom-centered grids.[?] The expressions for the interaction energy (eq. 5) and its first (eqs. 7 and 8) and second derivatives (eqs. 12 and 13) remain valid. Therefore, here we concentrate on the expression of the first and second derivatives of the ESPF charge operator and the extra contributions that arise from the grid derivatives. First of all, we define a grid centred on the QM atoms. We can redefine the previous grid points k by a new grid definition of points of a Lebedev sphere centered on quantum atoms A,

$$\mathbf{r}_k = \mathbf{r}_A + \mathbf{r}_l := \mathbf{r}_{A,l} . \quad (17)$$

In this section, we consider all grid points contribute equally, so that the grid weights are considered all equal to 1. The advantage of using an atom-centred grid is that we can compute the grid derivatives with respect to a quantum atom B is simply given by $\mathbf{r}_{A,l}^{x_B} = \delta_{AB} \mathbf{e}_x$, in which $\mathbf{e}_x$ is the x-axis standard unit vector. Accordingly, all grid derivatives can be expressed as QM atom position derivatives. In this atom-centered grid, the electrostatic kernel (eq. 1) can be re-expressed as

$$T_{A,l;B} = \frac{1}{|\mathbf{r}_{A,l} - \mathbf{r}_B|} . \quad (18)$$

Using this grid, one can re-write the atomic charge operator elements (eq. 4) as

$$\hat{Q}_{A,\mu\nu} = \sum_B^{N_{\text{QM}}} \sum_l^{N_{B,l}} [T_w^+]_{A;B,l} V_{B,l;\mu\nu} \\ := \sum_{B,l} [T_w^+]_{A;B,l} V_{B,l;\mu\nu} . \quad (19)$$

In the second equation, we defined the short-hand sum over B and l for the grid. Indeed, these two sums do not commute, since the number of Lebedev grid points per atom ($N_{B;l}$) is different for each atom.

a. Charge operator gradient: The first derivative of the electrostatic kernels with respect to an atom C are $T^{x_C}_{A,l;C} = -T^{x_C}_{C,l;A}$ are non-zero for $A \neq C$, and zero otherwise. It is worth noting that the "local" electrostatic kernel of atom A is independent of its position $T_{A,l;A} = |\mathbf{r}_l|^{-1}$ and $T^{x_C}_{A,l;A} = 0$ for $\forall$ C. The first derivative kernel in eq. 11 can now be written as

$$T^{x_C}_{A,l;B} = (\delta_{BC} - \delta_{AC}) \frac{x_{A,l} - x_B}{|\mathbf{r}_{A,l} - \mathbf{r}_B|^3} , \qquad (20)$$

including the grid derivatives, which introduces the second delta function in the equation. The first derivatives of the on-grid electrostatic integrals defined on atom-centred grids is given by

$$V^{x_C}_{B,l;\mu\nu} = \langle \chi^{x_C}_\mu | \frac{1}{|\mathbf{r} - \mathbf{r}_{B,l}|} | \chi_\nu \rangle + \langle \chi_\mu | \frac{1}{|\mathbf{r} - \mathbf{r}_{B,l}|} | \chi^{x_C}_\nu \rangle$$
$$+ \delta_{BC} \langle \chi_\mu | \frac{(x - x_{B,l})}{|\mathbf{r} - \mathbf{r}_{B,l}|^3} | \chi_\nu \rangle . \qquad (21)$$

b. Charge operator hessian: The second derivative of the electrostatic kernel previously defined in eq. 16 can now be written as

$$T^{x_C y_D}_{A,l;B} = (\delta_{BC} - \delta_{AC})(\delta_{BD} - \delta_{AD}) \frac{(x_{A,l} - x_B)(y_{A,l} - y_B) - \delta_{xy} |\mathbf{r}_{A,l} - \mathbf{r}_B|^2}{|\mathbf{r}_{A,l} - \mathbf{r}_B|^5} ,$$

thus including three extra terms due to the grid derivatives. The second derivatives of the electrostatic integrals on a grid are given by

$$V^{x_C y_D}_{B,l;\mu\nu} = \langle \chi^{x_C y_D}_\mu | \frac{1}{|\mathbf{r} - \mathbf{r}_{B,l}|} | \chi_\nu \rangle + \langle \chi_\mu | \frac{1}{|\mathbf{r} - \mathbf{r}_{B,l}|} | \chi^{x_C y_D}_\nu \rangle + \langle \chi^{x_C}_\mu | \frac{1}{|\mathbf{r} - \mathbf{r}_{B,l}|} | \chi^{y_D}_\nu \rangle + \langle \chi^{y_D}_\mu | \frac{1}{|\mathbf{r} - \mathbf{r}_{B,l}|} | \chi^{x_C}_\nu \rangle$$
$$+ \delta_{BD} \left[\langle \chi^{x_C}_\mu | \frac{(y - y_{B,l})}{|\mathbf{r} - \mathbf{r}_{B,l}|^3} | \chi_\nu \rangle + \langle \chi_\mu | \frac{(y - y_{B,l})}{|\mathbf{r} - \mathbf{r}_{B,l}|^3} | \chi^{x_C}_\nu \rangle \right] + \delta_{BC} \left[\langle \chi^{y_D}_\mu | \frac{(x - x_{B,l})}{|\mathbf{r} - \mathbf{r}_{B,l}|^3} | \chi_\nu \rangle + \langle \chi_\mu | \frac{(x - x_{B,l})}{|\mathbf{r} - \mathbf{r}_{B,l}|^3} | \chi^{y_D}_\nu \rangle \right]$$
$$+ \delta_{BC} \delta_{BD} \langle \chi_\mu | \frac{(x - x_{B,l})(y - y_{B,l}) - \delta_{xy} |\mathbf{r} - \mathbf{r}_{B,l}|^2}{|\mathbf{r} - \mathbf{r}_{B,l}|^5} | \chi_\nu \rangle . \qquad (22)$$

The last terms both in eq. 20 and 21 arise from the grid derivatives, and can be interpreted as the on-site induced dipole moments due to the variations of atom C. Using these definitions in eq. 9, one obtains the full expression for the gradient of the charge operator including grid derivatives (see eq. S1 in the Supporting Information). Equivalent expressions for the derivative of the charges were obtained by ? using CHELPG charges.[?] [?] However, our formula avoid the pseudoderivative of the inverse of the electrostatic kernel, which makes the derivative for the charges and the implementation simpler and lessens the computational cost for calculating charge derivatives. While the difference is not large for the gradient, it becomes important for the hessian.

The last five terms both in 22 arise from the grid derivative. Using these definitions in eq. 14, one obtains the full expression for the gradient of the charge operator including grid derivatives (see eq. S2 in the Supporting Information).

C. Charge conservation

The ESPF atomic charges have an error inherent to the finite size of the numerical grid. Usually, ESP-type methods conserve the total charge by adding a simple Lagrangian constraint in the solution of the ESP fitting equations. However, this strategy introduces extra electrostatic kernel pseudoinverses in the definition of the charges,[?] [?] which increase the complexity of equations and the computational cost of gradients and hessians.[?] In addition, the least-square fitting of charges with Lagrangian constraint is prone to rank deficiency problems.[?] To avoid this, we propose the definition of a new atomic charge operator that exactly conserves total charge without adding any extra pseudoinverses in the definition,

$$\hat{Q}'_{A,\mu\nu} = \hat{Q}_{A,\mu\nu} + N^{-1}_{QM} \left(S_{\mu\nu} - \sum_B^{N_{QM}} \hat{Q}_{B,\mu\nu} \right) . \qquad (23)$$

As we will show in the rest of this section, this operator conserves the total charge, keeping the computational efficiency of its energy derivatives by rewriting the charge conservation in the form of an average external potential. Hereafter, we use the primed charges as indication of charge-conserved versions using the operator defined in Eq. 23. In this new atomic charge operator, the "spurious charge" errors (i.e., the deviation of the ESPF charge from the total charge) are equivalently shared between all

atoms. For large grids, this difference is close to 0, and the charge operator falls back to the original definition in Eq. 3. The conservation of the total charge in the new operator can be easily verified by tracing this operator with the AO density matrix, thus obtaining

$$Q'_A = Q_A + N_{QM}^{-1} \left(N_{el} - \sum_B^{N_{QM}} Q_B \right), \qquad (24)$$

using the definition of eq. 4 and noting that $Tr_{AO}[\mathbf{PS}] = N_{el}$, where N_{el} is the total number of electrons of the QM subsystem. Therefore, the total charge conservation condition is satisfied exactly,

$$Tr_{QM}[\mathbf{Q'}] = N_{el}. \qquad (25)$$

Using this atomic charge operator in the energy expression (eq. 5) leads to the new interaction energy expression

$$\Delta E' = Tr_{QM}[\mathbf{q}\boldsymbol{\phi}] + \Phi_{av}(Tr_{QM}[\mathbf{Q}] - N_{el}), \qquad (26)$$

in which we defined the average external potential as $\Phi_{av} = N_{QM}^{-1} Tr_{QM}[\boldsymbol{\phi}]$. Similarly, the new definition of the atomic charge operator leads to additional terms in the ESPF hamiltonian (eq. 6),

$$h'_{\mu\nu} = \sum_A^{N_{QM}} (\Phi_{av} - \phi_A) \hat{Q}_{A,\mu\nu} - \Phi_{av} S_{\mu\nu}. \qquad (27)$$

In principle, one could equivalently add last term as a constant $-\Phi_{av} N_{el}$ in the energy. However, the $\Phi_{av} Tr_{QM}[\hat{\mathbf{Q}}]$ alone is large and would disrupt the wavefunction. The two terms together lead to a small contribution in the interaction operator and a correct wavefunction.

    a.   First derivative: The first derivative of the atomic charge-conserving operator is simply given by

$$\hat{Q}'^{x_C}_{A,\mu\nu} = \hat{Q}^{x_C}_{A,\mu\nu} + N_{QM}^{-1} \left(S^{x_C}_{\mu\nu} - \sum_B^{N_{QM}} Q^{x_C}_{B,\mu\nu} \right), \qquad (28)$$

in which the ESPF operator charge derivative is given in eq. 9. The atomic charge-conserving operator derivative now satisfies exactly the following condition for the full atomic charge derivative,

$$Tr_{QM}[\mathbf{Q'}^{x_C}] = \sum_A^{N_{QM}} Tr_{AO}[\mathbf{P}^{x_C}\mathbf{Q'}_A + \mathbf{PQ'}^{x_C}_A] = 0. \qquad (29)$$

It is worth to note that the sum of atomic charges first derivatives at fixed MO coefficient is given by

$$Tr_{QM}\left[\mathbf{Q}'^{(x_C)}\right] = \sum_A^{N_{QM}} Tr_{AO}\left[\mathbf{PQ}'^{x_C}_A\right]$$
$$= Tr_{AO}[\mathbf{PS}^{x_C}], \qquad (30)$$

thus different from 0. For a translational invariant gradient, the following conditions should be satisfied,[?][?][?]

$$\sum_C^{N_{QM}} Tr_{QM}\left[\mathbf{Q}'^{(x_C)}\right] = 0, \qquad (31)$$

and,

$$\sum_C^{N_{QM}} Tr_{QM}\left[\mathbf{Q}^{(x_C)}\right] = 0. \qquad (32)$$

The former is trivially satisfied, while the latter is only satisfied when grid derivatives are included (see below).

    The previous expressions for the first derivatives of the ESPF energy (eq. 7) and operator (eq. 6) have additional terms due to the charge conservation. The first derivative of the energy is given by

$$\Delta E'^{x_C} = Tr_{QM}[\mathbf{Q}^{x_C}(\Phi_{av} - \boldsymbol{\phi}) + \mathbf{q}'\boldsymbol{\phi}^{x_C}]. \qquad (33)$$

It is important to note that the main difference between the charge-conserved gradient and the initial gradient in Eq. 7 is only in the presence of the conserved charges $\mathbf{q}'$ and the average potential, which are straightforwardly calculated at virtually no additional computational cost. In both gradients, the charge operator first derivative is the same operator. The first derivative of the ESPF operator is given by

$$h'^{x_C}_{\mu\nu} = \sum_A^{N_{QM}} \hat{Q}^{x_C}_{A,\mu\nu}(\Phi_{av} - \phi_A) - \Phi_{av} S^{x_C}_{\mu\nu} \qquad (34)$$
$$+ \sum_A^{N_{QM}} \hat{Q}_{A,\mu\nu}(\Phi_{av} - \phi_A)^{x_C} - \Phi^{x_C}_{av} S_{\mu\nu}.$$

It is important to note that only the derivatives of the classical external potential and the AO overlap are required for charge conservation.

    b.   Second derivative: Similarly, the second derivative of the atomic charge-conserving operator is given by

$$Q'^{x_C y_D}_{A,\mu\nu} = Q^{x_C y_D}_{A,\mu\nu} + N_{QM}^{-1} \left(S^{x_C y_D}_{\mu\nu} - \sum_B^{N_{QM}} Q^{x_C y_D}_{B,\mu\nu} \right), \qquad (35)$$

in which the ESPF operator charge derivative is given in eq. 14. The atomic charge-conserving operator derivative now satisfies exactly the following condition,

$$Tr_{QM}[\mathbf{Q}'^{x_C y_D}] = 0. \qquad (36)$$

The second derivatives of the ESPF energy and operator with charge-conserving operators are thus given respectively by

$$\Delta E'^{x_C y_D} = Tr_{QM}[\mathbf{Q}^{x_C y_D}(\Phi_{av} - \boldsymbol{\phi}) + \mathbf{q}'\boldsymbol{\phi}^{x_C y_D}] \qquad (37)$$
$$+ Tr_{QM}[\mathbf{Q}^{x_C}(\Phi_{av} - \boldsymbol{\phi})^{y_D}] + Tr_{QM}[\mathbf{Q}^{y_D}(\Phi_{av} - \boldsymbol{\phi})^{x_C}],$$

and

$$h'^{x_C y_D}_{\mu\nu} = \sum_A^{N_{QM}} \hat{Q}^{x_C y_D}_{A,\mu\nu} (\Phi_{av} - \phi_A) - \Phi_{av} S^{x_C y_D}_{\mu\nu} \quad (38)$$

$$+ \sum_A^{N_{QM}} \hat{Q}^{y_D}_{A,\mu\nu} (\Phi_{av} - \phi_A)^{x_C} - \Phi^{x_C}_{av} S^{y_D}_{\mu\nu}$$

$$+ \sum_A^{N_{QM}} \hat{Q}^{x_C}_{A,\mu\nu} (\Phi_{av} - \phi_A)^{y_D} - \Phi^{y_D}_{av} S^{x_C}_{\mu\nu}$$

$$+ \sum_A^{N_{QM}} \hat{Q}_{A,\mu\nu} (\Phi_{av} - \phi_A)^{x_C y_D} - \Phi^{x_C y_D}_{av} S_{\mu\nu} .$$

D. Derivatives with respect to MM-type atoms

Up to now, we concentrated exclusively on atomic charge and energy derivatives induced by QM perturbations. Although the energy is uniquely defined for MM or QM atoms, the corresponding derivatives lead to different expressions. In Ref. ? , we showed the specific coupled-perturbed (CP) equations for MM atom perturbations, in order to construct the density derivatives with respect to MM atoms, necessary to build up the full Hessian. A new set of CP equations, dubbed Q-vector equations, were constructed to avoid the scaling of CP with the MM size (see Eq. 20 of Ref. ?). The same Q-vector equations are satisfied with the charge-conserving operator, in which Eq. 23 is used as perturbation operator instead. Hereafter, we show the energy gradient and hessian with respect to MM perturbations. In order to distinguish MM- from QM-type atom perturbations, a tilde is used for the former.

a. First derivative: The full gradient at the QM/MM level has two vectors, corresponding to the QM and MM derivatives.? The QM first derivative is given by Eq. 7, while the interaction energy gradient with respect to an MM atom is simply given by

$$\Delta E^{\tilde{x}_C} = Tr_{QM} \left[\mathbf{q}' \boldsymbol{\phi}^{\tilde{x}_C} \right] . \quad (39)$$

This equation stems from the fact that $Q^{x_C}_{A,\mu\nu} = 0$ for any MM atom C, since the electrostatic kernel and the electrostatic grid are independent of the MM atom positions. In addition, the trace of the density derivative with the Fock operator, $Tr_{AO} \left[\mathbf{P}^{\tilde{x}_C} \mathbf{F} \right]$, is substituted by the equivalent term $-Tr_{AO} \left[\mathbf{W} \mathbf{S}^{\tilde{x}_C} \right]$, in which $\mathbf{W}$ is the energy-weighted matrix.? Since $\mathbf{S}^{\tilde{x}_C} = 0$ for MM atoms, this terms cancel out exactly.

b. Second derivative: The full Hessian at the QM/MM level has four blocks, corresponding to the QM-QM, QM-MM, MM-QM and MM-MM derivatives.? ? The QM-QM block is given by Eq. 12. The QM-MM and MM-QM derivatives are equivalent due to Schwarz equality of partial derivatives ($\Delta E^{\tilde{y}_C x_D} = \Delta E^{x_C \tilde{y}_D}$). The MM-QM second derivative with charge conserving oper-

ators is thus given by

$$\Delta E^{\tilde{x}_C y_D} = Tr_{QM} \left[\mathbf{q}' \boldsymbol{\phi}^{\tilde{x}_C y_D} + \mathbf{Q}^{y_D} (\Phi_{av} - \boldsymbol{\phi})^{\tilde{x}_C} \right] , \quad (40)$$

and the MM-MM second derivative by

$$\Delta E^{\tilde{x}_C \tilde{y}_D} = Tr_{QM} \left[\mathbf{q}' \boldsymbol{\phi}^{\tilde{x}_C \tilde{y}_D} + \mathbf{Q}^{\tilde{y}_D} (\Phi_{av} - \boldsymbol{\phi})^{\tilde{x}_C} \right] . \quad (41)$$

III. COMPUTATIONAL DETAILS

The new charge conserving and grid derivative including ESPF model is capable of calculating the energy, its gradient and hessian with respect to all, QM and MM, degrees of freedom. This ESPF v2.0 is currently implemented in a local development version of GAUSSIAN16,? interfaced with a modified version of Tinker 8.7.1.? Unless otherwise stated, all calculations presented below are performed at the B3LYP/6-31G* and Amber99 levels of theory for QM and MM respectively.? ? ? ? ? The implementation has been tested by comparing numerical and analytic, which lead to an average absolute differences between $10^{-3} - 10^{-5}$ a.u. in all cases.

The atom-centered grids have been constructed from Lebedev spheres, in which all weight has been set to 1 and the weight derivatives to 0. Each atom features three spheres of 100 Lebedev points each, that have been equally spaced using the van der Waals radius of the atom. Each grid point inside the van der Waals radius of any of the atom of the molecule is excluded. No scaling of the QM-MM electrostatic interactions was applied, while we maintained the standard scaling factors for the purely MM electrostatic interactions.

The construction of the ESPF operator elements and its derivatives requires the computation of on-grid electrostatic integrals up to the second order in the interaction kernel, as well as integral derivatives up to the first order [see eqs. 21 and 22], which are already available in GAUSSIAN16. The derivatives of the electrostatic kernel matrix $\mathbf{T}$ are formed on the fly and contracted with the integrals and the pseudoinverse matrices. The pseudoinverse $\mathbf{T}^+_w$ has been computed by means of a Cholesky decomposition of $(\mathbf{T}^\dagger \mathbf{w} \mathbf{T})$ before its inversion. The intermediate matrix $B_{B,l} = \sum_A^{N_{QM}} \phi_A [T^+_w]_{A;B,l}$ is stored in memory and then contracted with the integrals with a sum over grid points, leading directly to an energy contribution.

IV. RESULTS AND DISCUSSION

A. Grid effect on the atomic charges

The grid type selected to derive atomic charges fitted to the electrostatic potential largely impacts their values and variations when the molecular geometry is modified.? This was clearly shown by ? in Ref. ? , who
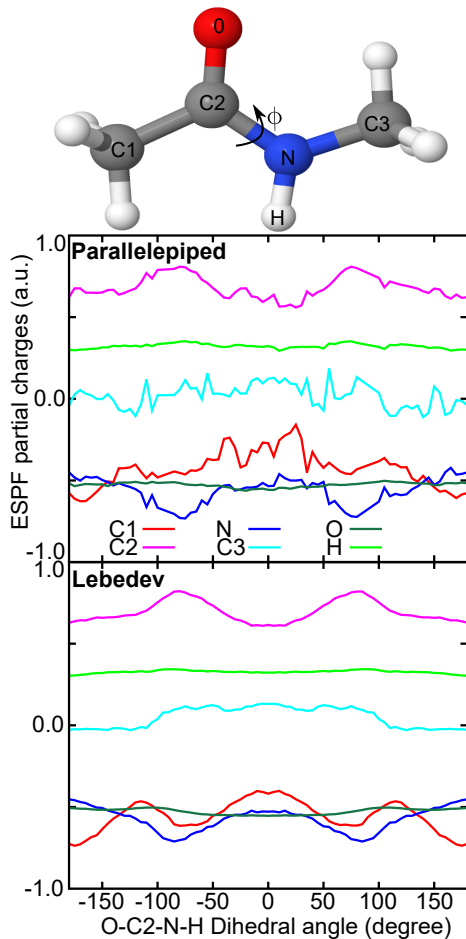
FIG. 1. ESPF partial atomic charges (in a.u.) along the O–C2–N–H dihedral angle of N-methylacetamide using a parallelepiped grid (1777 grid points) or a Lebedev grid (2572 grid points) using charge conserving operators. For simplicity, the charges of the methyl hydrogen atoms have been omitted.

computed atomic charges along one dihedral angle in N-methylacetamide. In Figure **??**, we take the same example and apply it to ESPF charges with (ESPF v1.0) parallelepiped and (ESPF v2.0) Lebedev grids. In this case, no MM atom is present in the calculation. As it can be clearly seen in Figure **??**, the numerical values of each atomic charge remain close whatever the grid type. However, their variations with the ϕ =O–C2–N–H dihedral angle are clearly different. Using the Lebedev grid leads to a rather continuous and smooth change of the charges while the parallelepiped grid introduces an appreciable numerical noise. In addition, the charge variation should be symmetric with respect to $\phi = 0°$, i.e. $q_A(\phi) = q_A(-\phi)$ for any atom A. This property is perfectly satisfied by the Lebedev grid but not by the parallelepiped grid. The reason for these differences is essentially rooted into the symmetry of the numerical grid. Since a Lebedev grid is constructed from spheres centered on nuclei, atomic displacements distort evenly the grid,

hence keeping a consistent overall molecular symmetry. Conversely, a parallelepiped grid is defined using three vectors from the most extreme atom Cartesian coordinates in the three space directions. If the atoms at the extremes of the molecule are kept fixed while other move, the orientation of the grid does not change, thus introducing a numerical noise in the computation of charges (or any other grid-dependent property).

B. Grid effect on the atomic charge derivatives

The consideration of the grid derivatives are important in order to satisfy certain sum rules of the atomic charge derivatives. The sum rule defined in eq. 29 ensures that the total charge of the QM subsystem is kept constant and no electron density spilling effect occurs between the QM and MM subsystems. In Table **??**, we computed this sum rule for a QM/MM water dimer (Figure **??**) using either a parallelepiped grid, hence excluding the grid derivatives, or a Lebedev grid, hence including or excluding the grid derivative terms.
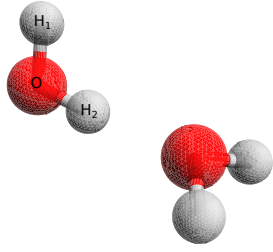
Only the selection of Lebedev grids, including the grid derivative contributions, is able to retain the translational invariance of the QM/MM Hamiltonian. When the grid derivatives are excluded, the corresponding charge and gradient sum rules are only satisfied approximately. The magnitude of the spurious extra charge is independent of the grid size. Indeed, these sum rule deviations are still observed when using a finer grid, with a similar magnitude as for the coarser grids. This shows that the grid derivatives are essential for satisfying the derivative sum rules. If one compares the grid-derivative formulas (eqs. 20 and 21) with the formulas without these contributions (eqs. 11 and 10), a physical interpretation can be given. Grid derivative terms introduce dipolar terms into the electrostatic integral derivatives, resulting in extra dipole-charge type interactions between the perturbed atom and the rest of atomic charges of the QM subsystem.

A consequence of not satisfying the charge derivative sum rule is the breaking of the QM/MM Hamiltonian translational invariance. When this condition is satisfied, it translates to
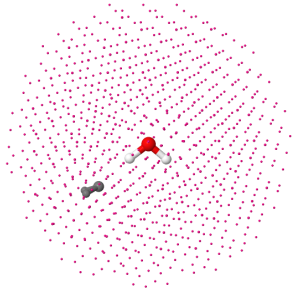
$$\sum_{C}^{N} E^{x_C} = 0. \tag{42}$$

The sum over N runs over all atoms. Table **??**(b) reports values for the gradient translational invariance for the different numerical grids. One could think that the use of atom-centred Lebedev grids, which are invariant under translations by construction, suffice to guarantee the exact translational invariance of the QM/MM Hamiltonian. As it can be observed, this is not the case, irrespective of the grid size. A similar trend is observed for parallelpiped grids. Only Lebedev grigds including grid derivative terms lead to a clear translational invariance of the QM/MM hamiltonian.
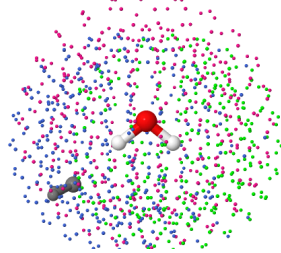
(a) Water dimer structure

(b) Grid types
Parallelpiped

Lebedev GD

Lebedev no GD

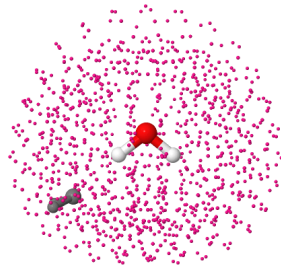FIG. 2. (a) (H$_2$O)$_2$ model used for the QM/MM calculations. The QM atoms are numbered. (b) Parallelpiped and Lebedev grids used to expand the electrostatic integrals. The QM and MM water molecules are shown respectively in colored and grey ball and sticks. For Lebedev grid with grid derivatives (GD), the grid points are colored in magenta, green and blue for grid points centred on O1, H1 and H2 respectively. When all points are in magenta, no atom origin is specified.

As an example of the effect of the grid derivatives, we compare in Table ?? the 12 frequencies characterizing the same water dimer (Figure ??) normal modes obtained with a full QM calculation with those obtained using QM/MM.

Whatever the grid type, most of the QM/MM frequen-

(a) Charge derivative conservation

Grid	# Points	x	y	z
Paral. (no GD)[1]	885	-0.180	-0.321	0.014
	56385	0.012	0.071	0.120
Lebedev (GD)[2]	928	0.000	0.000	0.000
	58895	0.000	0.000	0.000
Lebedev (no GD)[1]	928	0.057	-0.091	0.354
	58895	0.133	-0.001	0.314

(b) Gradient translational invariance

Grid	# Points	x	y	z
Paral. (no GD)[1]	885	18.560	-16.289	20.693
	56385	14.466	-7.813	22.430
Lebedev (GD)[2]	928	0.012	0.014	0.001
	58895	0.012	0.014	0.001
Lebedev (no GD)[1]	928	9.535	-9.859	7.111
	58895	10.067	-8.642	7.334

[1] No grid derivative included [2] Grid derivative included

TABLE I. (a) Deviation from the ideal sum rule of the charge derivatives (Eq. 32) for a QM/MM water dimer model (Figure ??) using different numerical grids including or not the grid derivative (GD) contributions. Units are 10^{-3} charge a.u./Bohr. (b) Sum of each atom QM/MM energy gradient contribution along one space direction. When the translational invariance is satisfied, each sum should be 0 (eq. ??). Units are meV/Å.

Mode	QM	Leb. GD	Leb. no GD	Paral. no GD
1	-0.04	-30.43	-41.68	-23.94
2	0.01	-20.94	-21.41	-16.30
3	0.03	-6.98	-12.45	-11.71
4	2.55	4.22	-4.66	-7.65
5	2.63	20.09	-0.02	10.94
6	17.89	30.01	10.63	29.17
7	59.26	68.78	49.32	69.20
8	165.54	160.80	160.23	160.92
9	180.32	202.77	201.64	199.75
10	218.66	228.55	225.29	227.12
11	434.46	412.38	402.99	404.42
12	659.11	656.41	653.15	655.76
13	1671.71	1708.27	1712.72	1711.37
14	1705.11	2340.07	2340.10	2340.08
15	3698.50	3659.94	3663.02	3660.91
16	3801.66	3706.01	3707.46	3706.44
17	3880.08	3739.61	3739.55	3739.58
18	3903.38	3872.34	3871.81	3871.26

TABLE II. Comparison of the full spectrum of harmonic frequencies for the water dimer (Figure ??) at the minimum energy structure of the full QM model and the QM/MM ESPF models using different grids. The Cartesian coordinates of each minimum energy structure can be found in supporting information. The first 6 vibrations correspond to translations and rotations and should be close to 0. Frequencies have been computed at the B3LYP/6-31G(2d,2p) level. The frequencies are given in cm^{-1}.

cies are very close to each other. Moreover, they are similar to (but not perfectly matching) the QM spectrum. The first 6 vibrations correspond to translations and rotations, and should be close to 0. In this case, the inclusion of grid derivatives brings the translations close to the 0 frequency, better than when grid derivatives are not included. In all QM/MM cases, rotations are not perfectly separated, since this would require the implementatino of a rotational invariant Lebedev grid.[?] For the rest of frequencies, the difference with the QM calculation can be attributed to the quantitatively different force constants in the selected classical force field. However, note that recent developments in polarizable force fields show a better comparison with respect to quantum calculations of force constants.[? ?]

C. Reoxidation of fully reduced anionic flavin adenine dinucleotide in cryptochrome

Cryptochromes (see Figure ??) are flavoproteins acting as blue light receptors, potentially at the origin of magnetoreception,[?] i.e., the sensing of the geomagnetic field. This mechanism could be possibly happening in the protein cavity upon blue light absorption, by generating a biradical between a tryptophan triad and the flavin adenine dinucleotide (FAD).[?] Alternatively, it has been proposed that biradicals can be formed during the reoxidation reaction of fully reduced flavin adendine dinucleotide (FADH$^-$) with molecular oxygen.[?] Recently, one of us showed that such radicals can be stabilized in the protein FAD pocket for a long time, due to strong electrostatic interactions.[?] The complete reaction is complex, since several electron transfers can occur between reduced FAD forms and oxygen when the complex is activated with light.[?] In the hypothesis of a mechanism featuring reactive oxygen species, the biradical should be stable for at least several milliseconds for the magnetoreception to be effective. Here, we study the final step of the oxidation reaction of reduced FAD by molecular oxygen in a cryptochrome protein, namely, the formation of hydrogen peroxide. This reaction step is irreversible, and thus, a rapid formation of H_2O_2 could contradict the possibility of oxygen-flavin-based magnetoreception.

The reaction mechanism considered in this study is depicted in Figure ??. After having optimized the geometry of the reactant, we have been able to locate the corresponding transition state, which corresponds to the simultaneous proton transfer from the secondary amine to one $-OOH$ oxygen, and from the same oxygen atom to the other one in $-OOH$. The QM/MM imaginary frequency of this mode is 1476i cm^{-1} in the protein. The activation free energy at standard conditions of temperature and pressure is 48.58 kcal·mol^{-1}, leading to a reaction rate of $1.59 \cdot 10^{-23}$ s^{-1}. In the gas phase, a similar transition state exists for lumiflavin (7,8,10-trimethylisoalloxazine) with an imaginary frequency of 1356i cm^{-1} and a free energy of activation

Parameter	Lumiflavin	Cryptochrome
O2-O4	1.522	1.588
O4-C6	1.759	2.484
C6-C7	1.501	1.454
H1-N5	1.220	1.657
H1-O2	1.289	1.019
H3-O2	1.302	1.086
H3-O4	1.311	1.147
N5-H1-O2	76.537	82.387

TABLE III. Comparison of main geometrical parameters of the transition states at the QM/MM level for cryptochrome and at the QM level for lumiflavin. Distances are in Å and angles in degrees. Atom numbering is shown in Figure ??. The structures of the transition states can be found in Supporting Information.

of 8.78 kcal·mol^{-1}, leading to a reaction rate of $2.27 \cdot 10^6$ s^{-1}. Indeed, this show that this reoxidation reaction is forbidden, while in gas phase is fast. This shows that reoxidation reaction between oxygen and FAD should occur via a different mechanism in protein (see Ref. ?). In Table ??, the geometries of the two transition states for lumiflavin in gas phase and FAD in cryptochrome are compared (for the Cartesian coordinates and the ESPF charges, see Supporting Information). As it can be seen, most geometrical parameters are similar. The O-O distance is almost the same in both cases, close to the equilibrium value of 1.45 Å for H_2O_2 (computed at the MP2/cc-pVQZ level in the gas phase). In protein, the oxygens are further apart from FAD ($\sim$0.5 Å further from carbon and $\sim$0.4 Å further from nitrogen) with respect to gas phase. The increase of the $C6 - O4$ distance destabilizes the transition state complex, and can be attributed to a redistribution of charge. The charge on O4 net changes by $\sim$-0.12 a.u. of charge between the minimum and the transition state, C6 by $\sim$-0.4 a.u., while C7 is $\sim$+0.14. Other atoms in the ribitol moiety also have a net positive charge increase (see Supporting Information). In addition, the hydrogens are closer to the equilibrium value of 0.97 Å for H_2O_2 (computed at the MP2/cc-pVQZ level in the gas phase) than the gas phase transition state.

D. Computational scaling: chignolin in water droplets

In Figure ??, we report the computational scaling of the QM/MM Hessian with respect to the number of MM atoms. As a test case, we have taken chignolin, a 10 aminoacid protein forming a double beta sheet. This molecule, consisting of 138 atoms treated at the QM level of theory. has been soaked in different microscopic water droplets with a radius of 10, 15, 20 and 25 Å around any of chignolin atoms. For the sake of safe comparisons, the complete geometries have been kept frozen. As it can be seen from the trend line drawn in the graph of Figure ??, the scaling is $\mathcal{O}^{1.22}$, hence close to linear scaling. However, note that Figure ?? reports timings for
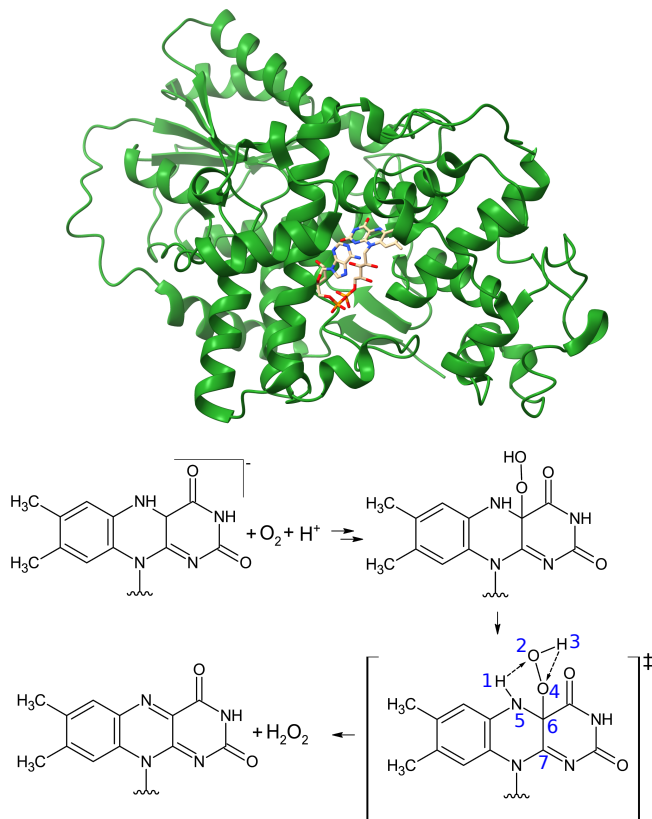
FIG. 3. (top) Structure of *Arabidopsis thaliana* cryptochrome 3 used in the QM/MM model (PDBID: 2J4D). The QM subsystem consists of flavin adenine dinucleotide (in ball and stick), while the rest of the protein has been treated at the MM level. The reduced form of FAD and the hydroperoxyl molecules have been placed in a second step, also treated at the QM level. (bottom) Reoxidation reaction considered in this study. The important atoms are numbered in blue on the representative structure of the transition state. The structures of the transition states can be found in Supporting Information.
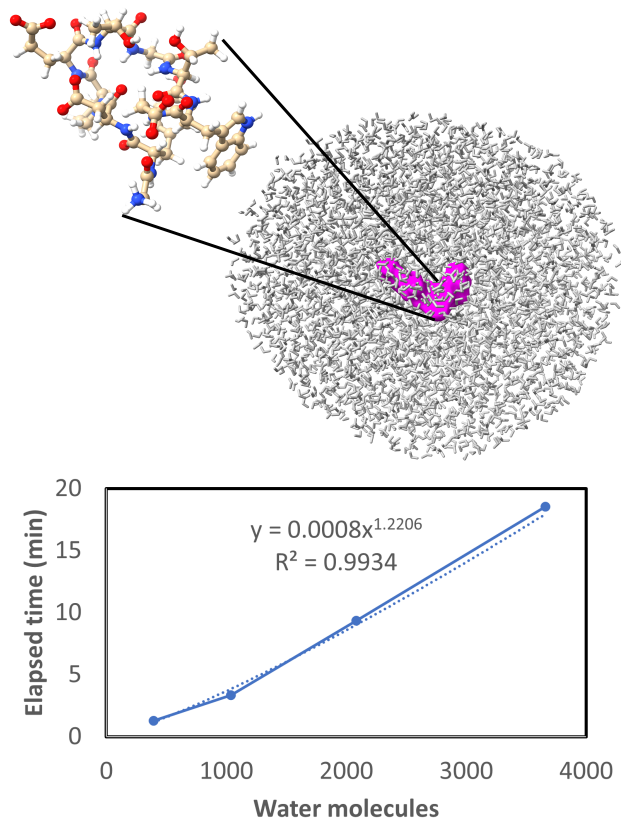

FIG. 4. Scaling of the QM/MM Hessian calculations with respect to the MM atom size. (top) QM/MM model used for the scaling calculations. The QM subsystem is formed by chignolin (PDBID: 1UAO, depicted in magenta soaked in water and its structure in the offset) treated at the HF/3-21G* level and spheres of water molecules around it are computed at the MM level using Amber99. (bottom) Plot of the elapsed time (in minutes) with respect to the number of MM atoms. From these, we have substracted the elapsed time of the pure QM hessian calculation, 15.5 min. Calculations have been performed on 64 CPUs Intel(R) Xeon(R) Platinum 8280 CPU @ 2.70GHz and 2800 Gb of RAM memory.

the building of all the QM/MM Hessian contributions. In other words, the I/O timings are not considered. Indeed, while the MM Hessian has a negligible computational cost with respect to the QM calculation, in our current implementation, Tinker transfers huge amounts of information to Gaussian16 by means of formatted text files. These I/O transfers increase the overall computational time to $\mathcal{O}^{1.85}$, ie. almost proportional to the size of the Hessian. This is however a technical problem, that could be solved by making the MM and QM programs share the same physical memory thanks to a suitable Application Programming Interface or by means of binary files. Finally, one should note that this scaling is possible thanks to the Q-vector method, which avoids the scaling of the coupled perturbed equations with the MM subsystem size.[?][?][?]

V. CONCLUSION

We have refined a QM/MM method based on new Electrostatic Potential Fitted (ESPF) atomic charge operators that are built to conserve the total molecular charge. This ESPF v2.0 approach is capable of calculating QM/MM energy, gradient and full Hessian. When interacting with any kind of external electrostatic potential, the ESPF operators can be added as expectation values to the total energy in a mechanical embedding approach, or directly included into the Fock operator of any self-consistent field procedure in an electrostatic embedding approach.

The QM atomic charges are fitted to reproduce electrostatic integrals computed on a grid surrounding the QM atoms. We showed that, unlike other grids, atom-centred

grids are superior to produce correct atomic charge and energy derivatives. Indeed, they avoid the numerical instabilities which are frequently appearing with other grid types when the QM geometry is distorted. In addition, the atomic charge derivative sum rules as well as the translational invariance condition are retrieved when atom-centred grids are used. Still, rotational invariance of the QM/MM Hamiltonian is not completely satisfied. This would require, for example, the use of rotational invariant Lebedev's grid.[?] However, this would require the implementation of new types of integrals, beyond the scope of the present work.

While some technical improvements are still possible for achieving a better computational efficiency (notably in terms of I/O), we have been able to locate the transition state characterizing a complex two proton transfer reaction occurring in a large cryptochrome protein.

The energy and analytic derivatives of the ESPF method with grid derivatives and charge conservation only affect the QM atomic charges, and thus they were derived for a general MM force field. In our applications, we considered only non-polarizable force fields (fixed point charge representation). The extension to polarizable force fields will change the analytic form for the external potential ϕ as well as its derivatives, depending on how polarization is introduced in the classical force field (Drude particles, multipolar expansion, etc.), but does not affect the ESPF expressions as long as the external MM field is independent of the QM charges.

To conclude, we formulated a general QM/MM method based on ESPF charges without approximations besides the truncation of the multipolar expansion. All the working equations shown here for the lowest order approximation (i.e., atomic charges) can be straightforwardly derived for higher-order terms.

ACKNOWLEDGMENTS

ADDITIONAL INFORMATION

Supporting information includes the full expressions for the first and second derivatives of the charge operator with grid derivatives, the cartesian coordinates of all optimized structures, the ESPF charges for the minimum structure and the transition state in protein, as well as the comparison between numerical and analytic gradient and hessian for all grid types. This information is available free of charge via the Internet at `http://pubs.acs.org`.