



An efficient electrostatic embedding QM/MM method using periodic boundary conditions based on particle-mesh Ewald sums and electrostatic potential fitted charge operators

Simone Bonfrate, Nicolas Ferré, Miquel Huix-Rotllant

► To cite this version:

Simone Bonfrate, Nicolas Ferré, Miquel Huix-Rotllant. An efficient electrostatic embedding QM/MM method using periodic boundary conditions based on particle-mesh Ewald sums and electrostatic potential fitted charge operators. *Journal of Chemical Physics*, 2023, 158 (2), pp.021101. 10.1063/5.0133646 . hal-04047850

HAL Id: hal-04047850

<https://hal.science/hal-04047850>

Submitted on 17 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An efficient electrostatic embedding QM/MM method using periodic boundary conditions based on particle-mesh Ewald sums and electrostatic potential fitted charge operators

Simone Bonfrate, Nicolas Ferré, and Miquel Huix-Rotllant
Aix-Marseille Univ, CNRS, ICR, Marseille, France.

(*Electronic mail: miquel.huix-rotllant@cnsr.fr)

(Dated: 17 November 2023)

Hybrid quantum mechanics / molecular mechanics (QM/MM) models are successful at describing the properties and reactivity of biological macromolecules. Combining *ab initio* QM/MM methods and periodic boundary conditions (PBC) is currently the optimal approach for modeling chemical processes in an infinite environment, but frequently these models are too time-consuming for general applicability to biological systems in solution. Here, we define a simple and efficient electrostatic embedding QM/MM model in PBC combining the benefits of electrostatic potential fitted (ESPF) atomic charges and particle-mesh Ewald sums, that can efficiently treat systems of arbitrary size at a reasonable computational cost. To illustrate this, we apply our scheme to extract the lowest singlet excitation energies from a model for *Arabidopsis thaliana* cryptochrome 1 containing circa 93000 atoms, reproducing accurately the experimental absorption maximum.

Embedding methods in quantum chemistry allow a substantial reduction of the overall computational cost by treating a small subsystem of atoms with an accurate theoretical method while treating the rest of the system in a cheaper and often less accurate approach.¹ In such embedding schemes, the total energy is computed as the sum of energies of the constituent subsystems plus some interaction terms between each fragment.² One of the most popular embedding methods for treating biological macromolecules is quantum mechanics / molecular mechanics (QM/MM),³ in which the energy is expressed as,

$$E = E_{QM} + E_{MM} + E^{int}, \quad (1)$$

where E_{QM} is the energy of the (small) QM subsystem, E_{MM} is the energy of the (large) MM subsystem, and E^{int} is the interaction term between them. Usually, the interaction is electrostatic, complemented with other pairwise atom-atom interactions. The majority of *ab initio* QM/MM methods have been formulated employing an electrostatic Coulomb interaction between QM and MM subsystems, the complete macromolecular system being in the gas phase.⁴ The accuracy of such QM embeddings depends on the locality of the property of interest, as well as parameters like the size of the QM subsystem or constraints imposed on the MM atom positions.⁵ There exist several attempts in the literature to formulate an *ab initio* QM/MM method for macromolecules surrounded by an extended environment (solvent, membrane, etc.), either using non-periodic continuum models⁶ or periodic boundary conditions (PBC) employing the Ewald summation technique.^{7–22} Most QM/MM PBC formulations rely on atomic point charges for efficiently representing the long-range QM-QM interactions such as Mulliken,^{7,8,11,12,15,19} ChElPG,^{14,16,21} ESP,¹⁷ or other types.^{20,22} Such methods mainly use Ewald pair potentials,^{7,14,16,17,21} standard Ewald,^{8,10,13,20} or exploit the efficiency of particle-mesh Ewald (PME) method.^{23–25} The PME, which is state-of-the-art algorithm for efficiently calculating long-range interactions in large MM systems, has mainly been implemented for semi-empirical QM/MM methods.^{11,12,15,19} We are aware of only two recent articles reporting its use in *ab initio* QM/MM methods^{18,22}.

Here, we combine the advantages of electrostatic potential

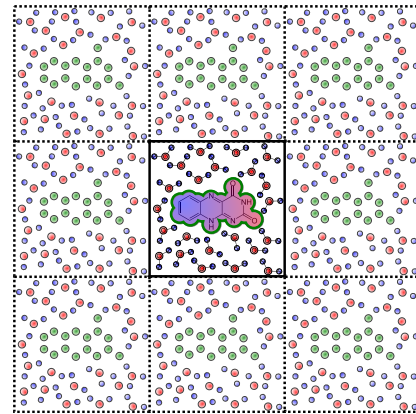


FIG. 1. Schematic representation of the electrostatic embedding QM/MM PBC method described here for a chromophore (QM) in water (MM). In the original cell (center), the QM subsystem is represented by a quantum charge density. In the replica cells, the QM atoms are represented as ESPF point charges (green circles). MM atoms are represented with blue and red point charges. All point charges are used to polarize the QM density in the original cell.

fitted (ESPF) charges,^{4,26} and PME potentials to formulate an efficient *ab initio* electrostatic embedding QM/MM+PBC method, defining a unified consistent embedding energy from an interaction Hamiltonian. Our formulation takes full computational advantage of PME, it reduces the number of integrals to be computed and can be applied to any *ab initio* self-consistent field method. The definition of the ESPF QM/MM+PBC interaction (see Fig. 1) and interaction hamiltonian is based on the pairwise electrostatic interaction energy between QM and MM subsystems (see the detailed derivation in the supporting information). The interaction energy in terms of potentials is defined as,

$$E^{int} = \sum_A^{N_{QM}} q_A \Phi_A^{MM} + \frac{1}{2} \sum_A^{N_{QM}} q_A \Phi_A^{QM}, \quad (2)$$

where the first term accounts for the QM-MM interactions and the second term for the QM-QM interactions, with the 1/2

factor to avoid double counting. N_{QM} is the number of QM atoms, $q_A = Z_A - Q_A$ are their partial charges, defined as the difference between the atomic charge Z_A and the electronic charge population $Q_A = \sum_{\mu\nu} P_{\mu\nu} Q_{A,\mu\nu}$, obtained as the contraction between the quantum density matrix $\mathbf{P}$ and an atomic charge operator matrix $\mathbf{Q}_A$ to be defined later on. The notation Φ_A is a short-hand notation for the external potential felt at atom position $\mathbf{r}_A$, that is, $\Phi_A = \Phi(\mathbf{r}_A)$. We can obtain an interaction operator by deriving the interaction energy with respect to any density matrix element $P_{\mu\nu}$, leading to

$$h_{\mu\nu}^{int} = - \sum_A^{N_{QM}} Q_{A,\mu\nu} (\Phi_A^{MM} + \Phi_A^{QM}). \quad (3)$$

Up to this point, this is a general formulation for a QM/MM embedding when using charge operators including QM-QM interactions. The different QM/MM models are then distinguished by defining the MM energy, the interaction energy and the external electrostatic potential from a classical electrostatic interaction. For example, in QM/MM models of pairwise Coulomb interactions without PBC, the QM/MM procedure is defined given an external electrostatic potential

$$\Phi_A = \Phi_A^{MM} = \sum_{j=1}^{N_{MM}} \frac{q_j}{|\mathbf{r}_{Aj}|}. \quad (4)$$

Here, $\mathbf{r}_{Aj} = \mathbf{r}_A - \mathbf{r}_j$ is the distance vector between the QM particle A and the MM particle j . In this case, $\Phi_A^{QM} = 0$, and therefore the interaction energy and operator matrix elements can be computed given the potential generated by MM atoms on QM centers.

The use of PBC allows one to account for the long-range interactions and to build a model of the macromolecule interactions with the solvent. However, the introduction of Coulomb interaction with replicas results in slow and conditionally convergent interaction energy.²⁷ The Ewald sums guarantee a faster convergence of the energy sums over replica cells. A general Ewald sum formulation in terms of potentials for N interacting charges can be written as

$$E^{Ew} = \frac{1}{2} \sum_{\alpha} q_{\alpha} (\Phi_{\alpha}^{short} + \Phi_{\alpha}^{long} + \Phi_{\alpha}^{self}). \quad (5)$$

Here, Φ_{α} has to be understood as the electrostatic potential calculated at the position of atom α in the original cell, that is, $\Phi_{\alpha} = \Phi(\mathbf{r}_{\alpha 0})$, where $\mathbf{r}_{\alpha 0} = \mathbf{r}_{\alpha} + \mathbf{0}L$, L being the length of the cubic unitary cell. The first term is computed in real space and recovers the short-range part of the electrostatic interaction. The short-range potential is given by

$$\Phi_{\alpha}^{short} = \sum_{\mathbf{n}=0}^N \sum_{\gamma=1}^N \frac{q_{\gamma}}{|\mathbf{r}_{\alpha\gamma\mathbf{n}}|} \text{erfc}(\beta |\mathbf{r}_{\alpha\gamma\mathbf{n}}|), \quad (6)$$

where $\mathbf{r}_{\alpha\gamma\mathbf{n}} = \mathbf{r}_{\alpha} - \mathbf{r}_{\gamma} + \mathbf{n}L$. The first summation over $\mathbf{n}$ vectors runs over the original box and all the replicas, whereas the sum over γ runs over all the N particles in the system. The prime in the second sum means that we are excluding

those terms for which $\alpha = \gamma$ when $\mathbf{n} = \mathbf{0}$. Finally, the complementary error function, $\text{erfc}(x) = 1 - \text{erf}(x)$, ensures the range separation and the convergence control through the parameter β . The second term in Eq. 5 corresponds to the long-range energy. The corresponding long-range potential, computed in the reciprocal space, is defined as

$$\Phi_{\alpha}^{long} = \frac{1}{\pi V} \sum_{\mathbf{m} \neq \mathbf{0}} \frac{e^{-\frac{\pi^2 \mathbf{m}^2}{\beta^2}}}{\mathbf{m}^2} \Re [e^{-2\pi i \mathbf{m} \cdot \mathbf{r}_{\alpha}} S(\mathbf{m})], \quad (7)$$

where $\Re[z]$ is the real part of z . The summation runs over all the reciprocal space vectors $\mathbf{m}$, using the so-called structure factors $S(\mathbf{m})$, which are defined as

$$S(\mathbf{m}) = \sum_{\gamma=1}^N q_{\gamma} e^{2\pi i \mathbf{m} \cdot \mathbf{r}_{\gamma}}. \quad (8)$$

Note that the long-range energy includes the spurious interaction of charges with themselves in the original cell. Such contribution is removed by the self-interaction energy, corresponding to the third term in Eq. 5. The associated self-interaction potential is

$$\Phi_{\alpha}^{self} = -\frac{2\beta}{\sqrt{\pi}} q_{\alpha}. \quad (9)$$

Applying the general Ewald sums formula in the QM/MM framework defined in Eqs. 2 and 3 requires new expressions for the MM energy, the interaction energy, and potential. The purely MM electrostatic energy contribution can be easily obtained by restricting indexes α and γ in Eqs. 5 to 9 to MM atoms only. The Ewald interaction energy can be obtained by extracting the terms defined in Eq. 2 from Eq. 5, which results in

$$E^{int,Ew} = \sum_A^{N_{QM}} q_A (\Phi_A^{short,MM} + \Phi_A^{long,MM}) + \frac{1}{2} \sum_A^{N_{QM}} q_A (\Phi_A^{short,QM} + \Phi_A^{long,QM} + \Phi_A^{self,QM}), \quad (10)$$

recalling that the first term includes the interaction of QM charges with MM charges, while the second term includes the interaction of QM charges in the original cell with QM charges in the replicas. The Ewald potentials in Eq. 10 need to be defined for such definition of the interaction. Note that, at variance with the non-PBC case, electrostatic potentials now includes both a MM and a QM parts. The MM potential contains short and long-range contributions, defined as

$$\Phi_A^{short,MM} = \sum_{\mathbf{n}=0}^N \sum_{i=1}^{N_{MM}} \frac{q_i}{|\mathbf{r}_{Ain}|} \text{erfc}(\beta |\mathbf{r}_{Ain}|) \\ \Phi_A^{long,MM} = \frac{1}{\pi V} \sum_{\mathbf{m} \neq \mathbf{0}} \frac{e^{-\frac{\pi^2 \mathbf{m}^2}{\beta^2}}}{\mathbf{m}^2} \Re [e^{-2\pi i \mathbf{m} \cdot \mathbf{r}_A} S_{MM}(\mathbf{m})]. \quad (11)$$

where $S_{MM}(\mathbf{m})$ refers to the structure factor (Eq. 8) in which the index α is restricted to MM charges. The self-correction

correction is not necessary in the MM potential, since the long-range term involves particles of different nature: one QM, A, and all the MM ones. The QM potential includes three terms,

$$\begin{aligned}\Phi_A^{short,QM} &= \sum_{\mathbf{n} \neq 0} \sum_{B=1}^{N_{QM}} \frac{q_B}{|\mathbf{r}_{AB\mathbf{n}}|} \text{erfc}(\beta |\mathbf{r}_{AB\mathbf{n}}|) \\ \Phi_A^{long,QM} &= \frac{1}{\pi V} \sum_{\mathbf{m} \neq 0} \frac{e^{-\frac{\pi^2 \mathbf{m}^2}{\beta^2}}}{\mathbf{m}^2} \Re [e^{-2\pi i \mathbf{m} \cdot \mathbf{r}_A} S_{QM}(\mathbf{m})] \\ &\quad - \sum_{B \neq A}^{N_{QM}} \frac{q_B}{|\mathbf{r}_{AB0}|} \text{erf}(\beta |\mathbf{r}_{AB0}|) \\ \Phi_A^{self,QM} &= -\frac{2\beta}{\sqrt{\pi}} q_A\end{aligned}\quad (12)$$

where $S_{QM}(\mathbf{m})$ refers to the structure factor (Eq. 8) in which the index α is restricted to QM charges. In the short-range potential, the sum $\mathbf{n}$ only runs over the replicas since the QM-QM interactions in the original cell are already included in the QM calculation. Similarly, the long-range potential needs to be corrected by subtracting the long-range QM-QM interactions in the original cell.⁷

While Ewald summation is undoubtedly useful for handling long-range interactions inside the PBC framework, its original formulation scales like $O(N^2)$ and becomes soon computationally unfeasible when the MM system is large. As a consequence, different approaches, aimed at reducing the algorithmic complexity have been proposed.^{28–31} Here, we make use of the Smooth Particle Mesh Ewald (SPME) method, firstly introduced by Pedersen and coworkers,^{23–25} which features a reduced complexity of $O(N \log N)$. The main idea behind SPME consists in approximating the structure factor (Eq. 8) by interpolating the complex exponentials on a numerical grid of B-splines.²⁵ In practice, the QM atomic charges are projected along with the MM atomic charges in the PME grid, from which one can construct the total long-range potential, Φ_A^{long} which corresponds to the sum of the long-range QM and MM contributions to the potential, besides the correction due to the long-range QM-QM interactions in the original cell, which can be added *a posteriori* as a correction term with the same expression as in Eq. 12. Thus, no modification of the original SPME algorithm is required.

The electrostatic potentials (Eqs. 11-12) can be used to construct the interaction operator (Eq. 3) once the charge operator is defined. Here, we define the charge operator via ESPF method which allows an efficient mapping of the quantum electron density to a classical point charge description compatible with the most common forcefields.³² In the ESPF procedure, the charge operator matrix elements are fitted to QM-only electrostatic integrals computed on a numerical grid constructed around the molecule.²⁶ For obtaining the charge operators used in Eq. 3, a system of linear equations,

$$\sum_A \frac{Q_{A,\mu\nu}}{|\mathbf{r}_k - \mathbf{r}_A|} = \int d^3\mathbf{r} \chi_\mu^*(\mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}_k|} \chi_\nu(\mathbf{r}), \quad (13)$$

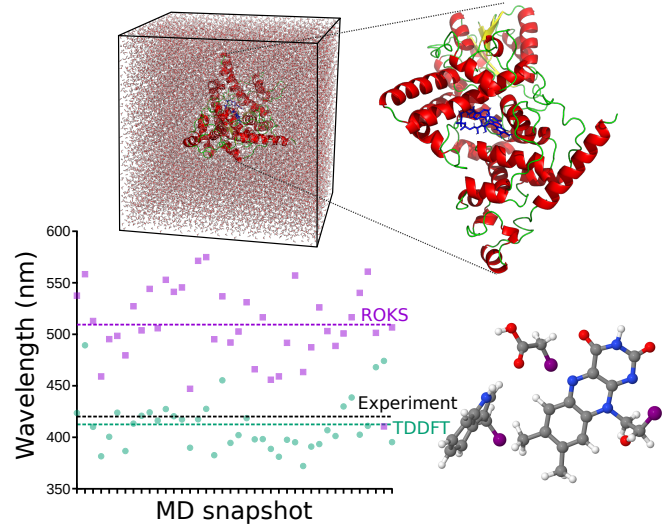


FIG. 2. (top) ESPF QM/MM PBC model of *Arabidopsis thaliana* cryptochrome 1 in a water box. The isoalloxazine, D396, and W400 (62 atoms in total) are treated at the QM level, with the link atoms shown in purple; (bottom, left) Excitation energies for several snapshots computed with restricted open-shell Kohn-Sham and time-dependent density functional theory. The mean values (dashed lines) are compared to experimental values from Ref. 33.

has to be solved. The $\mathbf{r}_k$ are the point coordinates of a Lebedev atom-centered grid defined around the molecule, and χ are the atomic orbitals. A correction is added to the charge operator matrix elements to ensure the conservation of the total charge of the QM subsystem.⁴

The presented methodology, combining SPME and ESPF charges, represents a consistent formulation between the energy and the Hamiltonian and an efficient method for computing the ground and excited state energies. To test this, we extract the lowest singlet excitation energies of 43 snapshots of *Arabidopsis thaliana* cryptochrome 1 (see Fig. 2 and the supporting information for the computational details). On the one hand, time-dependent density functional theory (TDDFT) has been performed on top of the QM/MM PBC ground state Kohn-Sham reference. In this model, we consider that the external potential is fixed in the excited state calculations, and therefore, the only excitation process occurs in the original cell, while the replicas remain in the ground state. No response terms have been added to the TDDFT equations. This approximate model, for which the average excitation energy is around $412 \text{ nm} \pm 26 \text{ nm}$, is in excellent accordance with the experimental absorption maximum of 420 nm .³³ The blue shift can be attributed to the lack of vibronic effects, which are known to be important in the absorption spectra of isoalloxazine.³⁴ On the other hand, we have implemented a restricted open-shell Kohn-Sham (ROKS) model to extract the lowest energy excited state directly from the SCF.³⁵ In this model, both the chromophore in the original cell and the replicas are excited. In this case, the average excitation energy is $518 \text{ nm} \pm 35 \text{ nm}$, thus underestimated by around 100 nm with respect to experiments. Of course, this model is limited

by the fact that excited states are represented by a single configuration, but also the probably unrealistic situation that the photoexcited protein is surrounded by simultaneously excited proteins in the replicas.

In conclusion, we have presented an efficient QM/MM formulation in periodic boundary conditions based on electrostatic potential fitted charges and smooth particle-mesh Ewald sums. It is worth noting that in the QM/MM framework defined by Eqs. 2 and 3 is general, and extra energy terms are easily included as potential sources (see for example the derivation of surface-dipole and the non-neutral cell potentials in the supporting information). The method scales approximately like $O(\alpha \cdot N_{MM}^{1.3})$ ($\alpha = 4 \cdot 10^{-6}$, see supporting information for further details), opening up the route for a general application of QM/MD simulations in large-sized periodic systems. This will require the computation of analytic energy first derivatives, which we plan to develop in the future.

ACKNOWLEDGMENTS

We acknowledge the support from “Agence Nationale de la Recherche” through the project MAPPLE (ANR-22-CE29-0014-01). Centre de Calcul Intensif d’Aix-Marseille is acknowledged for granting access to its high-performance computing resources.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- ¹L. O. Jones, M. A. Mosquera, G. C. Schatz, and M. A. Ratner, *Journal of the American Chemical Society* **142**, 3281 (2020).
- ²D. G. Fedorov, T. Nagata, and K. Kitaura, *Phys. Chem. Chem. Phys.* **14**, 7562 (2012).
- ³Q. Cui, T. Pal, and L. Xie, *The Journal of Physical Chemistry B* **125**, 689 (2021).
- ⁴M. Huix-Rotllant and N. Ferré, *Journal of Chemical Theory and Computation* **17**, 538 (2021).
- ⁵S. Gómez, T. Giovannini, and C. Cappelli, Multiple facets of modeling electronic absorption spectra of systems in solution, *ACS Physical Chemistry Au*, (2022), in press.
- ⁶E. Falbo, M. Fusè, F. Lazzari, G. Mancini, and V. Barone, *Journal of Chemical Theory and Computation* **18**, 6203–6216 (2022).
- ⁷K. Nam, J. Gao, and D. M. York, *Journal of Chemical Theory and Computation* **1**, 2 (2005).
- ⁸D. Riccardi, P. Shaefer, and Q. Cui, *The Journal of Physical Chemistry B* **109**, 17715–17733 (2005).
- ⁹T. Laino, F. Mohamed, A. Laio, and M. Parrinello, *Journal of Chemical Theory and Computation* **1**, 1176 (2005).
- ¹⁰T. Laino, F. Mohamed, A. Laio, and M. Parrinello, *Journal of Chemical Theory and Computation* **2**, 1370 (2006).
- ¹¹G. d. M. Seabra, R. C. Walker, M. Elstner, D. A. Case, and A. E. Roitberg, *The Journal of Physical Chemistry A* **111**, 5655 (2007).
- ¹²R. C. Walker, M. F. Crowley, and D. A. Case, *Journal of Computational Chemistry* **29**, 1019 (2008).
- ¹³C. F. Sanz-Navarro, R. Grima, A. García, E. A. Bea, A. Soba, J. M. Cela, and P. Ordejón, *Theoretical Chemistry Accounts* **128**, 825 (2011).
- ¹⁴Z. C. Holden, R. M. Richard, and J. M. Herbert, *The Journal of Chemical Physics* **139**, 244108 (2013).
- ¹⁵K. Nam, *Journal of Chemical Theory and Computation* **10**, 4175 (2014).
- ¹⁶Z. C. Holden, R. M. Richard, and J. M. Herbert, *The Journal of Chemical Physics* **142**, 059901 (2015).
- ¹⁷T. Vasilevskaya and W. Thiel, *Journal of Chemical Theory and Computation* **12**, 3561 (2016).
- ¹⁸T. J. Giese and D. M. York, *Journal of Chemical Theory and Computation* **12**, 2611 (2016).
- ¹⁹H. Nishizawa and H. Okumura, *Journal of Computational Chemistry* **37**, 2701 (2016).
- ²⁰Y. Kawashima, K. Ishimura, and M. Shiga, *The Journal of Chemical Physics* **150**, 124103 (2019).
- ²¹Z. C. Holden, B. Rana, and J. M. Herbert, *The Journal of Chemical Physics* **150**, 144115 (2019).
- ²²J. P. Pederson and J. G. McDaniel, *The Journal of Chemical Physics* **156**, 174105 (2022).
- ²³T. Darden, D. York, and L. Pedersen, *The Journal of Chemical Physics* **98**, 10089 (1993).
- ²⁴U. Essmann, L. Perera, Berkowitz, T. Max L., Darden, H. Lee, and L. G. Pedersen, *The Journal of Chemical Physics* **103**, 8577 (1995).
- ²⁵C. Sagui, L. G. Pedersen, and T. A. Darden, *The Journal of Chemical Physics* **120**, 73 (2004).
- ²⁶N. Ferré and J. G. Ángyán, *Chemical Physics Letters* **356**, 331 (2002).
- ²⁷P. P. Ewald, *Annalen der Physik* **369**, 253 (1921).
- ²⁸A. Y. Toukmaji and J. A. Board Jr., *Computer Physics Communications* **95**, 73 (1996).
- ²⁹D. R. Wheeler and J. Newman, *Chemical Physics Letters* **366**, 537 (2002).
- ³⁰Y. Shan, J. L. Klepeis, M. P. Eastwood, R. O. Dror, and D. E. Shaw, *The Journal of Chemical Physics* **122**, 054101 (2005).
- ³¹F. Nestler, M. Pippig, and D. Potts, *Journal of Computational Physics* **285**, 280 (2015).
- ³²F. Melaccio, M. Olivucci, R. Lindh, and N. Ferré, *International Journal of Quantum Chemistry* **111**, 3339 (2011).
- ³³M. Ahmad, N. Grancher, M. Heil, R. C. Black, B. Giovani, P. Galland, and D. Lardemer, *Plant Physiology* **129**, 774 (2002).
- ³⁴K. Schwinn, N. Ferré, and M. Huix-Rotllant, *Journal of Chemical Theory and Computation* **16**, 3816 (2020).
- ³⁵T. Kowalczyk, T. Tsuchimochi, P.-T. Chen, L. Top, and T. Van Voorhis, *The Journal of Chemical Physics* **138**, 164101 (2013).