



HAL
open science

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. 2022. hal-03846131

HAL Id: hal-03846131

<https://hal.science/hal-03846131v1>

Preprint submitted on 10 Nov 2022

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@cnr.fr)

(Dated: 8 November 2022)

Hybrid quantum mechanics / molecular mechanics (QM/MM) models successfully describe the properties of biological macromolecules. However, most QM/MM methodologies are constrained to unrealistic gas phase models, thus limiting their applicability. In the literature, several works have attempted to define a QM/MM model in periodic boundary conditions (PBC) but frequently the 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 reducing 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 There exist several attempts in the literature to formulate an *ab initio* QM/MM method for models of macromolecules surrounded by an extended environment (solvent, membrane, etc.), either using non-periodic continuum models⁵ or periodic boundary conditions (PBC) employing the Ewald summation technique.^{6–21} Most QM/MM PBC formulations rely on atomic point charges for efficiently representing the long-range QM-QM interactions such as Mulliken,^{6,7,10,11,14,18} ChEIPG,^{13,15,20} ESP,¹⁶ or other types.^{19,21} Such methods mainly use Ewald pair potentials,^{6,13,15,16,20} standard Ewald,^{7,9,12,19} or exploiting the efficiency of particle-mesh Ewald (PME) method.^{22–24} 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^{10,11,14,18} QM/MM methods. We are aware of only two recent articles reporting its use in *ab initio* QM/MM methods^{17,21}.

Here, we combine the advantages of electrostatic potential fitted (ESPF) charges,^{4,25} and PME potentials to formulate an efficient *ab initio* electrostatic embedding QM/MM PBC method, defining a unified consistent embedding energy from

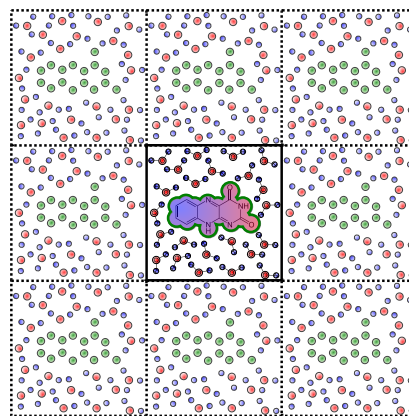


FIG. 1. Schematic representation of the electrostatic embedding QM/MM PBC method described here for a chromophore (QM) in water (MM). In the original cell (center), the QM subsystem is represented by a quantum charge density. In the replica cells, the QM atoms are represented as ESPF point charges (green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

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 ESPF QM/MM PBC interaction (see Fig. 1) is based on the pairwise electrostatic interaction energy between QM and MM subsystems defined in terms of potentials,

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

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. Hereafter, we use the common index notation upper and lower case letters to indicate QM and MM atoms respectively. The 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 A, that is, $\Phi_A = \Phi(\mathbf{r}_A)$.

It must be immediately emphasized that $\partial\Phi_A^{MM}/\partial P_{\mu\nu} = 0$ while $\partial\Phi_A^{QM}/\partial P_{\mu\nu} \neq 0$. Taking this into account, we can obtain an interaction operator by deriving the interaction energy with respect to any density matrix element $P_{\mu\nu}$, leading to

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

where we defined the total potential $\Phi_A = \Phi_A^{MM} + \Phi_A^{QM}$.

Up to this point, this is a general formulation for a QM/MM embedding when using charge operators. The different QM/MM models are then distinguished by defining the MM energy and the external electrostatic potential. For example, in the straightforward QM/MM implementations of pairwise Coulomb interactions without PBC, the QM/MM procedure is simple by defining the external electrostatic potential as

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

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

The use of PBC allows to account for the long-range interactions and to build a more realistic model of the macromolecule interactions with the solvent. However, the introduction of Coulomb interaction with replicas results in slow and conditionally convergent interaction energy. The employment of Ewald summation technique to reach a faster convergence by introducing a range-separated electrostatic interaction,²⁶ allows to split the pairwise Coulombic interaction energy for a system containing N point charges into three contributions,

$$E = E^{short} + E^{long} + E^{self}, \quad (5)$$

in which we define short-range energy as

$$E^{short} = \frac{1}{2} \sum_{\mathbf{n}=0} \sum_{\alpha,\beta=1}^{N'} \frac{q_\alpha q_\beta}{|\mathbf{r}_{\alpha\beta\mathbf{n}}|} \text{erfc}(\beta |\mathbf{r}_{\alpha\beta\mathbf{n}}|), \quad (6)$$

the long-range energy as

$$E^{long} = \frac{1}{2\pi V} \sum_{\mathbf{m} \neq 0} e^{-\frac{\pi^2 \mathbf{m}^2}{\beta^2}} \frac{1}{\mathbf{m}^2} |S(\mathbf{m})|^2, \quad (7)$$

and the self-interaction energy as

$$E^{self} = -\frac{\beta}{\sqrt{\pi}} \sum_{\alpha=1}^N q_\alpha^2. \quad (8)$$

In these formulas, $\mathbf{r}_{\alpha\beta\mathbf{n}} = \mathbf{r}_\alpha - \mathbf{r}_\beta + \mathbf{n}L$, where L is the length of the unitary cell (cubic) box and β is the range separation parameter that controls the rates of converging of the first two terms. E^{short} , which recovers the short-range part of the interactions, is computed in real space and contains a summation over the original box and all the replicas which are described by the \mathbf{n} vectors. The prime in the second sum of Eq. 6 means that we are excluding those terms for which $\alpha = \beta$ when $\mathbf{n} = \mathbf{0}$, while $\text{erfc}(x)$ is the complementary error function which is defined as $\text{erfc}(x) = 1 - \text{erf}(x)$, where $\text{erf}(x)$ is the error function. E^{long} , including the long-range part of the interactions, is computed in the reciprocal space, where the summation runs over all the reciprocal space vectors \mathbf{m} . The so-called structure factors are defined by $S(\mathbf{m}) = S_{QM}(\mathbf{m}) + S_{MM}(\mathbf{m})$, and

$$S_{QM}(\mathbf{m}) = \sum_{A=1}^{N_{QM}} q_A e^{2\pi i \mathbf{m} \cdot \mathbf{r}_A}; \quad S_{MM}(\mathbf{m}) = \sum_{i=1}^{N_{MM}} q_i e^{2\pi i \mathbf{m} \cdot \mathbf{r}_i}. \quad (9)$$

Note that, within the long-range energy expression, the interactions in the original cell for which $\alpha = \beta$ are not omitted from the summation and, instead, the contribution to the energy due to such terms is removed by the introduction of self-interaction energy (E^{self}) which is a constant correction term.

For the QM/MM model in PBC, it is necessary to define new expressions for MM energy and Φ_A out of the Ewald energy expression. The purely MM electrostatic energy contribution can be easily obtained from Eqs. 6 to 8 by restricting the summations to MM atoms only. The electrostatic potentials in QM/MM PBC procedure have both MM and QM parts. The MM potential contains short and long-range contributions,

$$\Phi_A^{MM} = \Phi_A^{short,MM} + \Phi_A^{long,MM}, \quad (10)$$

while the QM potential contains in addition to the short and long-range potentials, a self-interaction and correction terms,

$$\Phi_A^{QM} = \Phi_A^{short,QM} + \Phi_A^{long,QM} + \Phi_A^{self,QM} + \Phi_A^{corr,QM}. \quad (11)$$

Here, Φ_A has to be understood as the electrostatic potential calculated at the position of QM atom A in the original cell, that is, $\Phi_A = \Phi(\mathbf{r}_A + \mathbf{0}L)$. The expressions for the short-range MM and QM potentials are given by

$$\begin{aligned} \Phi_A^{short,MM} &= \sum_{\mathbf{n}=0} \sum_{i=1}^{N_{MM}} \frac{q_i}{|\mathbf{r}_{Ain}|} \text{erfc}(\beta |\mathbf{r}_{Ain}|) \\ \Phi_A^{short,QM} &= \sum_{\mathbf{n} \neq 0} \sum_{B=1}^{N_{QM}} \frac{q_B}{|\mathbf{r}_{ABn}|} \text{erfc}(\beta |\mathbf{r}_{ABn}|), \end{aligned} \quad (12)$$

and the long-range MM and QM potentials are given by

$$\begin{aligned} \Phi_A^{long,MM} &= \frac{1}{\pi V} \sum_{\mathbf{m} \neq 0} e^{-\frac{\pi^2 \mathbf{m}^2}{\beta^2}} \frac{1}{\mathbf{m}^2} \Re [e^{-2\pi i \mathbf{m} \cdot \mathbf{r}_A} S_{MM}(\mathbf{m})] \\ \Phi_A^{long,QM} &= \frac{1}{\pi V} \sum_{\mathbf{m} \neq 0} e^{-\frac{\pi^2 \mathbf{m}^2}{\beta^2}} e^{-2\pi i \mathbf{m} \cdot \mathbf{r}_A} S_{QM}(\mathbf{m}), \end{aligned} \quad (13)$$

where $\Re(z)$ is the real part of z . For constructing the total QM potential, two extra terms have to be considered, the self-interaction ($\Phi_A^{self,QM}$) and the correction potentials ($\Phi_A^{corr,QM}$). The self-interaction potential is defined as

$$\Phi_A^{self,QM} = -\frac{2\beta}{\sqrt{\pi}}q_A, \quad (14)$$

while the correction potential is defined as

$$\Phi_A^{corr,QM} = -\sum_{B=1}^{N_{QM}} \frac{q_B}{|\mathbf{r}_{AB0}|} \text{erf}(\beta |\mathbf{r}_{AB0}|). \quad (15)$$

These two terms arise from spurious interactions that need to be removed.⁶ In this framework, extra energy terms are easily included as extra potential sources (see for example the surface-dipole and the non-neutral cell correction terms in the supporting information).

The ESPF method ensures the uniqueness of the resulting QM/MM potential energy surface, an absolute requirement for obtaining accurate molecular gradients at a reasonable computational cost.²⁷ Therefore, ESPF guarantees that the correction term in eq. 15 is equivalent to the interaction term that should be removed from eq. 13. 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.²⁵ For obtaining the charge operators used in eq. 3, a system of equations,

$$\sum_A \frac{Q_{A,\mu\nu}}{|\mathbf{r}_k - \mathbf{r}_A|} = \int d^3\mathbf{r} \chi_\mu^*(\mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}_k|} \chi_\nu(\mathbf{r}), \quad (16)$$

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

While Ewald summation has been undoubtedly useful for handling long-range interactions inside the PBC framework, its original formulation scales like $\mathcal{O}(N^2)$ 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} The model described in this work makes use of the Smooth Particle Mesh Ewald (SPME) method, firstly introduced by Pedersen and coworkers,^{22–24} which features a reduced complexity of $\mathcal{O}(N \log(N))$. The main idea behind PME consists in approximating the structure factor (eq. 9) by interpolating the complex exponential. While the original Particle Mesh Ewald method makes use of Lagrangian interpolation, the SPME is based on cardinal B-spline interpolation, which allows analytic differentiation. Within SPME, the long-range potential could be expressed as

$$\Phi_A^{long} \approx \sum_{\mathbf{k}} \theta_p(u_{1j} - k_1) \theta_p(u_{2j} - k_2) \theta_p(u_{3j} - k_3) (G \star Q)(\mathbf{k}) \quad (17)$$

where θ_p is the p -th order cardinal B-spline functions, \mathbf{u} contains the scaled fractional coordinates of a point of coordinates

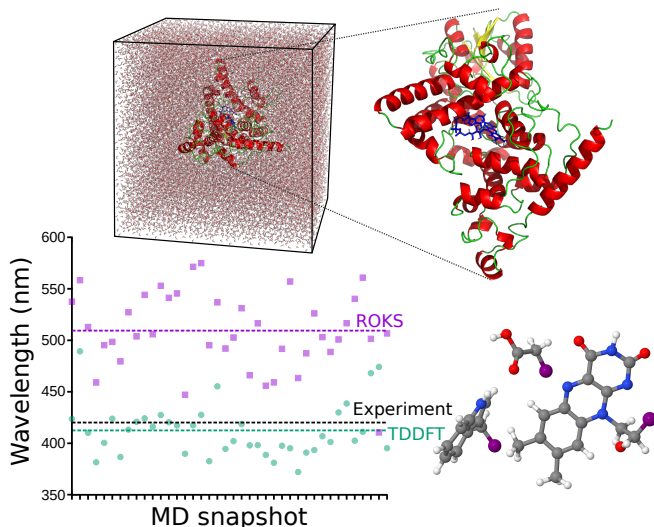


FIG. 2. (top) ESPF QM/MM PBC model of *arabidopsis thaliana* cryptochrome 1 in a water box. The isoalloxazine, D396 and W400 (64 atoms in total) are treated at the QM level, with the link atoms shown in purple; (bottom, left) Excitation energies for several snapshots computed with restricted open-shell Kohn-Sham and time-dependent density functional theory. The mean values (dashed lines) are compared to experimental values from Ref. 32.

\mathbf{r} in the original cell, \mathbf{n} are the real-space grid points vectors, $G \star Q$ is the convolution of G , the generalized influence function defined by its Fourier transform, and Q , the generalized grid multipolar array.²⁴

The presented methodology, combining PME 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 40 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 .³² The blue-shift can be attributed to the lack of vibronic effects, which are known to be important in the absorption spectra of isoalloxazine.³³ 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.³⁴ 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 con-

figuration, 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. 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

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