Analytic QM/MM atomic charge derivatives avoiding the scaling of coupled perturbed equations with the MM subsystem size

Karno Schwinn, Nicolas Ferré, Miquel Huix-Rotllant

Analytic QM/MM atomic charge derivatives avoiding the scaling of coupled perturbed equations with the MM subsystem size

Cite as: J. Chem. Phys. **151**, 041102 (2019); https://doi.org/10.1063/1.5115125
Submitted: 14 June 2019 . Accepted: 01 July 2019 . Published Online: 24 July 2019

Karno Schwinn, Nicolas Ferré, and Miquel Huix-Rotllant

The Journal of Chemical Physics | COMMUNICATION | scitation.org/journal/jcp

Analytic QM/MM atomic charge derivatives avoiding the scaling of coupled perturbed equations with the MM subsystem size

Cite as: J. Chem. Phys. 151, 041102 (2019); doi: 10.1063/1.5115125
Submitted: 14 June 2019 • Accepted: 1 July 2019 •
Published Online: 24 July 2019

Karno Schwinn, Nicolas Ferré, and Miquel Huix-Rotllant[a]

AFFILIATIONS

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

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

ABSTRACT

In electrostatic embedding mixed quantum and molecular mechanics (QM/MM) approaches, the QM charge distribution is polarized by the electrostatic interaction with the MM environment. Analytic derivatives of expectation values of operators are required to extract properties such as vibrational spectra. These derivatives usually require solving a set of coupled perturbed equations for each nucleus/atom in the system, thus becoming prohibitive when the MM subsystem contains thousands of atoms. In the context of Electrostatic Potential Fitting (ESPF) QM/MM, we can easily overcome this bottleneck by defining a set of auxiliary coupled perturbed equations called the Q-vector equations. The Q-vector method scales only with the size of the QM subsystem, producing an effective charge tensor that leads to the atomic charge derivative after contraction with the MM electrostatic potential gradient. As an example, we use the charge derivatives as an analysis tool to identify the most important chromophore-polarizing amino-acids in plant cryptochrome. This finding opens up the route of defining polarizable force fields and simulating vibrational spectroscopy using ESPF QM/MM electrostatic embedding at an affordable computational cost.

Mixed quantum and molecular mechanics (QM/MM) methods[1] have become standard approaches for treating quantum systems embedded in (large) molecular environments, usually modeled using classical interatomic potentials. The total energy of the system can be partitioned as

$$E = E_{\mathrm{QM}} + E_{\mathrm{MM}} + E_{\mathrm{QM/MM}}\,, \tag{1}$$

in which E_{QM} (respectively, E_{MM}) is the energy of the QM (respectively, MM) subsystem and $E_{\mathrm{QM/MM}}$ represents the interaction energy between the two subsystems. Among the different flavors of QM/MM couplings, e.g., mechanical embedding,[2] polarizable embedding,[3] etc., this communication focuses on the electrostatic embedding mode. In the latter, the QM/MM leading interaction yields a polarization of the QM charge distribution by the electrostatic potential generated by the MM subsystem.[4] Using the Electrostatic Potential Fitted (ESPF) method at its lowest order,[5] the interaction term reads

$$E_{\mathrm{QM/MM}} = \sum_{A}^{N_{\mathrm{QM}}} \phi_A \left(Z_A + \sum_{\mu\nu} P_{\mu\nu} \hat{Q}_{\mu\nu,A} \right) = \sum_{A}^{N_{\mathrm{QM}}} \phi_A q_A^0, \tag{2}$$

in which Z_A is the nuclear charge of the QM atom A, ϕ_A is the classical external potential felt by the QM atom, and $\hat{Q}_{\mu\nu,A}$ is the atomic charge operator.[5] Note the strict equivalence, hence uniqueness, between the classical and the quantum representations of this interaction energy.[6] The atomic charge operator, whose definition is independent of the MM coordinates, can be explicitly written as

$$\hat{Q}_{\mu\nu,A} = \sum_{k}^{N_{\mathrm{grid}}} \left[\left(\mathbf{T}^\dagger \mathbf{T} \right)^{-1} \mathbf{T}^\dagger \right]_{kA} \left\langle \mu \left| \frac{1}{|\mathbf{r} - \mathbf{r}_k|} \right| \nu \right\rangle, \tag{3}$$

in which $\langle \mu || \mathbf{r} - \mathbf{r}_k |^{-1} | \nu \rangle$ are the electrostatic integrals on a grid of N_{grid} points and coordinates $\mathbf{r}_k$, selected as a function of the QM atom coordinates, and $\mathbf{T}$ is the electrostatic kernel, a rectangular

matrix containing $N_{\text{grid}} \times N_{\text{QM}}$ terms $T_{Ak} = |\mathbf{r}_k - \mathbf{r}_A|^{-1}$.[7]

In this work, we focus on the QM atomic charge derivatives using density matrices derived from single determinant mean-field wavefunctions, in which the density matrix elements $P_{\mu\nu}$ are simply given by

$$P_{\mu\nu} = \sum_i C^*_{\nu i} C_{i\mu}\,, \tag{4}$$

in which i runs over the occupied molecular orbitals and $C_{\mu i}$ is the molecular orbital coefficient. The contraction of the density matrix with the atomic charge operator leads to the atomic charge q^0_A. These QM atomic charges play a central role in our QM/MM methodology. They are polarizable and can respond to a change in the MM electrostatic potential. Indeed, a Taylor development of the QM atomic charges leads to

$$q_A = q^0_A + \sum_i q^{\tilde{x}_i}_A \Delta\tilde{x}_i + \frac{1}{2!}\sum_{ij} q^{\tilde{x}_i\tilde{x}_j}_A \Delta\tilde{x}_i\Delta\tilde{x}_j + \cdots, \tag{5}$$

where $q^x_A \equiv \left.\frac{\partial q_A}{\partial x}\right|_{x=x_0}$ is a short-hand notation for the first derivative of the charge. Hereafter, superscripts are to be interpreted as derivatives, while tilde refers to MM atoms. The zero-order contribution $q_A \approx q^0_A$ is sufficient for computing the energy and its nuclear gradient.[5] However, second derivatives of the QM/MM energy [Eq. (1)] and vibrational properties such as infrared,[8] Raman,[9] and vibrational circular dichroism[10] require first-derivatives of atomic charges.[11]

The expression for the first derivative of the atomic charge with respect to a QM atom is given by

$$q^x_A = \sum_{\mu\nu}\left(P^x_{\mu\nu}\hat{Q}_{\mu\nu,A} + P_{\mu\nu}\hat{Q}^x_{\mu\nu,A}\right), \tag{6}$$

while the derivative with respect to an MM atom is given by

$$q^{\tilde{x}}_A = \sum_{\mu\nu} P^{\tilde{x}}_{\mu\nu}\hat{Q}_{\mu\nu,A}\,. \tag{7}$$

The derivative of the ESPF charge operator has a straightforward analytic formula that has been presented elsewhere.[5] Here, we focus on the derivatives of the electronic density derivatives over the QM and MM atoms. In order to construct such density derivatives over QM atoms, it is common to compute the MO coefficient derivative as

$$C^x_{p\mu} = \sum_q U^x_{pq} C_{q\mu}, \tag{8}$$

in which $\mathbf{U}^x$ is given by

$$U^x_{ij} = -\frac{1}{2}S^{(x)}_{ij}, \tag{9}$$

$$U^x_{ia} = -\left(S^{(x)}_{ai} + U^x_{ai}\right), \tag{10}$$

$$U^x_{ab} = -\frac{1}{2}S^{(x)}_{ab}, \tag{11}$$

for the occupied-occupied, occupied-virtual, and virtual-virtual blocks. The indexes $i, j, \ldots$ and $a, b, \ldots$ correspond to occupied and virtual orbitals, respectively, and the $S^{(x)}$ matrix is the derivative of the atomic overlap matrix at fixed MO coefficients [indicated as (x) in the superscript]. The occupied-virtual block is solved by a set of coupled perturbed equations,

$$\sum_{bj}(A-B)_{ai,bj}U^x_{bj} = -D^{(x)}_{ai}, \tag{12}$$

with the following definitions for **A**, **B**, and **D**:

$$\begin{aligned} A_{ai,bj} &= (\epsilon_a - \epsilon_i)\delta_{ij}\delta_{ab} + (ai|f_{Hxc}|bj), \\ B_{ai,bj} &= (ia|f_{Hxc}|bj), \\ D^{(x)}_{ai} &= F^{(x)}_{ai} - \epsilon_i S^{(x)}_{ai} - \sum_{kl}(ai|f_{Hxc}|kl)S^{(x)}_{kl}, \end{aligned} \tag{13}$$

in which ϵ_p are the canonical orbital energies, f_{Hxc} is the Hartree-exchange-correlation kernel, and $F^{(x)}$ is the derivative of the Fock operator. In principle, Eq. (12) has to be solved for each perturbation x. In the case of QM atoms, this requires solving the coupled perturbed self-consistent field (CPSCF) equation $3N_{QM}$ times. In addition, another set of CPSCF equations has to be solved for the MM atoms to obtain $P^{\tilde{x}}$. They take the form

$$\sum_{bj}(A-B)_{ai,bj}U^{\tilde{x}}_{bj} = -h^{(\tilde{x})}_{ai}\,, \tag{14}$$

in which the h operator is given by

$$h^{(\tilde{x})}_{ai} = \sum_A \phi^{\tilde{x}}_A \sum_{\mu\nu} C_{i\mu}\hat{Q}_{\mu\nu,A}C^*_{\nu a} = \sum_A \phi^{\tilde{x}}_A \hat{Q}_{ia,A}. \tag{15}$$

Usually, the number of MM atoms is so large that the numerical resolution of these equations becomes intractable. Several approximations have been proposed in order to reduce the computational cost of charge derivatives.[11,12] Here, we show that we can work around the computation of $P^{\tilde{x}}$ by applying the Z-vector method of Handy and Schaefer.[13] Indeed, we can rewrite the atomic charge derivative [Eq. (6)] as

$$q^{\tilde{x}}_A = \sum_{\mu\nu} P^{\tilde{x}}_{\mu\nu}\hat{Q}_{A,\mu\nu} = 2\sum_{ai}\hat{Q}_{A,ai}U^{\tilde{x}}_{ai}\,. \tag{16}$$

Employing the CPSCF over MM coordinates [Eq. (14)], we obtain

$$q^{\tilde{x}}_A = -2\sum_{ia}\hat{Q}_{A,ia}\sum_{jb}(A-B)^{-1}_{ia,jb}h^{(\tilde{x})}_{jb}, \tag{17}$$

and using the definition of the h operator [Eq. (15)], we arrive to

$$q^{\tilde{x}}_A = -2\sum_{B,ia}\hat{Q}_{A,ia}\phi^{\tilde{x}}_B\sum_{jb}(A-B)^{-1}_{ia,jb}\hat{Q}_{B,jb}. \tag{18}$$

We can now define a new $\tilde{Q}$-vector,

$$\tilde{Q}_{B,ia} = -2\sum_{jb}(A-B)^{-1}_{ia,jb}\hat{Q}_{B,jb}, \tag{19}$$

which is obtained as a solution from an auxiliary CPSCF equation defined as

$$\sum_{jb}(A-B)_{ia,jb}\tilde{Q}_{B,jb} = -2\hat{Q}_{B,ia}. \tag{20}$$

This allows us to write the final expression for the atomic charge derivative with respect to MM coordinates as

$$q_A^{\tilde{x}} = \sum_{B,ia} \hat{Q}_{A,ia}\phi_B^{\tilde{x}}\tilde{Q}_{B,ia} = \sum_B Q'_{AB}\phi_B^{\tilde{x}}. \tag{21}$$

This is the main results of the present work. The advantage of this definition is obvious since now we avoid solving MM coordinate times the CPSCF equations, and we only solve N_{QM} times the $\tilde{Q}$-vector equation defined in Eq. (20). Then, Q'_{AB} can be interpreted as a tensor that, given an electric charge Q_A on QM atom A interacting with a gradient of the external potential, will give the "effective" charge on QM center B that interacts with another gradient of the external field.

We have shown to this point that the first derivative of the ESPF atomic charges with respect to MM perturbations of single-determinant wavefunctions does not require the solution of $3 \cdot N_{\mathrm{MM}}$ times the CPSCF equation, but rather an auxiliary set of only N_{QM} CPSCF equations is necessary, in which the perturbation is given by the atomic charge operator. We call this the *Q-vector* method. In fact, this method is valid for higher orders of the multipolar expansion. Indeed, we can apply this technique as long as the perturbation in the CPSCF equation is given by a product of an operator that does not depend on the MM atom positions and an external field that does depend on the MM atom positions. For example, if we write the interaction energy [Eq. (2)] to the next order,

$$E_{\mathrm{QM/MM}} = \sum_A \left(q_A\phi_A - \boldsymbol{\mu}_A \cdot \nabla\boldsymbol{\phi}_A\right), \tag{22}$$

in which $\boldsymbol{\mu}_A$ is the dipole tensor and $\nabla\boldsymbol{\phi}_A$ is the external potential derivative. Applying the same strategy as shown for the charge operator, we can define a new set of independent CPSCF equations for each component of the dipole tensor operator of the type

$$\sum_{jb} (A-B)_{ia,jb}\tilde{\mu}_{\tau,A,jb} = -\hat{\mu}_{\tau,A,ia}, \tag{23}$$

in which $\tau = x, y, z$. Then, the charge derivative [Eq. (21)] would add the following contribution:

$$q_A'^{\tilde{x}} = q_A^{\tilde{x}} + \sum_{\tau,B,ia} \tilde{\mu}_{\tau,B,ia}\hat{\mu}_{\tau,A,ia} \cdot \nabla_\tau\phi_B^{\tilde{x}}. \tag{24}$$

Again, only N_{QM} auxiliary CPSCF equations for each tensor component need to be solved.

We have implemented the Q-vector method in a local development version of GAUSSIAN 16.[14] The implementation requires only the transformation of the first-order derivative of the ESPF operator [Eq. (3)] in atomic orbital basis to molecular orbital basis. The occupied-virtual block is then used to solve the CPSCF equation [Eq. (20)], using the standard algorithms for solving CPSCF equations for static perturbations.[15]

As a first application of the Q-vector method, we use the charge derivative with respect to MM atoms as a measure of the polarization of the QM subsystem. The derivative of atomic charges q^x and $q^{\tilde{x}}$ has the physical interpretation of the induced dipole moments by motions of the QM and MM atoms, respectively. In fact, we can use this as a criterion for selecting the "importance" of an atom, a molecule, or a cluster (hereafter simply denoted as residue) in polarizing the quantum system. If we define the atomic charge response modulus as

$$\delta q_{A,k} = \sqrt{\left(q_A^{\tilde{x}_k}\right)^2 + \left(q_A^{\tilde{y}_k}\right)^2 + \left(q_A^{\tilde{z}_k}\right)^2}, \tag{25}$$

in which the index k refers to MM atoms, we can define the residue induced polarization of QM atom A as

$$\tilde{\mu}_{A,\mathrm{res}} = \frac{1}{N_{\mathrm{res}}} \sum_{k\in\mathrm{res}}^{N_{\mathrm{res}}} \delta q_{A,k}, \tag{26}$$

in which "res" stands for residue (for example, an amino acid in the case of proteins), k runs over all the atoms in the residue, and N_{res} is the number of atoms in the residue. We have applied this criterion to investigate the residues, the motion of which induces the largest polarization of isoalloxazine chromophore in the plant cryptochrome 3 of *Arabidopsis thaliana*.[16] For each of the QM atoms of the chromophore, we search for the $\max_{\mathrm{res}}\{\tilde{\mu}_{A,\mathrm{res}}\}$. In Fig. 1, we associate each of the isoalloxazine atoms with the residue inducing the largest $\tilde{\mu}$. As we can observe, the isoalloxazine chromophore is polarized by 4 types of residues. The uracil-like ring of isoalloxazine is strongly affected by the motions of aspartate 422. The nitrogen N_5 is strongly affected by the motions of asparagine 428, while the o-xylene ring is mostly polarized by asparagine 431 and to a lesser extent by arginine 394. Interestingly, the N_5 position is strongly polarized by a single residue. This position becomes acidic when isoalloxazine is reduced, and a proton transfer from a nearby residue occurs.[17] With the proposed analysis, we can easily identify a good candidate as a proton donor. Furthermore, this criterion can be used to identify important residues for performing mutations in "site-directed evolution," that is, identifying the mutations that will affect the quantum properties of the chromophore in the desired manner.[18]

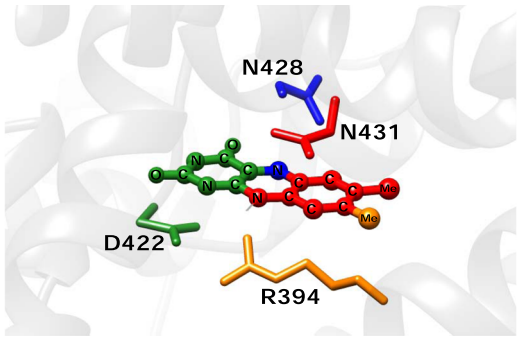

FIG. 1. Schematic representation of the residues (stick) showing the largest induced dipoles ($\max_{\mathrm{res}}\{\tilde{\mu}_{A,\mathrm{res}}\}$, see definition in the text) for each of the atoms of isoalloxazine chromophore (ball and stick). The color associates each residue inducing the largest polarization of the quantum atom. This calculation has been performed at the B3LYP/6-31G* level for the QM subsystem using the Amber99 force field for the MM subsystem.

In summary, we have shown that the Q-vector method allows us to obtain analytic atomic charge derivatives in electrostatic embedding ESPF QM/MM methods by solving an auxiliary number of equations that scales only with the number of QM atoms. We have shown that this is general for any order of the multipolar expansion for the QM/MM interaction. Indeed, the Q-vector method can be applied whenever the atomic charge operator does not depend on the MM coordinates. This opens up a broad perspective in the electrostatic embedding QM/MM methods, ranging from a definition of polarizable force fields, second- and higher-derivatives of the QM/MM energy, and calculation of property derivatives necessary for computing infrared, Raman, and other type of vibrational spectroscopies.

We acknowledge financial support from the project BIOMAGNET (Grant No. ANR-16CE29-0008-01) by the "Agence Nationale de la Recherche." This work was granted access to the HPC resources of Aix-Marseille Université financed by the project Equip@Meso (Grant No. ANR-10-EQPX-29-01) of the program "Investissements d'Avenir" supervised by the "Agence Nationale de la Recherche."

REFERENCES

[1] H. M. Senn and W. Thiel, Angew. Chem., Int. Ed. **48**, 1198 (2009).
[2] L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, H.-B. Li, L. Ding, and K. Morokuma, Chem. Rev. **115**, 5678 (2015).
[3] F. Lipparini and V. Barone, J. Chem. Theory Comput. **7**, 3711 (2011).
[4] H. Lin and D. G. Truhlar, Theor. Chem. Acc. **117**, 185 (2006).
[5] N. Ferré and J. G. Ángyán, Chem. Phys. Lett. **356**, 331 (2002).
[6] F. Melaccio, M. Olivucci, R. Lindh, and N. Ferré, Int. J. Quantum Chem. **111**, 3339 (2011).
[7] A. Stone, *The Theory of Intermolecular Forces*, International Series of Monographs on Chemistry (Clarendon Press, 1997).
[8] F. Lipparini, C. Cappelli, G. Scalmani, N. De Mitri, and V. Barone, J. Chem. Theory Comput. **8**, 4270 (2012).
[9] T. Giovannini, M. Olszówka, F. Egidi, J. R. Cheeseman, G. Scalmani, and C. Cappelli, J. Chem. Theory Comput. **13**, 4421 (2017).
[10] T. Giovannini, M. Olszówka, and C. Cappelli, J. Chem. Theory Comput. **12**, 5483 (2016).
[11] Q. Cui and M. Karplus, J. Chem. Phys. **112**, 1133 (2000).
[12] A. Ghysels, H. L. Woodcock, J. D. Larkin, B. T. Miller, Y. Shao, J. Kong, D. V. Neck, V. V. Speybroeck, M. Waroquier, and B. R. Brooks, J. Chem. Theory Comput. **7**, 496 (2011).
[13] N. C. Handy and H. F. Schaefer, J. Chem. Phys. **81**, 5031 (1984).
[14] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, GAUSSIAN 16 Revision A.01, Gaussian, Inc., Wallingford, CT, 2016.
[15] R. McWeeny, Phys. Rev. **126**, 1028 (1962).
[16] R. Pokorny, T. Klar, L.-O. Essen, and A. Batschauer, Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun. **61**, 935 (2005).
[17] I. A. Solov'yov, T. Domratcheva, A. R. Moughal Shahi, and K. Schulten, J. Am. Chem. Soc. **134**, 18046 (2012).
[18] R. S. McIsaac, M. K. M. Engqvist, T. Wannier, A. Z. Rosenthal, L. Herwig, N. C. Flytzanis, E. S. Imasheva, J. K. Lanyi, S. P. Balashov, V. Gradinaru, and F. H. Arnold, Proc. Natl. Acad. Sci. U. S. A. **111**, 13034 (2014).