The Journal of Chemical Physics

The fragment molecular orbital method combined with density-functional tight-binding and periodic boundary conditions

Cite as: J. Chem. Phys. **154**, 111102 (2021); https://doi.org/10.1063/5.0039520
Submitted: 03 December 2020 . Accepted: 22 February 2021 . Published Online: 15 March 2021

Yoshio Nishimoto, and Dmitri G. Fedorov

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

The fragment molecular orbital method combined with density-functional tight-binding and periodic boundary conditions

Cite as: J. Chem. Phys. 154, 111102 (2021); doi: 10.1063/5.0039520
Submitted: 3 December 2020 • Accepted: 22 February 2021 •
Published Online: 15 March 2021

Yoshio Nishimoto[1] and Dmitri G. Fedorov[2,a)]

AFFILIATIONS

[1]Graduate School of Science, Kyoto University, Kitashirakawa Oiwakecho, Sakyoku, Kyoto 606-8502, Japan
[2]Research Center for Computational Design of Advanced Functional Materials (CD-FMat), National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan

[a)]**Author to whom correspondence should be addressed:** d.g.fedorov@aist.go.jp

ABSTRACT

The density-functional tight-binding (DFTB) formulation of the fragment molecular orbital method is combined with periodic boundary conditions. Long-range electrostatics and dispersion are evaluated with the Ewald summation technique. The first analytic derivatives of the energy with respect to atomic coordinates and lattice parameters are formulated. The accuracy of the method is established in comparison to numerical gradients and DFTB without fragmentation. The largest elementary cell in this work has 1631 atoms. The method is applied to elucidate the polarization, charge transfer, and interactions in the solution.

I. INTRODUCTION

Quantum-mechanical (QM) calculations can describe many useful properties of materials such as heat of chemical reactions, electronic structure, and multipole moments. To reduce the large cost of QM calculations, various methods have been proposed, taking advantage of the locality of the electron density. Among low scaling methods,[1] there are a large number of various fragment-based approaches.[2–12]

One of them is the fragment molecular orbital (FMO) method.[13–17] FMO has been developed for many *ab initio* QM methods and used to evaluate single point energies and interactions between fragments,[18,19] especially between residues in proteins and ligands.[20,21] However, traditional QM methods combined with FMO are rather expensive for full geometry optimizations and molecular dynamics (MD), although some progress has been achieved[22,23] in reducing the cost.

To enable full structure relaxation, density-functional tight-binding (DFTB) is a good compromise between accuracy and efficiency. DFTB is parameterized to reproduce either density functional theory (DFT) properties[24–26] or other high-level methods,[27–29] but at a very small fraction of cost. FMO has been combined with DFTB for both two and three-body expansions in FMO and second- and third-order of DFTB.[30–33]

A few of the fragment-based methods[7,34,35] can be used with periodic boundary conditions (PBC). Although some work has been done for formulating FMO with PBC,[36,37] these Hartree–Fock approaches are costly for large scale applications and lack a proper summation of the long-ranged electrostatic interaction and analytic derivatives with respect to lattice parameters.

In this work, by employing DFTB, a practical FMO–DFTB/PBC method is formulated, in which the periodic electrostatics is properly accounted for using the Ewald[38] summation technique, and the analytic gradients with respect to both atomic coordinates and lattice parameters are derived. Besides FMO, there are other low scaling formulations of DFTB.[39–41]

FMO/PBC is intended to be used with large elementary cells. Typical examples are solutions, crystals with defects, adsorption on crystal surfaces, and crystals of macromolecules such as proteins. The concentration of solutes, defects, or guest molecules is often low, which means that a large elementary cell should be used.

There is a strong demand for the QM method with PBC that can be used for large periodic cells, and the method in this work opens many future fields of applications. Also, the developed method delivers interaction energies between fragments in a periodic system, providing quantitative means for gaining physical insight into the forces driving processes such as adsorption or chemical reactions.

II. METHODOLOGY

A. Density-functional tight-binding theory with periodic boundary conditions

The total energy per unit cell in the third-order DFTB3/PBC method[42,43] can be written as

$$E = \sum_{\mathbf{k}} w_{\mathbf{k}} \sum_{\mu\nu\in\mathbf{0}} D^{\mathbf{k}}_{\mu\nu} H^{0,\mathbf{k}}_{\mu\nu} + \frac{1}{2}\sum_{AB}\sum_{\mathbf{R}} \gamma_{A\in\mathbf{0},B\in\mathbf{R}} \Delta q_A \Delta q_B + \frac{1}{6}\sum_{AB}\sum_{\mathbf{R}} (\Gamma_{A\in\mathbf{0},B\in\mathbf{R}}\Delta q_A + \Gamma_{B\in\mathbf{0},A\in\mathbf{R}}\Delta q_B)\Delta q_A \Delta q_B + \sum_{\mathbf{R}}\sum_{A\in\mathbf{0}}\sum_{B\in\mathbf{R}(B<A)} E^{\mathrm{rep}}_{AB}, \quad (1)$$

where $\mathbf{k}$ is the wave vector, $w_{\mathbf{k}}$ is the weight factor, $D^{\mathbf{k}}_{\mu\nu}$ is the electron density matrix, Δq_A is the Mulliken atomic charge of atom A, and $\mathbf{0}$ and $\mathbf{R}$ are the original and replicated (image) cell, respectively. γ_{AB} and Γ_{AB} are the interatomic distance dependent functions. The second and third terms in Eq. (1) describe the electrostatic interaction with some exchange–correlation taken into account[43] as related to the second- and third-order of DFTB, respectively.

In DFTB2, all Γ_{AB} values are zero; DFTB2 equations can thus be trivially obtained from DFTB3. In this work, both DFTB2 and DFTB3 are implemented.

The $\mathbf{k}$-dependent 0-order Hamiltonian is

$$H^{0,\mathbf{k}}_{\mu\nu} = \sum_{\mathbf{R}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}} H^{0}_{\mu\in\mathbf{0},\nu\in\mathbf{R}}, \quad (2)$$

where $\mathbf{R}$ runs over replicated cells including the original cell $\mathbf{0}$. $H^{0}_{\mu\in\mathbf{0},\nu\in\mathbf{R}}$ is computed as for molecules using atomic orbitals (AOs) $\mu \in \mathbf{0}$ and $\nu \in \mathbf{R}$.

Because atomic charges are independent of the cell $\mathbf{R}$, one can sum the γ terms in Eq. (1) as

$$\tilde{\gamma}_{AB} = \sum_{\mathbf{R}} \gamma_{A\in\mathbf{0},B\in\mathbf{R}}. \quad (3)$$

As the distance r_{AB} between atoms A and B increases, γ_{AB} asymptotically behaves as $1/r_{AB}$, and the summation over $\mathbf{R}$ has to be performed over replicated cells to infinity. In this study, the three-dimensional Ewald summation technique[38] is used as follows:

$$\sum_{\mathbf{R}} \frac{1}{r_{A\in\mathbf{0},B\in\mathbf{R}}} = \sum_{\mathbf{R}} \frac{\mathrm{erfc}(\alpha|\mathbf{r}_{AB}+\mathbf{R}|)}{|\mathbf{r}_{AB}+\mathbf{R}|} + \frac{4\pi}{V}\sum_{\mathbf{G}\neq 0}\frac{\cos(\mathbf{G}\cdot\mathbf{r}_{AB})}{|\mathbf{G}|^2} \times \exp\left(-\frac{|\mathbf{G}|^2}{4\alpha^2}\right) - \frac{\pi}{V\alpha^2} - \delta_{AB}\frac{2\alpha}{\sqrt{\pi}}, \quad (4)$$

where $\mathbf{r}_{AB}$ is the difference between Cartesian coordinates of atoms A and B, α is a parameter that regulates the separation of the short and long distance terms, $\mathbf{G}$ is a unit cell vector in the reciprocal space, and V is the volume of the unit cell.

Because the electrostatic interaction is conditionally convergent, the original Ewald summation requires that the system should be neutralized. However, the third term in Eq. (4), as a result of the conducting or tinfoil boundary conditions,[44] can compensate for a charge in the periodic cell. Nevertheless, it is known that these boundary conditions lead to an overstabilization of ions,[45] so PBC is usually applied to systems with the total zero charge.

In the universal force-field (UFF)-type dispersion model,[46,47] one has to sum $1/r^m_{AB}$ terms for some m over replicated cells. The Ewald summation is used to do that.[48] Another variant of the dispersion model[49] has been implemented as well, but this model is not a simple polynomial function of $1/r_{AB}$, so the dispersion contribution is computed by directly summing the values over the original and a few neighboring image cells. In this study, the UFF-type dispersion model is employed throughout.

The Γ_{AB} function is relatively short-ranged and decays exponentially,[43] so it can be computed directly by summing over a few neighboring cells $\mathbf{R}$,

$$\tilde{\Gamma}_{AB} = \sum_{\mathbf{R}} \Gamma_{A\in\mathbf{0},B\in\mathbf{R}}. \quad (5)$$

The same direct summation over a few cells can be done for the repulsion (rep) term in Eq. (1),

$$\tilde{E}^{\mathrm{rep}} = \sum_{A>B} \tilde{E}^{\mathrm{rep}}_{AB} = \sum_{\mathbf{R}}\sum_{A\in\mathbf{0}}\sum_{B\in\mathbf{R}(B<A)} E^{\mathrm{rep}}_{AB}. \quad (6)$$

In this work, only the Γ point (the center of the Brillouin zone) is used, that is, $w_{\mathbf{k}}$ is 1 for $\mathbf{k} = \mathbf{0}$ and 0 otherwise, so that Eq. (1) can be simplified as

$$E = \sum_{\mu\nu\in\mathbf{0}} D_{\mu\nu}\tilde{H}^{0}_{\mu\nu} + \frac{1}{2}\sum_{AB\in\mathbf{0}}\tilde{\gamma}_{AB}\Delta q_A\Delta q_B + \frac{1}{6}\sum_{AB\in\mathbf{0}}\left(\tilde{\Gamma}_{AB}\Delta q_A + \tilde{\Gamma}_{BA}\Delta q_B\right)\Delta q_A\Delta q_B + \tilde{E}^{\mathrm{rep}}, \quad (7)$$

where

$$\tilde{H}^{0}_{\mu\nu} = \sum_{\mathbf{R}} H^{0}_{\mu\in\mathbf{0},\nu\in\mathbf{R}}. \quad (8)$$

This matrix is relatively short-ranged and can be computed by summing over a few neighboring cells.

B. FMO–DFTB with periodic boundary conditions

In FMO, an elementary cell is divided into fragments. During each fragment calculation, the atomic charges for atoms in the corresponding fragment are updated, whereas the charges of the other atoms (not in the fragment) are fixed. After performing all fragment calculations, the total embedding is updated using the atomic charges of all fragments. Then fragment calculations are repeated with the updated embedding until convergence of the fragment energies with respect to the embedding. Finally, fragment pairs (in FMO3, also triples) are computed in the embedding to describe interfragment charge transfer.[18]

The workflow[40,50] for FMO–DFTB with PBC is essentially the same as without PBC. There are only two basic differences: (a) the

embedding is due to the atoms in the original and all replicated cells with (PBC) vs atoms in the molecule (no PBC) and (b) dimers cover fragment pairs in the original cell and also when one fragment is in an image cell with (PBC) vs dimers within the molecule (no PBC). However, the overall scheme is the same, as the summation over image cells can be done internally (see below) hiding the PBC treatment in the gamma summations [Eq. (3)].

The total energy for the two-body FMO method (FMO2) is

$$E = \sum_{I}^{N} E'_I + \sum_{I>J}^{N} \Delta E_{IJ}, \tag{9}$$

where pair interaction energies (PIEs) between fragments I and J are

$$\Delta E_{IJ} = E'_{IJ} - E'_I - E'_J + \Delta E^V_{IJ}. \tag{10}$$

The internal energy of X ($X = I$ for fragments and $X = IJ$ for pairs) is

$$E'_X = \sum_{\mu\nu\in X} D^X_{\mu\nu}(\widetilde{H}^{0,X}_{\mu\nu} + P^X_{\mu\nu}) + \frac{1}{2}\sum_{AB\in X}\widetilde{\gamma}_{AB}\Delta q^X_A \Delta q^X_B + \frac{1}{6}\sum_{AB\in X}\left(\widetilde{\Gamma}_{AB}\Delta q^X_A + \widetilde{\Gamma}_{BA}\Delta q^X_B\right)\Delta q^X_A \Delta q^X_B + \widetilde{E}^{\mathrm{rep},X}, \tag{11}$$

where $P^X_{\mu\nu}$ is the hybrid orbital projection (HOP) operator for covalent boundaries between fragments.[51] In the AO basis, the operator is

$$P^X_{\mu\nu} = B\sum_{\rho\sigma} S^X_{\mu\rho}\widetilde{D}^X_{\rho\sigma}S^X_{\sigma\nu}, \tag{12}$$

where B is a universal constant (usually 10^6 a.u.), $\widetilde{D}^X_{\rho\sigma}$ is the density of the hybrid orbitals,[52] and $S_{\mu\rho}$ is the overlap matrix in the AO basis.

The energy representing the coupling of the second ($\Delta\Delta q^{IJ}_A$) and third ($\Delta\Delta Q^{IJ}_A$) order charge transfer terms[33] between I and J with the embedding potential V (due to fragments K) is

$$\Delta E^V_{IJ} = \sum_{A\in IJ}\sum_{K\neq I,J}^{N}\sum_{B\in K}\left\{\widetilde{\gamma}_{AB}\Delta q^{IJ}_A\Delta q^K_B + \frac{1}{3}\widetilde{\Gamma}_{AB}\Delta\Delta Q^{IJ}_A\Delta q^K_B + \frac{1}{3}\widetilde{\Gamma}_{BA}\Delta\Delta q^{IJ}_A(\Delta q^K_B)^2\right\}. \tag{13}$$

If two fragments are far from each other, there is no charge transfer between them and the electrostatic dimer (ES-DIM) approximation can be used to compute the energy for such dimer IJ as

$$E'_{IJ} \approx E'_I + E'_J + \sum_{A\in I}\sum_{B\in J}\left\{\widetilde{\gamma}_{AB} + \frac{1}{3}\left(\widetilde{\Gamma}_{AB}\Delta q^I_A + \widetilde{\Gamma}_{BA}\Delta q^J_B\right)\right\}\Delta q^I_A\Delta q^J_B. \tag{14}$$

C. Analytic derivatives with respect to atomic coordinates

By differentiating the total energy with respect to a nuclear coordinate a, one obtains[32]

$$\frac{dE}{da} = \sum_I\left(1 - N^I_{\mathrm{md}}\right)E'^a_I + \sum_{I>J}^{N}\left(E'^a_{IJ} + \Delta E^{V,a}_{IJ}\right) + 4\sum_{K}^{N}\sum_{k\in K}^{\mathrm{virt}}\sum_{l\in K}^{\mathrm{occ}} U^{K,a}_{kl}\mathcal{L}^K_{kl}, \tag{15}$$

where N^I_{md} is the number of short-range dimers including monomer I, "virt" and "occ" stand for virtual and occupied molecular orbitals, respectively. $U^{K,a}_{kl}$ is the orbital response term,[32] and $\mathcal{L}^K_{kl}$ is the orbital Lagrangian.[32] In the derivatives of E'^a_X and $\Delta E^{V,a}_{IJ}$, one has to sum over replicated cells as in Eqs. (2) and (3); otherwise, the expressions are the same as for molecules.[32]

To evaluate the last term in Eq. (15), coupled-perturbed equations are solved using the self-consistent Z-vector method,[32]

$$\sum_{L}^{N}\sum_{i\in L}^{\mathrm{virt}}\sum_{j\in L}^{\mathrm{occ}} Z^L_{ij}\mathcal{A}^{L,K}_{ij,kl} = \mathcal{L}^K_{kl}, \tag{16}$$

where $\mathcal{A}^{X,Y}_{ij,kl}$ defined for molecules[32] is modified by summing over replicated cells. Solving Eq. (16), one obtains Z^L_{ij} and the gradient is computed using the Z-vector technique[53] as

$$\frac{dE}{da} = \sum_I\left(1 - N^I_{\mathrm{md}}\right)E'^a_I + \sum_{I>J}^{N}\left(E'^a_{IJ} + \Delta E^{V,a}_{IJ}\right) + 4\sum_{K}^{N}\sum_{k\in K}^{\mathrm{virt}}\sum_{l\in K}^{\mathrm{occ}} Z^K_{kl}\mathcal{B}^{K,a}_{kl}, \tag{17}$$

where $\mathcal{B}^{X,a}_{ij}$ is defined elsewhere[32] for molecules and a summation over replicated cells has to be used in it for PBC.

In practice, the most difficult part of implementing DFTB/PBC in the Γ point approximation is the replacement of the long-ranged γ_{AB} with $\widetilde{\gamma}_{AB}$, which has to be summed properly to infinity; in addition, one has to do an explicit summation of nearby cells for short-ranged terms, such as the 0-order Hamiltonian. In the FMO/PBC equations, the PBC part is hidden inside tilde terms and the equations [e.g., Eq. (9)] are similar to those without PBC.

D. Analytic derivatives with respect to lattice parameters

The stress tensor is

$$\Pi_{\alpha\beta} = -\frac{1}{V}\sum_{\gamma}\frac{dE}{da_{\alpha\gamma}}a_{\beta\gamma}, \tag{18}$$

where $a_{\alpha\beta}$ is the lattice parameter, $\alpha, \beta = x, y$ or z ($\mathbf{a}_\alpha$ is the lattice vector).

To compute the stress tensor in FMO, one needs to obtain the tensor contribution from the internal energy of X [see Eqs. (9) and (18)],

$$\sum_{\gamma}\frac{dE'_X}{da_{\alpha\gamma}}a_{\beta\gamma} = \sum_{AB\in X}\frac{dE'_X}{dr^\alpha_{AB}}r^\beta_{AB}, \tag{19}$$

$$= \sum_{AB\in X}\sum_{\mu\in A}\sum_{\nu\in B}\left\{D^X_{\mu\nu}\frac{\partial\widetilde{H}^{0,X}_{\mu\nu}}{\partial r^\alpha_{AB}} - W'^X_{\mu\nu}\frac{\partial\widetilde{S}^X_{\mu\nu}}{\partial r^\alpha_{AB}} + \frac{1}{2}\frac{\partial\widetilde{S}^X_{\mu\nu}}{\partial r^\alpha_{AB}}\Omega^X_{AB}\right\}r^\beta_{AB} + \frac{1}{2}\sum_{AB\in X}\left\{\frac{\partial\widetilde{\gamma}_{AB}}{\partial r^\alpha_{AB}} + \frac{1}{3}\left(\frac{\partial\widetilde{\Gamma}_{AB}}{\partial r^\alpha_{AB}}\Delta q^X_A + \frac{\partial\widetilde{\Gamma}_{BA}}{\partial r^\alpha_{AB}}\Delta q^X_B\right)\right\}\Delta q^X_A\Delta q^X_B r^\beta_{AB} + \sum_{A>B\in X}\frac{\partial\widetilde{E}^{\mathrm{rep}}_{AB}}{\partial r^\alpha_{AB}}r^\beta_{AB}, \tag{20}$$

where r^{α}_{AB} is the α component of the $\mathbf{r}_{AB}$ and $W^{IX}_{\mu\nu}$ and Ω^{X}_{AB} are defined elsewhere.[32]

The derivatives of ΔE^{V}_{IJ} are obtained similarly. After summing all fragment terms in Eq. (9), one obtains the total stress tensor $\Pi_{\alpha\beta}$ for FMO, from which the derivatives with respect to lattice parameters are obtained by multiplying with a reciprocal cell matrix.[42]

Response contributions described above are also needed for a proper evaluation of the stress tensor. As Z^{L}_{ij} is obtained in FMO–DFTB for analytic gradients anyway, the extra cost for PBC is very little. Neglect of the response can make the stress tensor non-symmetric.

E. Extension to FMO3 and interfragment distances for PBC

The energy in the three-body FMO3–DFTB method is[33]

$$E^{\mathrm{FMO3}} = \sum_{I}^{N} E'_I + \sum_{I>J}^{N} \Delta E_{IJ} + \sum_{I>J>K}^{N} (\Delta E_{IJK} - \Delta E_{IJ} - \Delta E_{JK} - \Delta E_{KI}). \quad (21)$$

In this work, FMO3–DFTB was also developed (the formulation closely follows FMO2–DFTB, and the details are omitted). Both the energy and its analytic derivatives can be computed for FMO3–DFTB/PBC.

There are two types of approximations used in FMO–DFTB, ES-DIM, and I-TRIM (I-TRIM stands for ignoring separated trimers) applied to fragments pairs and triples, respectively. Both approximations rely on a definition of a distance r_{IJ} between fragments I and J. For molecules, it is

$$r_{IJ} = \min_{A\in I, B\in J} \frac{r_{AB}}{R_A + R_B}, \quad (22)$$

where R_A is the van der Waals radius of atom A and r_{AB} is the geometrical distance between atoms A and B.

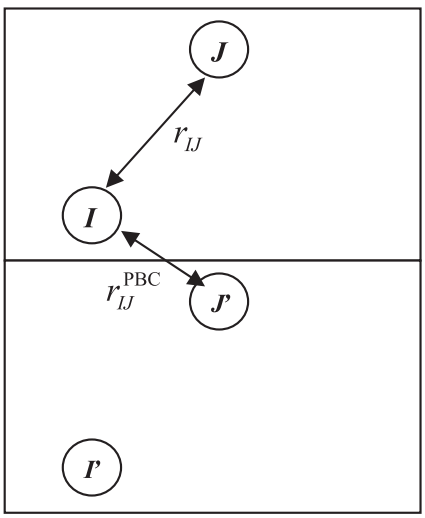

FIG. 1. Two copies of the elementary cell are shown. Without PBC, the distance between fragments I and J is r_{IJ}. With PBC, the distance is reduced to r^{PBC}_{IJ} because of the presence of replicated images of fragments, I' and J'.

However, for PBC, there are replicated images of the elementary cell, atoms which may be close to the atoms in the original cell (Fig. 1). Therefore, the distance should be defined as

$$r^{\mathrm{PBC}}_{IJ} = \min_{A\in I(\mathbf{0}), B\in J(\mathbf{R}),\ \forall\ \mathbf{R}} \frac{r_{AB}}{R_A + R_B}, \quad (23)$$

where $\mathbf{0}$ and $\mathbf{R}$ are the original and replicated cells, respectively (the original cell is included in the replicated cell set $\mathbf{R}$).

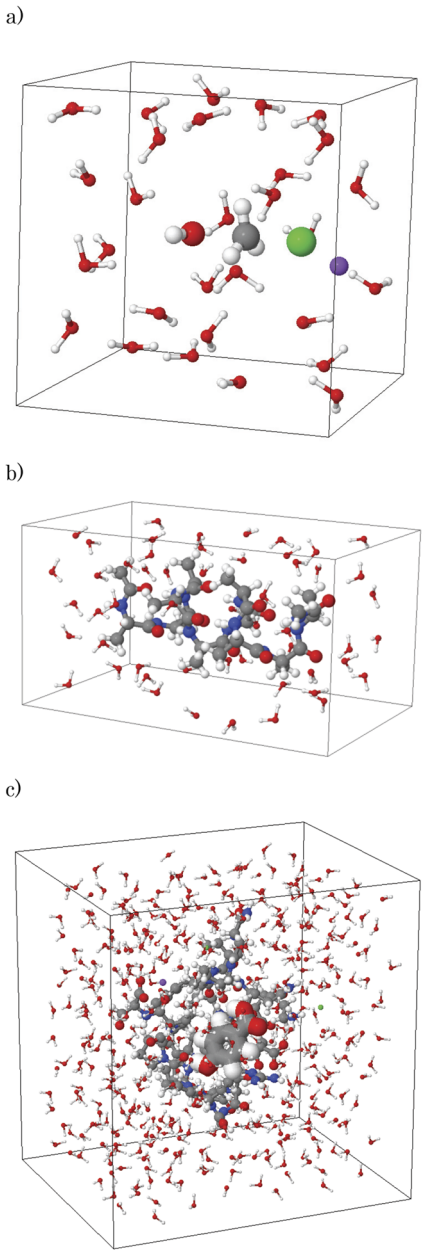

FIG. 2. Molecular structures for the smallest cell size, (a) S_N2, (b) α-helix of $(ALA)_{10}$, and (c) a complex of Trp-cage (1L2Y) with a ligand (in an explicit solvent).

In practice, for a wrapped original cell, it is enough to limit the replicated set to the original cell and its nearest neighbors in the directions of the three lattice vectors. Using Eq. (23) yields smaller distances, in general, than Eq. (22) (see Fig. 1). Once these distances r_{IJ}^{PBC} are computed, ES-DIM and I-TRIM may be used in a straightforward way by comparing the distances to the thresholds.

III. COMPUTATIONAL DETAILS

FMO–DFTB/PBC was implemented in a development version of GAMESS-US[54] and parallelized with the generalized distributed data interface (GDDI).[55] DFTB/PBC without fragmentation was also implemented in this work.

DFTB3 with 3ob-3-1[27] parameters and UFF dispersion correction[46,47] was used in all calculations. HOP boundary treatment was employed for polypeptide fragmentation. Solvent boxes were added using FU.[56] The polypeptide and water were fragmented as one residue and molecule per fragment, respectively.

The following systems were used: (a) an S_N2 reaction $CH_3Cl + OH^- \rightarrow CH_3OH + Cl^-$, (b) a capped α-helix of $(ALA)_{10}$ (α-$(ALA)_{10}$), and (c) a complex of Trp-cage (PDB: 1L2Y) with a ligand, *p*-phenolic acid, all solvated in explicit water. The structures for the smallest box size are shown in Fig. 2.

The geometry of the S_N2 reaction was prepared by locating the transition state using unfragmented DFTB in vacuum. Then, explicit water molecules were added, and, to neutralize the system, a sodium cation was added and assigned to the reaction center fragment. The structure of the α-$(ALA)_{10}$ was obtained by a geometry optimization with FMO2–DFTB in vacuum. Then, water molecules were added. These geometries of S_N2 and α-$(ALA)_{10}$ were used for accuracy tests. For an analysis of solute–solvent interactions, the structure of S_N2 was optimized while freezing the coordinates of the reaction site ($CH_3Cl + OH^-$) at the values for the transition state and fully optimizing water and the counterion, with PBC while keeping the box size constant, with the threshold of 10^{-3} a.u./bohr.

The geometry for the protein–ligand complex was optimized with FMO2–DFTB/PCM.[57] Then, the system was neutralized by adding one sodium and two chloride ions to Ser-20, Asn-1, and Lys-8 fragments, respectively; the ions were assigned to the respective fragments. Asp-9 and Arg-16 retained their charges (−1 and +1, respectively). The geometry of all atoms in the protein–ligand complex was fully optimized with PBC (fixed box) with the threshold of 5×10^{-4} a.u./bohr; then, the geometry of the protein and ligand was further optimized to 1×10^{-4} a.u./bohr while keeping the solvent frozen.

IV. RESULTS AND DISCUSSION

The accuracy results of FMO in comparison to the full calculations are shown in Table I. The errors in the energy and gradient with respect to DFTB/PBC without fragmentation are reasonably small, and increasing the order of FMO from two to three substantially reduces the errors. FMO3 errors are within the chemical accuracy of 1 kcal/mol. A comparison of the analytic and numerical gradients in Table II shows that the analytic gradients are accurate for both FMO2 and FMO3.

To gain some insight into the transition state stabilization by the solvent, the total solute–solvent interaction energy ΔE_{int} (the sum of pair interactions ΔE_{IJ} over solvent fragments J for the fixed solute fragment $I = 1$) and the destabilization component of the polarization (pold) energy[58] of the reaction fragment ΔE_{pold} (the difference of the energy of the reaction fragment E_I' in solution and vacuum) were computed for the solution (described with PBC) and a water droplet (no PBC). The concentration was computed based on the cell volume. The subsystem analysis[59] was applied by defining two subsystems, solute and solvent.

PIEs ΔE_{IJ} are a useful measure of the role of fragments in studying binding processes.[20,21] For PBC, PIEs are defined in Eq. (10); they describe interactions between fragments polarized by the whole periodic system (a crystal field polarization; "crystal" here denotes that a system is periodic like a crystal, although it may be a solution). A PIE in FMO/PBC incorporates a sum over cells [via the cell-summed γ, see Eq. (3)] so that a PIE value for a pair of fragments I and J includes a sum of interactions between fragment I in the original cell **0** and fragment J in all cells **R** (including **0**).

The results are shown in Table III. The neutral zwitterionic solute has a strong solute–solvent interaction energy, between −164 and −213 kcal/mol; the pold component is between 8 and 14 kcal/mol. Stronger interactions lead to a larger polarization, as expected.[58] The PBC effect (the difference between solution and droplet) does not exceed 8 kcal/mol; it decreases with the cell size (the effect is weaker for lower concentrations), as expected: the thinner the solvent layer, the stronger the crystal field from replicated

TABLE I. Errors in the FMO*n* ($n = 2, 3$) energies E (kcal/mol) and gradients G (hartree/bohr) with respect to atoms (a) and lattice (l) parameters in comparison to full unfragmented DFTB/PBC.[a]

System	a	N_{at}	E^2	E^3	G_{a}^2, max	G_{a}^2, rmsd	G_{a}^3, max	G_{a}^3, rmsd	G_{l}^2, max	G_{l}^2, rmsd	G_{l}^3, max	G_{l}^3, rmsd
S_N2	12	95	−2.0	0.1	0.002 16	0.000 29	0.000 31	0.000 03	0.000 59	0.000 33	0.000 05	0.000 02
S_N2	18	341	−4.9	0.4	0.004 94	0.000 29	0.000 26	0.000 06	0.001 11	0.000 54	0.000 09	0.000 05
S_N2	23	812	−4.9	0.7	0.001 56	0.000 12	0.000 28	0.000 03	0.000 94	0.000 51	0.000 13	0.000 07
α-$(ALA)_{10}$	24	217	1.4	0.0	0.004 73	0.000 24	0.000 20	0.000 02	0.000 78	0.000 29	0.000 01	0.000 00
α-$(ALA)_{10}$	28	514	0.3	0.2	0.000 75	0.000 09	0.000 18	0.000 02	0.000 13	0.000 06	0.000 04	0.000 02
α-$(ALA)_{10}$	33	1171	−3.2	0.7	0.001 22	0.000 10	0.000 28	0.000 03	0.000 73	0.000 33	0.000 12	0.000 05

[a]The largest of the three lattice parameters is shown as a (Å).

TABLE II. Deviations between analytic and numerical FMO*n* derivatives (hartree/bohr) with respect to atoms (a) and lattice (l) parameters, for the S_N2 reaction in an explicit solvent (water) with the box size of 12 Å.

n	G_a, max	G_a, rmsd	G_l, max	G_l, rmsd
2	0.000 000 7	0.000 000 06	0.000 000 04	0.000 000 02
3	0.000 009 9	0.000 000 75	0.000 000 37	0.000 000 14

TABLE III. Effect of the concentration *c* (M) on the solute–solvent interaction (int) and the solute polarization (pold) energies (kcal/mol) for the transition state in the S_N2 reaction computed with FMO2–DFTB/PBC.

c	$\Delta E_{\text{int}}^{\text{solution}}$	$\Delta E_{\text{pold}}^{\text{solution}}$	$\Delta E_{\text{int}}^{\text{droplet}}$	$\Delta E_{\text{pold}}^{\text{droplet}}$
0.96	−172.1	7.7	−163.8	10.9
0.28	−214.4	13.8	−212.7	14.3
0.15	−181.2	10.8	−179.2	11.2

cells. The largest PBC effect for pold and interaction values is for the lowest concentration, which is because for higher concentrations a thicker solvent layer damps the polarization of the solute. For more reliable energetics, one has to average various conformations, which will be the subject of future work.

The solute–solvent charge transfer is substantial, 0.7–0.9 electrons, which in this system is mainly between Na and water (i.e., Na pulls the electron density of lone pairs of oxygens), because the other charged ion, Cl, is busy in the chemical reaction and shares little of its charge with the solvent.

For the complex of Trp-cage with its ligand, PIEs are shown in Fig. 3 (the concentration of the protein or ligand is 0.067 M). Calculations were done with PBC (solution) and without PBC (droplet). There is little difference in PIEs, which indicates that the droplet of this size (see Fig. 2) is sufficient to describe solute–solute (protein–ligand) interactions.

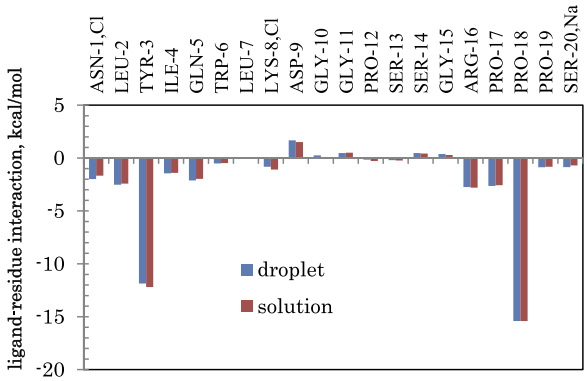

FIG. 3. Residue–ligand interactions ΔE_{lJ} in the Trp-cage complex in an explicit solvent, computed with PBC (solution) and without PBC (droplet) (*J* is the ligand, and *l* runs over all residues).

The reason why a reasonably sized droplet is enough for PIEs is that protein images are well separated from each other by a layer of water so that there is little additional polarization of the protein and ligand by image cells, which makes little difference on the interactions between fragments. For a higher concentration, the effect is expected to be larger.

V. CONCLUSIONS

The density-functional tight-binding approach based on the fragment molecular orbital method has been extended to treat periodic boundary conditions, at the levels of second- and third-order DFTB and two and three-body FMO. Energy and analytic derivatives with respect to atomic coordinates and lattice parameters have been developed.

Long-range electrostatics has been properly summed using the Ewald technique. FMO/PBC uses a polarizable embedding, and the gradient and stress tensor require computing response terms related to the embedding, obtained from the coupled-perturbed equations (Z-vector equations); the developed formalism may be of use to other fragment methods with a polarizable embedding.

The developed method makes it possible to use large elementary cells with the parameterized quantum-mechanical method DFTB (in this work, the largest cell containing a protein, a ligand, and counterions has 1631 atoms), which describes polarization and charge transfer based on the wave function in DFTB.

It has been shown by comparison to DFTB calculations without fragmentation and FMO numerical gradients that energy and analytic gradients in FMO can be accurately calculated using FMO–DFTB/PBC. In the applications to the S_N2 reaction and the protein–ligand complex, it has been shown how PBC affects the pair interactions between fragments. It has been demonstrated how the developed method can be used to gain physical insight into the polarization of the solute, solute–solvent, and solute–solute (protein–ligand) interactions.

FMO–DFTB/PBC can be used to describe solution with explicit solvent as an alternative to continuum models.[57] In the future, FMO–DFTB/PBC may be extended to treat improved DFTB models such as DFTB with long-range corrections.[60] It would be useful to develop analytic second derivatives[61] to study phonons.

SUPPLEMENTARY MATERIAL

See the supplementary material for the energy conservation in an NVE MD simulation performed with FMO2–DFTB3/PBC demonstrating that the energy gradient is analytic. Timings show the efficiency of the developed FMO–DFTB/PBC.

ACKNOWLEDGMENTS

The authors acknowledge financial support by JSPS KAKENHI (Grant No. 19H02682).

DATA AVAILABILITY

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

REFERENCES

[1] A. V. Akimov and O. V. Prezhdo, Chem. Rev. **115**, 5797 (2015).
[2] M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, Chem. Rev. **112**, 632 (2012).
[3] P. Otto and J. Ladik, Chem. Phys. **8**, 192 (1975).
[4] J. Gao, J. Phys. Chem. B **101**, 657 (1997).
[5] K. Kiewisch, C. R. Jacob, and L. Visscher, J. Chem. Theory Comput. **9**, 2425 (2013).
[6] H. Yu, H. R. Leverentz, P. Bai, J. I. Siepmann, D. G. Truhlar, and J. Friedrich, J. Phys. Chem. Lett. **5**, 666 (2014).
[7] H. Nakai, A. W. Sakti, and Y. Nishimura, J. Phys. Chem. B **120**, 217–221 (2016).
[8] C. Bertoni and M. S. Gordon, J. Chem. Theory Comput. **12**, 4743 (2016).
[9] T. Fang, Y. Li, and S. Li, Wiley Interdisp. Rev: Comput. Mol. Sci. **7**, e1297 (2017).
[10] M. B. Ahirwar, S. R. Gadre, and M. M. Deshmukh, J. Phys. Chem. A **124**, 6699 (2020).
[11] K.-Y. Liu and J. M. Herbert, J. Chem. Theory Comput. **16**, 475–487 (2020).
[12] B. Thapa, J. Erickson, and K. Raghavachari, J. Chem. Inf. Model. **60**, 2924 (2020).
[13] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999).
[14] D. G. Fedorov and K. Kitaura, J. Phys. Chem. A **111**, 6904 (2007).
[15] D. G. Fedorov, T. Nagata, and K. Kitaura, Phys. Chem. Chem. Phys. **14**, 7562 (2012).
[16] S. Tanaka, Y. Mochizuki, Y. Komeiji, Y. Okiyama, and K. Fukuzawa, Phys. Chem. Chem. Phys. **16**, 10310 (2014).
[17] D. G. Fedorov, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **7**, e1322 (2017).
[18] D. G. Fedorov and K. Kitaura, J. Phys. Chem. A **122**, 1781 (2018).
[19] D. G. Fedorov, J. Phys. Chem. A **124**, 10346 (2020).
[20] I. Morao, D. G. Fedorov, R. Robinson, M. Southey, A. Townsend-Nicholson, M. J. Bodkin, and A. Heifetz, J. Comput. Chem. **38**, 1987 (2017).
[21] H. Lim, A. Baek, J. Kim, M. S. Kim, J. Liu, K. Nam, J. Yoon, and K. T. No, Sci. Rep. **10**, 16862 (2020).
[22] D. G. Fedorov, Y. Alexeev, and K. Kitaura, J. Phys. Chem. Lett. **2**, 282 (2011).
[23] H. Nakata and D. G. Fedorov, Int. J. Quantum Chem. **120**, e26414 (2020).
[24] S. Jahangiri, G. Dolgonos, T. Frauenheim, and G. H. Peslherbe, J. Chem. Theory Comput. **9**, 3321 (2013).
[25] M. Kubillus, T. Kubař, M. Gaus, J. Řezáč, and M. Elstner, J. Chem. Theory Comput. **11**, 332 (2015).
[26] M. Gaus, H. Jin, D. Demapan, A. S. Christensen, P. Goyal, M. Elstner, and Q. Cui, J. Chem. Theory Comput. **11**, 4205 (2015).
[27] M. Gaus, A. Goez, and M. Elstner, J. Chem. Theory Comput. **9**, 338 (2013).
[28] M. Gaus, X. Lu, M. Elstner, and Q. Cui, J. Chem. Theory Comput. **10**, 1518 (2014).
[29] X. Lu, M. Gaus, M. Elstner, and Q. Cui, J. Phys. Chem. B **119**, 1062 (2015).
[30] Y. Nishimoto, D. G. Fedorov, and S. Irle, J. Chem. Theory Comput. **10**, 4801 (2014).
[31] Y. Nishimoto, D. G. Fedorov, and S. Irle, Chem. Phys. Lett. **636**, 90 (2015).
[32] Y. Nishimoto, H. Nakata, D. G. Fedorov, and S. Irle, J. Phys. Chem. Lett. **6**, 5034 (2015).
[33] Y. Nishimoto and D. G. Fedorov, J. Comput. Chem. **38**, 406 (2017).
[34] H. Li, H. M. Netzloff, and M. S. Gordon, J. Chem. Phys. **125**, 194103 (2006).
[35] C. Liu, J. Zhao, Z.-Z. Yang, and D.-X. Zhao, J. Chem. Theory Comput. **16**, 7618 (2020).
[36] T. Fujita, T. Nakano, and S. Tanaka, Chem. Phys. Lett. **506**, 112 (2011).
[37] K. R. Brorsen, N. Minezawa, F. Xu, T. L. Windus, and M. S. Gordon, J. Chem. Theory Comput. **8**, 5008 (2012).
[38] P. P. Ewald, Ann. Phys. **369**, 253 (1921).
[39] T. J. Giese, H. Chen, T. Dissanayake, G. M. Giambaşu, H. Heldenbrand, M. Huang, E. R. Kuechler, T.-S. Lee, M. T. Panteva, B. K. Radak, and D. M. York, J. Chem. Theory Comput. **9**, 1417 (2013).
[40] T. J. Giese, M. Huang, H. Chen, and D. M. York, Acc. Chem. Res. **47**, 2812 (2014).
[41] H. Nishizawa, Y. Nishimura, M. Kobayashi, S. Irle, and H. Nakai, J. Comput. Chem. **37**, 1983 (2016).
[42] B. Aradi, B. Hourahine, and T. Frauenheim, J. Phys. Chem. A **111**, 5678 (2007).
[43] M. Gaus, Q. Cui, and M. Elstner, J. Chem. Theory Comput. **7**, 931 (2011).
[44] F. Figueirido, G. S. Del Buono, and R. M. Levy, J. Chem. Phys. **103**, 6133 (1995).
[45] J. S. Hub, B. L. de Groot, H. Grubmüller, and G. Groenhof, J. Chem. Theory Comput. **10**, 381 (2014).
[46] A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard III, and W. M. Skiff, J. Am. Chem. Soc. **114**, 10024 (1992).
[47] L. Zhechkov, T. Heine, S. Patchkovskii, G. Seifert, and H. A. Duarte, J. Chem. Theory Comput. **1**, 841 (2005).
[48] B. A. Wells and A. L. Chaffee, J. Chem. Theory Comput. **11**, 3684 (2015).
[49] M. Elstner, P. Hobza, T. Frauenheim, S. Suhai, and E. Kaxiras, J. Chem. Phys. **114**, 5149 (2001).
[50] K. Kitaura, T. Nagata and D. G. Fedorov, "Mathematical formulation of the fragment molecular orbital method," in *Linear-Scaling Techniques in Computational Chemistry and Physics*, edited by R. Zalesny, M. G. Papadopoulos, P. G. Mezey, and J. Leszczynski (Springer, New York, 2011), pp. 17–64.
[51] T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayasi, and K. Kitaura, Chem. Phys. Lett. **318**, 614 (2000).
[52] T. Nagata, D. G. Fedorov, and K. Kitaura, Chem. Phys. Lett. **492**, 302 (2010).
[53] T. Nagata, K. Brorsen, D. G. Fedorov, K. Kitaura, and M. S. Gordon, J. Chem. Phys. **134**, 124115 (2011).
[54] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, J. Comput. Chem. **14**, 1347 (1993).
[55] D. G. Fedorov, R. M. Olson, K. Kitaura, M. S. Gordon, and S. Koseki, J. Comput. Chem. **25**, 872 (2004).
[56] D. G. Fedorov and K. Kitaura, "Modeling and visualization for the fragment molecular orbital method with the graphical user interface FU, and analyses of protein-ligand binding," in *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*, edited by M. S. Gordon (Wiley, Hoboken, 2017), pp. 119–139.
[57] Y. Nishimoto and D. G. Fedorov, Phys. Chem. Chem. Phys. **18**, 22047 (2016).
[58] D. G. Fedorov and K. Kitaura, J. Comput. Chem. **28**, 222 (2007).
[59] D. G. Fedorov and K. Kitaura, J. Phys. Chem. A **120**, 2218 (2016).
[60] V. Q. Vuong, Y. Nishimoto, D. G. Fedorov, B. G. Sumpter, T. A. Niehaus, and S. Irle, J. Chem. Theory Comput. **15**, 3008 (2019).
[61] H. Nakata, Y. Nishimoto, and D. G. Fedorov, J. Chem. Phys. **145**, 044113 (2016).