Density-Functional Tight-Binding Combined with the Fragment Molecular Orbital Method

Yoshio Nishimoto,[†] Dmitri G. Fedorov,*,[§] and Stephan Irle*,[†,‡]

[†]Department of Chemistry and [‡]Institute of Transformative Bio-Molecules (WPI-ITbM), Nagoya University, Nagoya 464-8602, Japan

[§]Nanosystem Research Institute (NRI), National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba 305-8565, Japan

ABSTRACT: We developed the energy and its gradient for the self-consistent-charge density-functional tight-binding (DFTB) method, combined with the fragment molecular orbital (FMO) approach, FMO-DFTB, including an optional a posteriori treatment for dispersion interaction, and evaluated its accuracy as well as computational efficiency for a set of representative systems: polypeptides, a DNA segment, and a small protein. The error in the total energy of FMO-DFTB versus full SCC-DFTB was below 1 kcal/mol for the polyalanine system consisting of about 2000 atoms partitioned

into fragments containing 2 residues, and the optimized structures had root-mean-square deviations below 0.1 Å. The scaling of FMO-DFTB with the system size N is only marginally larger than linear $[O(N^{1.2})$ in the worst case$]$. A parallelization efficiency of 94% was achieved using 128 CPU cores, and we demonstrate the applicability of FMO-DFTB for systems containing more than one million atoms by performing a geometry optimization of a fullerite cluster.
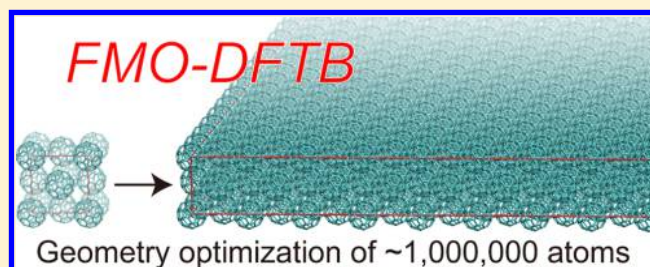
1. INTRODUCTION

In biochemistry, nanoelectronics, and many other areas of molecular science, it is often necessary to deal with systems containing a large number of atoms, typically between 10^3 to 10^7. Simulations of chemical reactions, electron transfer, electronically excited states, and even accurate structure predictions for polar or hydrogen-bonded systems require an accurate treatment of the electronic structure, which can, in principle, be accomplished with quantum-chemical methods based on quantum mechanics (QM methods). However, traditional QM methods scale very steeply with the system size. For instance, even the fastest ab initio Hartree–Fock (HF) method scales as N^{2-3}, where N is the number of atoms, and for coupled cluster (CC) theory, the scaling can be as high as N^7. This means that, although QM calculations of small and medium size systems can be routinely conducted, increasing the system size by a factor of 100 results in an HF calculation that is up to a million time more expensive, and 100 trillion times more expensive in the case of CC. Such huge computational effort is accompanied by a similarly huge increase in required computer memory and disk storage.

At present, only classical molecular mechanics (MM) is capable of treating systems consisting of millions of atoms at the full atomistic level, yet molecular dynamics (MD) simulations using MM force fields for such tremendously large systems require the most powerful, massively parallel supercomputers available today.[1] At the same time, there is an ongoing effort to develop efficient traditional QM methods[2–4] with linear scaling[5–7] and linear scaling semiempirical

approaches, partially designed to replace MM-MD simulations in the future by QM-MD.[8,9]

Alternatively, methods using fragmentation are also becoming increasingly popular.[10–26] Although these methods do succeed in enabling much larger calculations than traditional QM approaches, the reported maximum system sizes are still only on the order of 30 000 atoms[27–32] for full QM applications. We note that, using a piecewise approach, a geometry optimization of a 150 000 atomic system has been reported.[33] Even though fragment-based approaches have been successfully applied to QM calculations at 0 K, MD simulations with full QM fragment-based methods of large systems are rather limited.[34,35]

The density-functional tight-binding (DFTB) method is a semiempirical QM approach applicable toward large systems. It has its origin in the computational materials sciences,[36] and recently, it has found broad applications in organic and biological systems.[37] The method itself has been described in the literature on several occasions, and we therefore suffice to only briefly describe the main features and options related to our work in section 2 and refer the interested reader to recent reviews.[37–40] Essentially, DFTB is an approximate density functional theory (DFT)[41,42] method utilizing the tight binding approach within the framework of an optimized minimal Slater-type all-valence basis set and a two-center approximation for Hamiltonian matrix elements. There are several options

ACS Publications © XXXX American Chemical Society

A

dx.doi.org/10.1021/ct500489d | J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

available. One is the self-consistent-charge (SCC) formalism, in which the interaction between atomic charges is included in the energy expression, and the atomic charges are determined iteratively until self-consistency is reached.[43] This option is sometimes called "DFTB2",[37] as it is a second-order expansion of the DFT energy around a reference density. Higher-order expansions, such as third order corrections ("DFTB3"), have been recently developed and shown to be important in polar and hydrogen-bonded systems.[44] The SCC-DFTB method is similar in spirit to other DFTB-based self-consistent charge transfer tight-binding models.[45,46] These iterative SCC methods are all related to the ideas of density expansion in tight-binding, proposed in 1989 by Foulkes and Haydock.[47] The earliest, and methodologically most simple approach of DFTB, is the nonself-consistent-charge (NCC) option, in which the atomic charge—atomic charge interaction is neglected.[48] This option, nowadays sometimes called "DFTB1",[37] is approximately 5 to 20 times faster than the SCC option, since it is noniterative, and may not be a bad approximation for homogeneous systems with little charge polarization. Although DFTB is fast, its scaling with system size is unfavorable with $O(N^3)$ because even though no integrals are calculated, it is still necessary to diagonalize the DFTB Hamiltonian. Its memory requirement is N^2, and combined with the cubic scaling of the computational efforts, these limitations prevent the use of DFTB for calculations of truly large systems (loosely defined as systems containing more than 10 000 atoms). Recently, however, advances have been made to reduce the cubic scaling required for full Hamiltonian matrix diagonalization: for instance, an $O(N \log N)$-scaling divide-and-conquer[5] formulation[49] of DFTB (mDC-DFTB) has been developed by the York group. Using their code, they were able to obtain a converged electronic density for a 1 million atom containing water cluster in around 5 min on 8 Intel Xeon E5520 cores.[50] Using density matrix purification methods, true $O(N)$ linear scaling with low prefactor was demonstrated for a similar density functional based tight binding method in energy conserving Born–Oppenheimer MD simulations of systems containing fewer than 1000 atoms.[51]

An a posteriori treatment for dispersion interaction within the framework of DFTB was first proposed by Elstner et al. to describe base pair stacking in DNA.[52] It employs lists of van der Waals-active interatomic contacts and a switching function to turn off the van der Waals potential at short distances. Similar corrections are nowadays added routinely in DFT calculations and have widely become known as "Grimme dispersion".[53] In 2005, Zhechkov et al. developed an alternative dispersion interaction model based on the parametrization by Rappé et al.,[54] which is available for all elements of the periodic table and does not require interatomic contact lists. In both approaches, the van der Waals interactions are evaluated as a sum over pair potentials between all atoms, and hence, the dispersion term is a constant for a given nuclear geometry. It is independent of the electronic structure calculation and hence easy to implement, yet has been recognized as important in computational biological and materials sciences.

The fragment molecular orbital (FMO) method[55−59] is one of the fragment-based approaches.[10] In FMO, the system is divided into fragments (also called monomers), which are calculated in the embedding electrostatic potential (ESP), obtained from the density of all fragments. These monomer calculations are repeated iteratively until convergence is reached, followed by the calculation of fragment pairs (dimers)

in the presence of the embedding ESP. FMO has been applied to a variety of molecular systems including proteins,[60,61] DNA,[62−64] polymers,[65] and inorganic systems.[66−70]

The accuracy and scaling of FMO based on regular DFT has been discussed.[71−74] FMO-DFT has been used for a geometry optimization[75] of chignolin, containing 138 atoms. The total energy in FMO is expanded in the many-body series.[76,77] In this work, we use only the two-body expansion (FMO2), often reliable enough for energetics and gradients, and develop the energy and gradient of FMO-based DFTB (FMO2-DFTB). We employ both nonself-consistent-charge (first-order, DFTB1, here called NCC-DFTB)[48,78] and self-consistent charge (second-order, DFTB2, here called SCC-DFTB)[43] DFTB. The accuracy of FMO-DFTB is tested on a variety of representative molecules in terms of energy and optimized geometry, and the computational efficiency as well as scaling behavior with respect to system size and parallelism are discussed in detail.

2. METHODOLOGY

2.1. Density-Functional Tight-Binding (DFTB). The DFTB method has been described in detail elsewhere;[37−40,43,44] therefore, we only provide a brief summary here. The total energy of NCC-DFTB, E^{NCC}, is written as

$$E^{\mathrm{NCC}} = \sum_i \sum_{\mu\nu} n_i c_{\mu i}^* c_{\nu i} H_{\mu\nu}^0 + \sum_{A>B} E_{AB}^{\mathrm{rep}} \tag{1}$$

where μ and ν denote atomic orbitals (AO), A and B label atoms, and n_i is the occupation number of the i-th molecular orbital (MO). The $c_{\mu i}$ are the expansion coefficients of the i-th MO in an AO basis. $H_{\mu\nu}^0$ is the nonperturbed (i.e., atomic- and diatomic-based) Hamiltonian in the basis of optimized AOs, precomputed for atoms and diatomic systems over a range of bond lengths in reference DFT calculations. E_{AB}^{rep} is the two-body repulsive energy term, which is precomputed and tabulated as a function of interatomic distance R_{AB} from DFT calculations of model systems containing the chemical elements of the A and B atoms. As a result of this approach, E_{AB}^{rep} is independent of the electronic structure and only a function of the interatomic distances. It is normally short-ranged and extends usually only up to maximally 1.5 times the equilibrium bond lengths.

The MO coefficients **c** in eq 1 are variationally determined by solving the equation, $\mathbf{Hc} = \mathbf{Sc}\epsilon$, where **S** is the overlap matrix, readily precomputed and tabulated for the optimized AOs used in the construction of $H_{\mu\nu}^0$ as a function of R_{AB}. Because $H_{\mu\nu}^0$ does not depend on **c**, NCC-DFTB is a noniterative method and hence computationally very economical. This method has been successfully applied to a number of systems,[48,79−82] mostly for homopolar systems such as carbon-based nanostructures, although the bandstructure of heteropolar systems can also be described reasonably well.[40,80,83]

In many polar systems one has to consider self-consistent, second-order terms of charge fluctuations.[43] In this case, the calculation of the charge density becomes iterative. Using routine density matrix update algorithms, typically between 5 and 20 SCC iteration cycles are required to achieve convergence. The total energy of SCC-DFTB, E^{SCC}, is written as

$$E^{\mathrm{SCC}} = E^{\mathrm{NCC}} + \frac{1}{2} \sum_{AB} \gamma_{AB} \Delta q_A \Delta q_B \tag{2}$$

where γ_{AB} depends on the distance R_{AB} between two atoms A and B and their chemical hardness.[84] γ_{AB} in the limit of long distances behaves as $1/R_{AB}$ and at short distances accounts for the chemical hardness of the chemical element in question, expressed by a Hubbard parameter U_A that is computed using DFT as a numerical derivative of the DFT atomic orbital energy (normally the highest occupied AO) with respect to its orbital occupation. The Hubbard U_A values are related to atomic ionization potentials and electron affinities. Charge fluctuations $\delta\rho$ around the reference electron density ρ_0 in SCC-DFTB are expressed by a spherical point charge model with partial atomic charges Δq_A. It is the deviation of the Mulliken population on atom A from that in the neutral atom q_A^0, that is $\Delta q_A = q_A - q_A^0$, where

$$q_A = \frac{1}{2} \sum_i n_i \sum_{\mu \in A} \sum_\nu (c_{\mu i}^* c_{\nu i} S_{\mu\nu} + c_{\nu i}^* c_{\mu i} S_{\nu\mu}) \tag{3}$$

Clearly, the last term in eq 2 requires an iterative calculation since atomic populations depend on the MO coefficients **c**. We note that other charge partitioning schemes have been applied and reported for SCC-DFTB,[85] but Mulliken population analysis remains the most widely used method for determining atomic charges.

The derivative of eq 2 with respect to MO coefficients defines the SCC-DFTB Hamiltonian matrix in the AO basis,

$$H_{\mu\nu} = H_{\mu\nu}^0 + \frac{1}{2} S_{\mu\nu} \sum_C (\gamma_{AC} + \gamma_{BC}) \Delta q_C \qquad \mu \in A, \quad \nu \in B \tag{4}$$

The derivative of the SCC-DFTB energy with respect to the x coordinate of atom α (note that throughout this paper x in the derivatives can be replaced by y or z) is

$$\frac{\partial E^{SCC}}{\partial R_{\alpha x}} = \sum_{A \neq \alpha} \sum_{\mu \in A} \sum_{\nu \in \alpha} \sum_i n_i c_{\mu i}^* c_{\nu i} \left\{ 2\frac{\partial H_{\mu\nu}^0}{\partial R_{\alpha x}} - 2\varepsilon_i \frac{\partial S_{\mu\nu}}{\partial R_{\alpha x}} \right.$$
$$\left. + \frac{\partial S_{\mu\nu}}{\partial R_{\alpha x}} \sum_C (\gamma_{AC} + \gamma_{\alpha C}) \Delta q_C \right\}$$
$$+ \Delta q_\alpha \sum_{A \neq \alpha} \Delta q_A \frac{\partial \gamma_{A\alpha}}{\partial R_{\alpha x}} + \sum_{A \neq \alpha} \frac{\partial E_{A\alpha}^{rep}}{\partial R_{\alpha x}} \tag{5}$$

2.2. Formulation of FMO-DFTB. The total energy E in the two-body FMO expansion (FMO2)[55] is

$$E = \sum_I^N E_I + \sum_{I>J}^N (E_{IJ} - E_I - E_J) \tag{6}$$

where N is the number of fragments, and the E_X terms are the energies of fragments ($X = I$) and their pairs ($X = IJ$). For SCC-DFTB, they are defined as

$$E_X = E_X' + E_X^V \tag{7}$$

whereas for NCC-DFTB, the expression is

$$E_X = E_X' \tag{8}$$

The internal energy of X is defined as

$$E_X' = \sum_{i \in X} \sum_{\mu\nu \in X} n_i^X c_{\mu i}^{X*} c_{\nu i}^X H_{\mu\nu}^{0,X} + \sum_{A>B \in X} E_{AB}^{rep}$$
$$+ \frac{1}{2} \sum_{AB \in X} \gamma_{AB} \Delta q_A^X \Delta q_B^X \tag{9}$$

and the embedding energy is

$$E_X^V = \sum_{A \in X} \sum_{K \neq X}^N \sum_{C \in K} \gamma_{AC} \Delta q_A^X \Delta q_C^K \tag{10}$$

Δq_A^X is the charge on atom A in fragment X. Note that A and C can be the same atom, if fragments X and K are connected via a covalent bond. The internal energy of fragment X includes the Coulomb interaction of charges within X, while the embedding energy is the interaction between charges in X and those outside of X (in all other fragments).

Using these definitions, in a same way as in other FMO methods,[86] it is possible to rewrite eq 6 for SCC-DFTB as follows,

$$E = \sum_I^N E_I' + \sum_{I>J}^N (E_{IJ}' - E_I' - E_J') + \sum_{I>J}^N \Delta E_{IJ}^V \tag{11}$$

For NCC-DFTB the expression is (compare eqs 6 and 8)

$$E = \sum_I^N E_I' + \sum_{I>J}^N (E_{IJ}' - E_I' - E_J') \tag{12}$$

In SCC-DFTB, the coupling of the charge transfer to the embedding potential is

$$\Delta E_{IJ}^V = \sum_{A \in IJ} \sum_{K \notin I,J}^N \sum_{C \in K} \gamma_{AC} \Delta\Delta q_A^{IJ} \Delta q_C^K \tag{13}$$

and the charge transfer between fragments I and J for atom A is

$$\Delta\Delta q_A^{IJ} = \Delta q_A^{IJ} - \Delta q_A^I \delta_{A \in I} - \Delta q_A^J \delta_{A \in J} \tag{14}$$

where $\delta_{A \in I} = 1$ when atom A belongs to fragment I; otherwise, it is zero.

In NCC-DFTB, the charge interaction is not considered, and thus, $E_{IJ}' \approx E_I' + E_J'$. On the other hand, in SCC-DFTB, one has to evaluate the fragment interactions for separated pairs, which scales in principle quadratic with system size N. However, the interfragment charge transfer is short-ranged and can be neglected if the interfragment distance R_{IJ} is larger than a certain threshold. In this case, a self-consistent field (SCF) calculation of the far separated dimer is not required, and the interfragment interaction is evaluated using the electrostatic dimer (ES-DIM) approximation instead.[86] Thus, for dimers IJ calculated with this approximation, $\Delta\Delta q_A^{IJ} = \Delta E_{IJ}^V = 0$. Different from quantum-effects such as charge transfer and exchange-repulsion, the electrostatic interaction decays with distance R slowly as $1/R$ and has to be evaluated for all separated dimers.

The energy of a separated dimer in FMO-DFTB should in principle be given by

$$E_{IJ}' \approx E_I' + E_J' + \sum_{A \in I} \sum_{B \in J} (E_{AB}^{rep} + \gamma_{AB} \Delta q_A^I \Delta q_B^J) \tag{15}$$

However, the repulsion energy E_{AB}^{rep} quickly decays with the interatomic distance and with the typical value of the ES-DIM threshold, and therefore can be neglected. Therefore, the

C

dx.doi.org/10.1021/ct500489d | J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

energy of far separated dimers in FMO-based SCC-DFTB is evaluated as

$$E'_{IJ} \approx E'_I + E'_J + \sum_{A \in I} \sum_{B \in J} \gamma_{AB} \Delta q_A^I \Delta q_B^J \tag{16}$$

For the energy gradient, one needs to define the Hamiltonian matrix of fragment X, which for SCC-DFTB is

$$H_{\mu\nu}^X = H_{\mu\nu}^{SCC,X} + V_{\mu\nu}^X + P_{\mu\nu}^X \tag{17}$$

where $H_{\mu\nu}^{SSC,X}$ is the Hamiltonian matrix of fragment X according to eq 4, which includes the Coulomb interaction for charges within X. For NCC-DFTB, one has

$$H_{\mu\nu}^X = H_{\mu\nu}^{0,X} + P_{\mu\nu}^X \tag{18}$$

where $H_{\mu\nu}^{0,X}$ corresponds to $H_{\mu\nu}^0$ of fragment X in eq 4. The ESP acting on X is given by

$$V_{\rho\sigma}^X = \frac{1}{2} S_{\rho\sigma}^X \sum_{K \neq X}^N \sum_{D \in K} (\gamma_{AD} + \gamma_{BD}) \Delta q_D^K \qquad \rho \in A, \quad \sigma \in B \tag{19}$$

$P_{\mu\nu}^X$ is the hybrid orbital projection (HOP) operator,[87] which is introduced in FMO for fragments connected by detached covalent bonds (HOP is not used in molecular clusters). The HOP representation in AO basis is

$$P_{\mu\nu}^X = B \sum_{\rho\sigma} S_{\mu\rho}^X \tilde{D}_{\rho\sigma}^X S_{\sigma\nu}^X \tag{20}$$

where $\tilde{D}_{\mu\nu}^X$ is the density matrix of the hybrid orbitals[88] and B is a universal constant, usually set to 10^6 Hartree. The hybrid orbitals are precomputed for each atom type using a suitable model compound, for example, CH_4 when detaching bonds at sp^3 carbons.

2.3. Gradient of FMO-DFTB. In the following, we derive the gradient for a closed shell singlet, that is $n_i = 2$ for all occupied MOs, which simplifies the formulation of FMO-DFTB with the ES-DIM approximation. By taking the derivative of eq 9, one obtains

$$\frac{\partial E'_X}{\partial R_{\alpha x}} = \sum_{(A \neq \alpha) \in X} \sum_{\mu \in A} \sum_{\nu \in \alpha} \left\{ 2 D_{\mu\nu}^X \frac{\partial H_{\mu\nu}^{0,X}}{\partial R_{\alpha x}} - 2 W_{\mu\nu}'^X \frac{\partial S_{\mu\nu}^X}{\partial R_{\alpha x}} \right.$$
$$\left. + D_{\mu\nu}^X \frac{\partial S_{\mu\nu}^X}{\partial R_{\alpha x}} \sum_{C \in X} (\gamma_{AC} + \gamma_{\alpha C}) \Delta q_C^X + 2 D_{\mu\nu}^X \frac{\partial P_{\mu\nu}^X}{\partial R_{\alpha x}} \right\}$$
$$+ \Delta q_\alpha^X \sum_{(A \neq \alpha) \in X} \Delta q_A^X \frac{\partial \gamma_{A\alpha}}{\partial R_{\alpha x}} + \sum_{(A \neq \alpha) \in X} \frac{\partial E_{A\alpha}^{rep}}{\partial R_{\alpha x}} \tag{21}$$

The derivative of the HOP matrix $P_{\mu\nu}^X$ in FMO-DFTB is the same as in regular FMO.[88] In the SCC-DFTB energy gradient for FMO, we did not include the derivatives of MO coefficients with respect to nuclear coordinates. The internal Lagranian $W_{\mu\nu}'^X$ in eq 21 is defined by subtracting the ESP contribution from the usual Lagrangian $W_{\mu\nu}^X$

$$W_{\mu\nu}'^X = W_{\mu\nu}^X - \frac{1}{2} \sum_{\rho\sigma} D_{\mu\rho}^X V_{\rho\sigma}^X D_{\sigma\nu}^X \tag{22}$$

and

$$W_{\mu\nu}^X = \frac{1}{2} \sum_{\rho\sigma} D_{\mu\rho}^X H_{\rho\sigma}^X D_{\sigma\nu}^X \tag{23}$$

where $D_{\mu\nu}^X$ is the density matrix of fragment X.

Second, we need to calculate the derivative of the embedding energy in eq 13. For atom α in dimer IJ ($\alpha \in IJ$) one obtains

$$\frac{\partial \Delta E_{IJ}^V}{\partial R_{\alpha x}} = \Delta \Delta q_\alpha^{IJ} \sum_{K \neq I,J}^N \sum_{C \in K} \Delta q_C^K \frac{\partial \gamma_{\alpha C}}{\partial R_{\alpha x}}$$
$$+ \sum_{\mu \in \alpha} \sum_\nu \left(\Delta \tilde{W}_{\mu\nu}^{IJ,\alpha} S_{\mu\nu}^{IJ} + \Delta D_{\mu\nu}^{IJ} \frac{\partial S_{\mu\nu}^{IJ}}{\partial R_{\alpha x}} \right) \sum_{K \neq I,J}^N \sum_{C \in K} \gamma_{\alpha C} \Delta q_C^K \tag{24}$$

and for $\alpha \in (K \neq I, J)$,

$$\frac{\partial \Delta E_{IJ}^V}{\partial R_{\alpha x}} = \Delta q_\alpha^K \sum_{A \in IJ} \Delta \Delta q_A^{IJ} \frac{\partial \gamma_{A\alpha}}{\partial R_{\alpha x}}$$
$$+ \sum_{\mu \in \alpha} \sum_\nu \left(\tilde{W}_{\mu\nu}^{K,\alpha} S_{\mu\nu}^K + D_{\mu\nu}^K \frac{\partial S_{\mu\nu}^K}{\partial R_{\alpha x}} \right) \sum_{A \in IJ} \gamma_{A\alpha} \Delta \Delta q_A^{IJ} \tag{25}$$

where

$$\tilde{W}_{\mu\nu}^{X,\alpha} = -\frac{1}{2} \sum_{\rho\sigma} D_{\mu\rho}^X \frac{\partial S_{\rho\sigma}^X}{\partial R_{\alpha x}} D_{\sigma\nu}^X \tag{26}$$

$$\Delta \tilde{W}_{\mu\nu}^{IJ,\alpha} = \tilde{W}_{\mu\nu}^{IJ,\alpha} - (\tilde{W}_{\mu\nu}^{I,\alpha} \oplus \tilde{W}_{\mu\nu}^{J,\alpha}) \tag{27}$$

and

$$\tilde{W}_{\mu\nu}^{I,\alpha} \oplus \tilde{W}_{\mu\nu}^{J,\alpha} = \tilde{W}_{\mu\nu}^{I,\alpha} + \tilde{W}_{\mu\nu}^{J,\alpha} \quad \text{for} \quad \mu\nu \in I, J$$
$$= \tilde{W}_{\mu\nu}^{I,\alpha} \qquad \text{for} \quad \mu\nu \in I \text{ and } \mu\nu \notin J$$
$$= \tilde{W}_{\mu\nu}^{J,\alpha} \qquad \text{for} \quad \mu\nu \in J \text{ and } \mu\nu \notin I$$
$$= 0 \qquad \text{for} \quad \text{all other cases}$$

The density difference matrix $\Delta D_{\mu\nu}^{IJ}$ is

$$\Delta D_{\mu\nu}^{IJ} = D_{\mu\nu}^{IJ} - (D_{\mu\nu}^I \oplus D_{\mu\nu}^J) \tag{28}$$

The gradient of the ES-DIM approximation is obtained by differentiating eq 16. For the derivative with respect to the x coordinate of atom $\alpha \in I$,

$$\frac{\partial E'_{IJ}}{\partial R_{\alpha x}} = \Delta q_\alpha^I \sum_{B \in J} \Delta q_B^J \frac{\partial \gamma_{\alpha B}}{\partial R_{\alpha x}} + \sum_{\mu \in \alpha} \sum_\nu \left(\tilde{W}_{\mu\nu}^{I,\alpha} S_{\mu\nu}^I \right.$$
$$\left. + D_{\mu\nu}^I \frac{\partial S_{\mu\nu}^I}{\partial R_{\alpha x}} \right) \sum_{B \in J} \Delta q_B^J \gamma_{\alpha B} \tag{29}$$

and for $\alpha \in J$,

$$\frac{\partial E'_{IJ}}{\partial R_{\alpha x}} = \Delta q_\alpha^J \sum_{A \in I} \Delta q_A^I \frac{\partial \gamma_{A\alpha}}{\partial R_{\alpha x}} + \sum_{\mu \in \alpha} \sum_\nu \left(\tilde{W}_{\mu\nu}^{J,\alpha} S_{\mu\nu}^J \right.$$
$$\left. + D_{\mu\nu}^J \frac{\partial S_{\mu\nu}^J}{\partial R_{\alpha x}} \right) \sum_{A \in I} \Delta q_A^I \gamma_{A\alpha} \tag{30}$$

Equations 29 and 30 describe the gradients of the electrostatic (charge−charge) interaction between two fragments.

2.4. Dispersion Interaction for FMO-DFTB. We implemented the dispersion interaction model based on the formulation by Zhechkov et al.[89] The dispersion interaction E^{disp} is independent of the electronic structure and is given as a

D

dx.doi.org/10.1021/ct500489d | J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

sum of all atomic pairwise contributions E_{AB}^{disp} for atoms A and B,[89]

$$E^{disp} = \sum_{A>B} E_{AB}^{disp} \qquad (31)$$

In FMO, the above equation becomes for the dispersion energy of fragment X,

$$E_X^{disp} = \sum_{A>B \in X} E_{AB}^{disp} \qquad (32)$$

E_X^{disp} is added to the internal energy in eq 9. Also, in the separated dimer approximation, we add the dispersion contribution to $E_{IJ}' - E_I' - E_J'$ following eq 16,

$$\Delta E_{IJ}^{disp} = \sum_{A \in I} \sum_{B \in J} E_{AB}^{disp} \qquad (33)$$

Parameters for the dispersion interaction are taken from the work by Rappé et al.[54] We denote the dispersion interaction with the suffix of "-D" hereafter. Although we implemented also the dispersion interaction formulated by Elstner et al.,[52] we use in the present work the formulation by Zhechkov et al. for actual calculations. We verified by numeric tests that the dispersion energy in FMO-DFTB-D is identical to that in DFTB-D, as it should be because it is a parametrized ad hoc correction independent of fragmentation.

2.5. Computational Details. The DFTB and FMO-DFTB methods were implemented in GAMESS-US.[90] In the DFTB implementation, the Broyden charge mixing[91] was used to accelerate the convergence of SCF calculations with the same type of thresholds as in the rest of GAMESS. In both FMO-DFTB and full DFTB gradient calculations, the derivatives of the Hamiltonian and overlap matrix elements were calculated numerically, whereas other derivatives were computed analytically.

FMO-DFTB was parallelized with the generalized distributed data interface (GDDI)[92] by assigning one monomer or dimer calculation per GDDI group. The main cost of a DFTB calculation is the matrix diagonalization, which is hard to efficiently parallelize for the matrix sizes we encounter in typical FMO fragments. Therefore, we always used the group size of 1 CPU core, that is, individual monomer and dimer calculations in FMO-DFTB were executed sequentially on different CPU cores, whereas the load balancing of distributing these calculations was dynamic, as implemented in GAMESS/GDDI.[92]

All computations of accuracy and scaling tests were performed using a PC cluster consisting of several nodes equipped with dual E5-2650 Xeon CPUs (2.0 GHz, 8 cores each), 64 GB of DDR3-1600 memory and a SATA hard disk. The nodes were connected with Infiniband (used in the IP mode). For the parallel efficiency tests we used 1 to 8 nodes (16−128 CPU cores).

We used the mio[43] set of DFTB parameters, except when performing calculations on DNA sequences, for which we used the matsci[93] set. Both parameter sets are freely available from the DFTB Web site.[94] The ES-DIM approximation was used with the threshold value of 2.0 (unitless[86]) applied to the interfragment distance, defined as the separation between two closest atoms in the two fragments divided by the sum of the atomic van der Waals radii.[86] For a typical case of the contact atoms being O and H, the distance of 2.0 corresponds to 9.83 bohr. The DFTB parameters are such that the interaction for

distances exceeding 10 bohr are usually zero, which justifies the neglect of the repulsive energy term in the ES-DIM approximation (see eqs 15 and 16).

Geometry optimizations were performed until the root-mean-square (rms) and maximum gradient values became smaller than $^1/_3 \times 10^{-4}$ and 10^{-4} Hartree/Bohr, respectively (OPTTOL $= 10^{-4}$ in GAMESS). Hybrid orbitals for HOP operators for sp^3 carbons were generated by performing the Pipek−Mezey localization[95] of SCC-DFTB MOs using the mio parameters for methane in its equilibrium geometry at the same level of theory. The fragmentation of polypeptides was performed at the C_α atoms, as usual in FMO. For DNA segments, we compared two fragmentation schemes, as described below.

The accuracy of single point energy calculations was evaluated for extended polyalanine (COMe−(Ala)$_n$−NHMe, $n = 10, 20, ..., 200$) partitioned into fragments containing 1, 2, 4, and 5 amino acid residues. The accuracy of optimized geometries is discussed for (a) the α-helix, β-turn, and extended form of polyalanine COMe−(Ala)$_{20}$−NHMe, (b) a short double stranded DNA segment of CGATCG nucleotides (PDB: 1AGL), (C$_{115}$P$_{10}$N$_{46}$H$_{136}$O$_{78}$), and (c) a small Trp-cage protein (PDB: 1L2Y) with all residues neutralized using GaussView 5.0.[96]

For Trp-cage we used the PDB structure as the initial geometry, whereas for all other systems the initial structures were constructed using the "protein" and the "nucleic" programs implemented in TINKER 6.0,[97] and they were subsequently optimized using the "minimize" program with the AMBER99 force field parameters.[98] Geometry optimizations[99] with full SCC-DFTB and FMO-DFTB were then performed using RUNTYP=OPTIMIZE in GAMESS, except for the optimization of the ~1 million atoms system, for which we used RUNTYP=OPTFMO, and LBFGS when dispersion interaction is included.

For the evaluation of the FMO-DFTB scaling in terms of the system size, we used a series of globular water clusters ((H$_2$O)$_n$, $n = 256, 512, 1024, 2048, 3072, 4096, 5120,$ and 6144), the α-helix of polyalanine (COMe−(Ala)$_n$−NHMe, $n = 100, 200, ...,$ 1900, and 1998), and a quadratic slab of C$_{60}$ fullerite with the square slabs containing from 1×1 to 7×7 unit cells of face centered cubic (fcc) fullerite with the fixed thickness of 3 layers of fullerenes (only its square length was varied, not the thickness). Water clusters were produced and optimized using the "addxyz" and the "minimize" modules in TINKER. For modeling the fullerite slab, a single C$_{60}$ was first optimized with SCC-DFTB, then the optimized C$_{60}$ molecules were aligned according to the fcc symmetry of fullerite with the unit cell length[100] of 14.04 Å.

The parallel efficiency was evaluated on a 40×40 lattice of fullerite, containing 590 520 atoms. Finally, we performed a full geometry optimization of a 53×53 unit cells of fullerite, containing 1 030 440 atoms, using 128 CPU cores (8 nodes). We performed this optimization with and without the dispersion interaction. This fullerite slab has the size of 74.4 nm $\times$ 74.4 nm. For comparison, the only inorganic system of comparable size fully optimized with FMO−DFT is the boron nitride (BN) nanoribbon ring with a diameter of 105 nm,[68] which contains a much smaller number of atoms (7878).

3. RESULTS AND DISCUSSION

3.1. Accuracy of FMO-DFTB. The accuracy of FMO-DFTB compared to full SCC-DFTB was evaluated for the
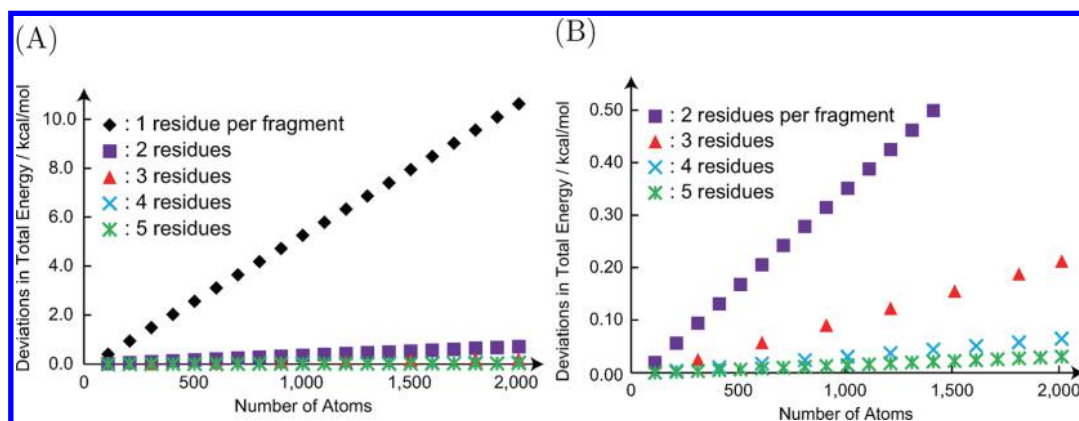
Figure 1. (A) Deviations of the total energy of FMO-DFTB from full SCC-DFTB for the extended polyalanine (COMe−(Ala)$_n$−NHMe, n = 10, 20, ..., 200). To show the deviations better, the same set of errors is plotted in (B) with a magnified scale.

Table 1. FMO-DFTB and Full SCC-DFTB Total Energies (Hartree) and rmsd (Å) for COMe−(Ala)$_{20}$−NHMe between Geometries Optimized with the Two Methods

	conformer	1 res./fragm.	2 res./fragm.	4 res./fragm.	5 res./fragm.	full SCC-DFTB
energy[a]	α-helix	−268.521608	−268.521370	−268.521072	−268.521342	−268.521568
	β-turn	−268.489373 (+20.2)	−268.490882 (+19.1)	−268.491555 (+18.7)	−268.491621 (+18.7)	−268.491889 (+18.6)
	extended	−268.445252 (+47.9)	−268.445374 (+47.7)	−268.445371 (+47.7)	−268.445394 (+47.7)	−268.445325 (+47.8)
rmsd[b]	α-helix	0.031 (0.025)	0.044 (0.039)	0.028 (0.020)	0.008 (0.008)	0.000
	β-turn	0.118 (0.119)	0.082 (0.032)	0.081 (0.033)	0.031 (0.021)	0.000
	extended	0.148 (0.055)	0.124 (0.050)	0.099 (0.000)	0.175 (0.000)	0.000

[a]The values in parentheses are the energies relative to the α-helix (kcal/mol). [b]rmsd values are shown for structures optimized with FMO-DFTB and full SCC-DFTB, starting from the same initial geometry. The values in parentheses almost always correspond to different local minima, obtained by taking FMO-DFTB structures as initial structures and reoptimizing them with full SCC-DFTB.

extended form of polyalanine (Figure 1) using SCC-DFTB. We note that throughout this section, all numerical results for FMO-DFTB were obtained with SCC-DFTB combined with FMO, so we use FMO-DFTB as a short-hand notation instead of FMO-SCC-DFTB. The number of atoms ranges from 112 to 2012, and we compared five fragmentations of n residues per fragment (n = 1, ..., 5). The error decreases with the fragment size, as expected.

The largest error is observed for the fragmentation of 1 residue per fragment: for 2012 atoms the error in total energy is about 10.6 kcal/mol. When the fragment size is doubled, this error is dramatically reduced to only 0.72 kcal/mol for the largest system, and by using a fragmentation of 3 (the edge fragments contain 4 residues), 4 and 5 residues per fragment, the errors become 0.21, 0.07, and 0.03 kcal/mol, respectively. The error is affected by the semantic periodicity (i.e., not a strict geometric periodicity) of the extended form, with the period of two residues. The CO groups face up and down alternatively, resulting in the two adjacent fragments having an opposite dipole moment.

To test the accuracy of the ES-DIM approximation (eq 16), we performed a full calculation of the interfragment Coulombic interactions for comparison. The difference of the total energy computed with and without the approximation was negligibly small, 1.02×10^{-3} kcal/mol for COMe−(Ala)$_{200}$−NHMe of the extended form with 1 residue per fragment. Thus, we see that the ES-DIM approximation is very accurate in FMO-DFTB, in agreement with other FMO studies.[72] It is also very efficient computationally, reducing the timing for the extended form of COMe−(Ala)$_{200}$−NHMe with 1 residue per fragment from 171 to 10 s on 1 CPU core.

We note that the error of FMO in general and FMO−DFT in particular is known to increase with the basis size, which is related to the increased importance of interfragment quantum effects (such as charge transfer).[72,75,101] On the other hand, FMO errors for minimum basis sets such as STO-3G are known to be very small.[72] DFTB also uses a minimum basis set, hence the accuracy of FMO-DFTB is high.

To demonstrate that the dispersion in FMO-DFTB-D is exact, we performed a single point calculation with FMO-DFTB-D and full DFTB-D for COMe−(Ala)$_{200}$−NHMe of the extended form with n residues per fragment (n = 1, ..., 5). The dispersion energy in full DFTB-D and FMO-DFTB-D was exactly identical, 8.894040160 Hartree. The gradient of our dispersion implementation is also exact.

Next, to check the accuracy of optimized structures, we performed geometry optimizations with FMO-DFTB and full SCC-DFTB using fragments containing 1, 2, 4, and 5 residues and dividing COMe−(Ala)$_{20}$−NHMe into 20, 10, 5, and 4 fragments, respectively. Table 1 shows the relative energies and root-mean-square deviation (rmsd) for the optimized structures of the α-helix, β-turn, and extended form. Hydrogen atoms are included and mass-weighting is not used when rmsd values are calculated.

Full SCC-DFTB predicts the α-helix to be the most stable isomer, with the β-turn and extended form being less stable by 18.6 and 47.8 kcal/mol, respectively. This trend is in qualitative agreement with FMO-MP2 results[102] for COMe−(Ala)$_{10}$−NHMe. To QM contributions, one can add solvent[102] and finite temperature[103] corrections. The relative energies of the three isomers of polyalanine obtained with FMO-DFTB are in good agreement with full SCC-DFTB: the largest deviation for relative energies of the β-turn is 1.6 kcal/mol for 1 residue per

Table 2. rmsd between FMO-DFTB and Full SCC-DFTB Optimized Geometrical Parameters for COMe−(Ala)$_{20}$−NHMe[a]

conformer	n	bond length (Å)	bond angle (deg)	ϕ (deg)	ψ (deg)	ω (deg)
α-helix	1	0.0004 (0.0014)	0.065 (0.398)	0.80 (2.50)	0.47 (0.95)	0.22 (0.74)
	2	0.0003 (0.0010)	0.045 (0.178)	0.86 (2.53)	0.49 (1.36)	0.31 (0.69)
	4	0.0002 (0.0005)	0.017 (0.056)	0.33 (0.84)	0.20 (0.51)	0.13 (0.30)
	5	0.0001 (0.0003)	0.009 (0.030)	0.19 (0.46)	0.09 (0.21)	0.05 (0.08)
β-turn	1	0.0004 (0.0018)	0.124 (0.880)	2.83 (10.57)	1.52 (4.56)	0.73 (1.70)
	2	0.0002 (0.0009)	0.075 (0.336)	0.92 (2.05)	1.03 (3.71)	0.40 (1.08)
	4	0.0002 (0.0010)	0.065 (0.414)	0.90 (2.79)	1.01 (3.39)	0.39 (0.99)
	5	0.0001 (0.0006)	0.036 (0.216)	0.52 (1.76)	0.35 (0.87)	0.18 (0.33)
extended	1	0.0002 (0.0008)	0.055 (0.231)	0.35 (0.68)	0.27 (0.56)	0.21 (0.36)
	2	0.0001 (0.0004)	0.028 (0.121)	0.21 (0.48)	0.35 (0.63)	0.15 (0.29)
	4	0.0001 (0.0002)	0.022 (0.101)	0.21 (0.49)	0.30 (0.58)	0.13 (0.27)
	5	0.0001 (0.0004)	0.037 (0.154)	0.32 (0.62)	0.34 (0.65)	0.21 (0.58)

[a]n indicates the number of fragments per residue, and the values in parentheses are maximum deviations. ϕ, ψ, and ω define the dihedral angles of C′(i−1)−N(i)−C$_\alpha$(i)−C′(i), N(i)−C$_\alpha$(i)−C′(i)−N(i+1), and C$_\alpha$(i)−C′(i)−N(i+1)−C$_\alpha$(i+1), respectively, where i is the residue number.

fragment and it is 0.1 kcal/mol for 5 residues per fragment. The relative energies of the extended form have errors of only 0.1 kcal/mol. With the partitioning of 2 residues per fragment, an accuracy of 1 kcal/mol with respect to the full SCC-DFTB calculation is achieved for all relative energies.

The optimized structures are also quite accurate with rmsd values not exceeding 0.175 Å, as shown in Table 1. The presence of many local minima complicates a systematic comparison, because the minimum search may converge to different minima in FMO-DFTB and full SCC-DFTB. Although the difference of gradient elements between these two methods is small, the geometry optimization pathway becomes notably different for the case of flat potential energy surfaces (PESs). In order to eliminate the problem associated with the existence of many local minima, in addition to the optimization using FMO-DFTB and full SCC-DFTB using the same initial geometry, we reoptimized the structures at the full SCC-DFTB level starting from FMO-DFTB minima. The rmsd values obtained in this way gives the true deviation in the quality of optimized geometry between FMO-DFTB and full SCC-DFTB. These values, given in parentheses in Table 1, becomes much smaller than those outside parentheses for the fragmentations of 2 residues per fragment or more, whereas for 1 residue per fragment the improvement is not uniform. This shows that for 1 residue-fragmentation the error is not always related to the fact that the local minima are different, consistent with the larger error in total energies for this particular choice of fragment size. For the partitioning of 2 residues per fragment the rmsd values for the same local minima are 0.050 Å or smaller. This improvement is especially pronounced for the extended form, which is particularly flexible: the rmsd values of 0.124, 0.099, and 0.175 (Å) for 2, 4, and 5 residues per fragment when a different set of minima is compared become 0.050, 0.000, and 0.000 (Å) when the optimized FMO-DFTB structure is used as initial geometry in full SCC-DFTB geometry optimizations.

Table 2 shows a comparison of bond lengths and angles computed at the minima obtained with FMO-DFTB and full SCC-DFTB. The accuracy in reproducing bond lengths and bond angles is satisfactory, whereas some dihedral peptide angles vary by as much as 10.6° for the fragmentation of 1 residue per fragment, reduced to 3.7° by doubling the fragment size. The accuracy of FMO-DFTB in terms of rmsd, bond lengths and angles is similar or better compared to FMO-RHF/

3-21G,[99] although we note that we use a polypeptide with twice as many residues, so the results are not directly comparable.

For the DNA benchmark system, a CGATCG double-stranded helix segment was optimized with both full SCC-DFTB and FMO-DFTB methods. We investigated two alternative fragmentation schemes by detaching bonds C−C bonds at the 4′ and 5′ carbon atoms as schematically shown in Figure 2A. The two resulting optimized structures are superimposed on the full SCC-DFTB geometry in Figure 2B. Compared to full SCC-DFTB, the difference of the total energy and the rmsd for optimized FMO-DFTB structures with the fragmentation at 5′ carbon atoms was +2.12 kcal/mol and 0.071 Å, respectively. On the other hand, for the fragmentation at 4′ carbon atoms the energy difference and rmsd of the optimized structures was −2.22 kcal/mol and 0.402 Å, respectively. The full SCC-DFTB single point calculation at the FMO-DFTB optimized geometry (with the fragmentation at 4′ carbon atoms) shows that FMO-DFTB found a more stable local minimum by 3.92 kcal/mol than full SCC-DFTB. The geometry optimization with full SCC-DFTB starting with the FMO-DFTB optimized structure (with the fragmentation at 4′ carbon atoms) resulted in a different local minimum, for which the rmsd between FMO-DFTB and full SCC-DFTB was 0.041 Å. The larger rmsd value obtained above (0.402 Å) stems from the difference in the local minima.

We also performed geometry optimizations for the neutralized Trp-cage (PDB: 1L2Y) protein. At the minima, obtained with FMO-DFTB and full SCC-DFTB, the rmsd values were 0.067 and 0.044 Å, for the fragmentation of 1 and 2 residues per fragment, respectively, whereas the errors in the total energy were +1.20 and +0.82 kcal/mol, respectively. The small rmsd values imply that the character of the minima found by FMO-DFTB is similar to those found using full SCC-DFTB (see Figure 3). The deviation of the optimized structures at FMO-DFTB is very small as demonstrated by their small rmsd values. Most importantly, the positions of backbone atoms and hydrogen bonds are well reproduced by FMO-DFTB.

3.2. Scaling of FMO-DFTB. For a comparison of timings, we performed FMO-DFTB and full SCC-DFTB calculations of water clusters, and the corresponding data is shown in Table 3. The calculation for a water cluster containing 12 288 atoms took 5.9 days for full SCC-DFTB, while FMO-DFTB took 109.7 s, with a speed-up factor of 4646. An extrapolation suggests that full SCC-DFTB calculation will take about 3
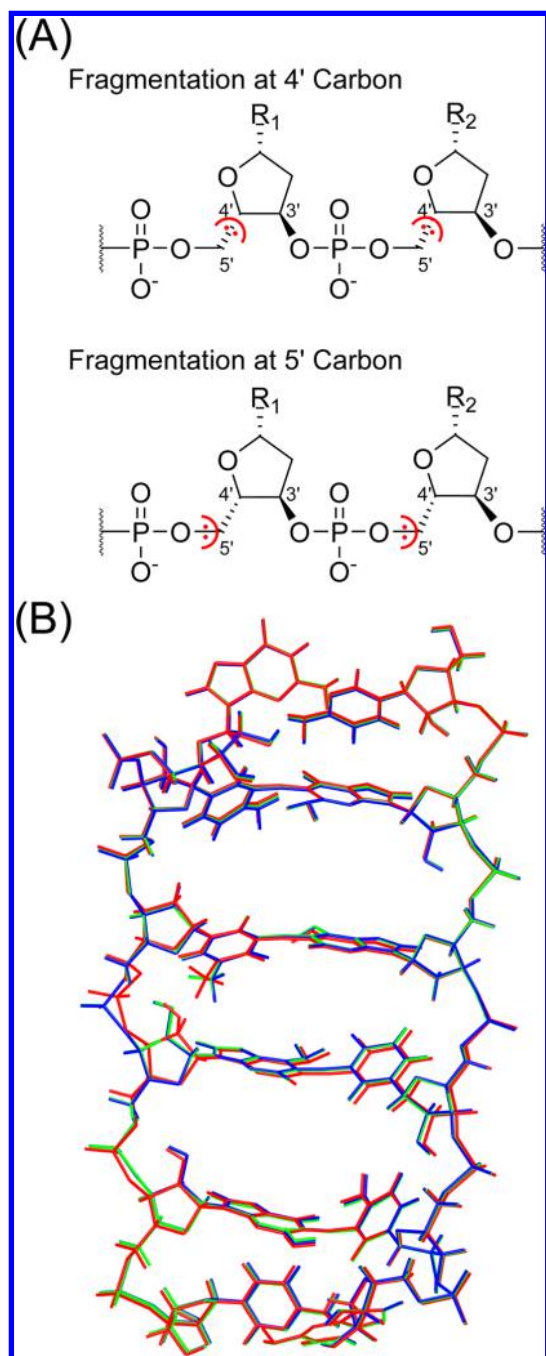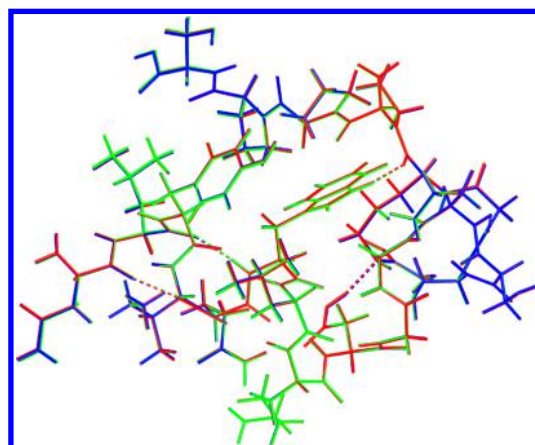
Figure 2. (A) Two alternative fragmentation schemes detaching C−C bonds at 4′ and 5′ carbon atoms, where red dots and curves show the fragment to which the detached covalent bond is assigned. (B) Overlay of the optimized structures of the CGATCG DNA. The blue, red and green lines represent the optimized structures obtained with full SCC-DFTB and FMO-DFTB with the fragmentation at 4′ and 5′ carbon atoms, respectively.

Figure 3. Superposition of the optimized structures of the neutralized Trp-cage protein. The blue, red and green lines are used to plot the optimized structures obtained with full SCC-DFTB and FMO-DFTB with 1 and 2 residues per fragment, respectively. The rmsd between FMO-DFTB and full SCC-DFTB was 0.067 and 0.044 Å for 1 and 2 residues per fragment, respectively. Dotted lines show hydrogen bonds.

Table 3. Comparison of Wall-Clock Timings (Seconds) for $(H_2O)_n$ Clusters Calculated with Full and FMO-Based SCC-DFTB on 1 Xeon CPU Core[a]

n	full SCC-DFTB	FMO-DFTB
256	74.2	4.6
512	865.4	9.8
1024	7725.2	20.8
2048	63075.0	48.6
3072	221243.4	78.5
4096	509745.0	109.7
5120	(1030224.1)	154.1
6144	(1810687.4)	198.8
Scaling	$O(N^{3.09})$	$O(N^{1.21})$

[a]The values in parentheses are extrapolated estimates.

weeks to finish the single point energy calculation of 18 432 atoms, whereas FMO-DFTB took only about 3 min.

The scaling of FMO-DFTB was evaluated for the α-helix of polyalanine, fullerite slabs, and water clusters containing up to about 20 000 atoms. The fragmentation was 1 residue per fragment for polyalanine and 1 molecule per fragment otherwise. The total wall-clock timings are shown in Figure 4, divided for the purpose of analysis into four steps: preparation (reading input file, processing fragmentation, etc.), SCF1 (monomer SCF), SCF2 (dimer SCF), and ES2 (ES-DIM approximation).

The scalings based on the total wall-clock time of the α-helix, fullerite slab, and water cluster are $O(N^{1.10})$, $O(N^{1.14})$, and $O(N^{1.21})$, respectively. The preparation and ES2 steps do not require much time; thus, the main time-consuming steps of FMO-DFTB are the calculations of monomers (SCF1) and dimers (SCF2), as in other FMO implementations. The number of monomers is N and the number of SCF dimers is proportional to N within the ES-DIM approximation. However, although the number of SCF calculations scales linearly with system size, in each of them we have to consider the ESP of the whole system. It is the inclusion of the embedding ESP that causes the increase in FMO-DFTB scaling from linear $O(N^{1.0})$ to the observed behavior of $O(N^{1.1-1.2})$. The computational bottleneck for all three examples given above was the SCF2 dimer calculation step.

3.3. Parallelization of FMO-DFTB and Its Application to a Fullerite Geometry Optimization. As described above, we parallelized FMO-DFTB in a fashion similar to other FMO methods, with each group of CPU cores doing a monomer or dimer calculation. In FMO-DFTB, we chose to assign 1 CPU core per group. To test the parallelization efficiency, we used a fullerite slab with 40 × 40 unit cells, $(C_{60})_{9842}$, containing 590
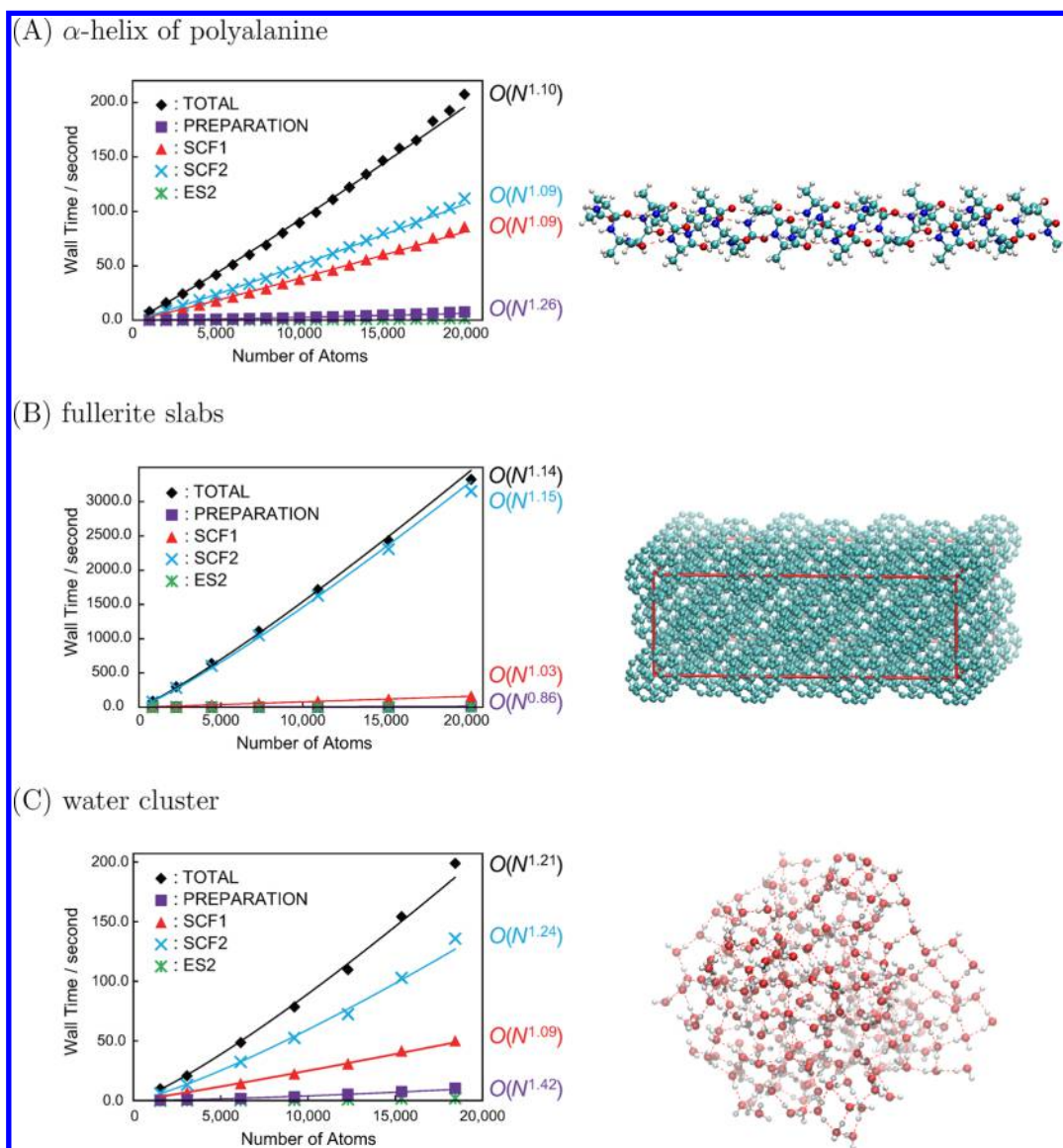
Figure 4. Timing of FMO-DFTB for (A) α-helices of polyalanine, (B) fullerite slabs, and (C) water cluster (in left), as measured on 1 Xeon CPU core. The representative structures of COMe$-$(Ala)$_{20}$$-$NHMe, 3 $\times$ 3 unit cells of fullerite, and (H$_2$O)$_{256}$ are shown on the right.

520 atoms. As in the above-described three benchmark systems, the computational bottleneck for the fullerite calculations was again the SCF2 dimer calculation step, and the required time for this step was even more significant, since each dimer contained 120 atoms. The SCF2 step took 92$-$95% of the total time, whereas the preparation and ES2 steps are almost negligible. We performed a single point energy calculation using 16, 32, 64, 96, and 128 CPU cores (1$-$8 nodes). The speed-up measured by the wall-clock time with a 1-node (16 CPU cores) calculation as reference is shown in Figure 5.

We achieved a 7.5 time speed-up on 128 CPU cores (the parallel efficiency of 93.8%). There is some loss of efficiency because of the load balancing and because a few steps remain sequential. In addition, there are no intragroup communications, so the deviation of the observed speed-up from the perfect results is mainly due to of the data exchange between groups, because of the large number of atoms and fragments. The data communication took 68.1 s for the total amount of 2.35 gigabytes of data. Table 4 shows the timings for the

geometry optimizations of some representative systems using our parallelized FMO-DFTB.

For demonstration purposes, we calculate the fullerite cluster consisting of 53 $\times$ 53 unit cells and containing 1 030 440 atoms. A single point calculation on 128 cores (8 nodes) took 4993 s (83.2 min). This relatively long wall-clock time is related to the fact that the fragment size is big: 60 atoms per fragment (one fullerene molecule per fragment). The equilibrium geometry of the fullerite containing over one million atoms was found in 16 steps, which took 99.4 h on 8 Xeon nodes (128 cores). C$_{60}$ is a nonpolar system and the interaction between fullerene molecules is governed by van der Waals forces. Thus, we repeated the geometry optimization of the fullerite system with a posteriori dispersion using FMO-DFTB-D. To accelerate geometry optimization, we constructed a new initial structure using the optimized cell length of 4 $\times$ 4 fullerite slab. Inclusion of dispersion interaction increased the wall-clock time by approximately 1.5% per geometry optimization cycle, and a minimum geometry with shorter cell length was found after 60 iterations. We found that the experimental cell length[100] of

I

dx.doi.org/10.1021/ct500489d | J. Chem. Theory Comput. XXXX, XXX, XXX$-$XXX
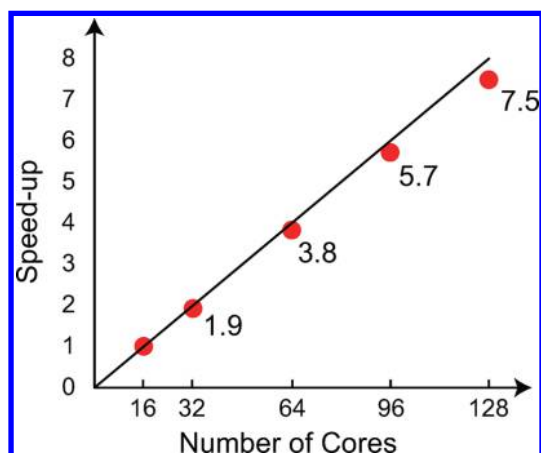
Figure 5. Speed-up of parallel FMO-DFTB calculations for the $(C_{60})_{9842}$ fullerite model system with respect to the number of cores are shown as red filled circles. The solid black line represents the ideal scaling with 100% parallel efficiency. The speed-up on 1 node (16 CPU cores) is set to 1.0.

Table 4. Numbers of Atoms, N_{AT}, Fragments, N, Geometry Steps, N_{steps}, and Wall-Clock Timings, T (min), for FMO-DFTB Geometry Optimizations, Performed on 1 Xeon Node (16 CPU Cores) Using RUNTYP=OPTFMO

system	N_{AT}	N	N_{steps}	T (min)
COMe−(Ala)$_{20}$−NHMe (α-helix)	212	20	431	2.8
COMe−(Ala)$_{20}$−NHMe (β-turn)	212	20	265	1.7
COMe−(Ala)$_{20}$−NHMe (extended)	212	20	191	1.0
CGATCG DNA (fragmentation at 4′ carbon)	385	12	1788	41.1
CGATCG DNA (fragmentation at 5′ carbon)	385	12	2659	58.4
Trp-cage protein	303	20	675	7.0
Trp-cage protein	303	10	674	11.9

14.04 Å after geometry optimization became 13.91 and 14.02 Å with and without dispersion, respectively.

4. CONCLUSIONS

We have derived and implemented the combination of the density-functional tight-binding method (DFTB) with the fragment molecular orbital (FMO) method, for both SCC and NCC versions of DFTB, including a posteriori dispersion interaction. Full DFTB and FMO-DFTB methods have been implemented in GAMESS-US. We have parallelized FMO-DFTB with GDDI and demonstrated its reasonable parallel efficiency.

We have shown that the total energy of full SCC-DFTB calculations is reproduced in FMO-DFTB within 1 kcal/mol for a 2000 atom polypeptide, when the system is partitioned with fragment sizes of two or more residues. The structures optimized with FMO-DFTB had rmsd values of 0.1 Å or less, compared to the minima obtained with full SCC-DFTB. The comparison was somewhat complicated by the existence of many local minima on the corresponding PESs, allowing FMO-DFTB and SCC-DFTB to converge to different stationary points when starting geometry optimizations from the same initial geometries. FMO-DFTB-D exactly reproduces the dispersion interaction of full DFTB-D calculations.

The computational scaling of FMO-DFTB with the number of atoms N was shown to be $O(N^{1.1-1.2})$ for 1D, 2D, as well as 3D systems, a little bit higher than the linear scaling. This is a huge improvement over the scaling of full SCC-DFTB, which is approximately cubic. The energy of a water cluster with 18 432 atoms was calculated with FMO-DFTB in about 3 min, whereas the corresponding full SCC-DFTB calculation is estimated to require about 3 weeks, which is about 4 orders of magnitude longer.

We have demonstrated the efficiency of FMO-DFTB by optimizing the structure of fullerite containing more than 1 million of atoms. Clearly, although purely periodic systems can be computed with periodic boundary conditions, the cluster approach is very useful for the structures with defects, doping and adsorption on the surface as well as for amorphous and other nonperiodic systems, which are very frequent in material science. In future, we plan to implement and develop the third-order DFTB,[44,104] important for practical applications in particular in relation to biosystems. FMO has been applied to a number of biochemical[105] and inorganic[66−70] systems,[59] although these simulations required considerable computational resources, sometimes top-class supercomputers.[27] With the introduction of FMO-DFTB, these simulations can be performed on small scale PC clusters or even single desktop computers, greatly increasing the accessibility of FMO to the end users.

■ AUTHOR INFORMATION

Corresponding Authors
*Phone: +81 (0)29 851-5426. Email: d.g.fedorov@aist.go.jp.
*Phone: +81 (0)52 747-6397. Email: sirle@chem.nagoya-u.ac.jp.

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P. Nature 2013, 497, 643−646.
(2) Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. J. Chem. Theory Comput. 2011, 2, 1789−1793.
(3) Sena, A. M. P.; Miyazaki, T.; Bowler, D. R. J. Chem. Theory Comput. 2011, 7, 884−889.
(4) Hirano, T.; Sato, F. Phys. Chem. Chem. Phys. 2014, 16, 14496−14503.
(5) Yang, W. Phys. Rev. Lett. 1991, 66, 1438−1441.
(6) Goedecker, S.; Scuseria, G. E. Comp. Sci. Eng. 2003, 5, 14−21.
(7) Linear-Scaling Techniques in Computational Chemistry and Physics; Zalesny, R., Papadopoulos, M. G., Mezey, P. G., Leszczynski, J., Eds.; Springer: New York, 2011.
(8) Nikitina, E.; Sulimov, V.; Zayets, V.; Zaitseva, N. Int. J. Quantum Chem. 2004, 97, 747−763.

(9) Stewart, J. J. P. *J. Mol. Model.* **2009**, *15*, 765−805.
(10) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. *Chem. Rev.* **2012**, *112*, 632−672.
(11) Otto, P.; Ladik, J. *Chem. Phys.* **1975**, *8*, 192−200.
(12) Gao, J. *J. Phys. Chem. B* **1997**, *101*, 657−663.
(13) Kobayashi, M.; Nakai, H. *J. Chem. Phys.* **2013**, *138*, 044102.
(14) He, X.; Merz, K. M. *J. Chem. Theory Comput.* **2010**, *6*, 405−411.
(15) Yu, H.; Leverentz, H. R.; Bai, P.; Siepmann, J. I.; Truhlar, D. G. *J. Phys. Chem. Lett.* **2014**, *5*, 660−670.
(16) Tong, Y.; Mei, Y.; Zhang, J. Z. H.; Duan, L. L.; Zhang, Q. G. *J. Theor. Comp. Chem.* **2009**, *8*, 1265.
(17) Söderhjelm, P.; Kongsted, J.; Ryde, U. *J. Chem. Theory Comput.* **2010**, *6*, 1726−1737.
(18) Gao, J.; Wang, Y. *J. Chem. Phys.* **2012**, *136*, 071101.
(19) Collins, M. A. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7744−7751.
(20) Frank, A.; Möller, H. M.; Exner, T. E. *J. Chem. Theory Comput.* **2012**, *8*, 1480−1492.
(21) Gordon, M. S.; Smith, Q. A.; Xu, P.; Slipchenko, L. V. *Annu. Rev. Phys. Chem.* **2013**, *64*, 553−578.
(22) Kiewisch, K.; Jacob, C. R.; Visscher, J. *J. Chem. Theory Comput.* **2013**, *9*, 2425−2440.
(23) Howard, J. C.; Tschumper, G. S. *J. Chem. Phys.* **2013**, *139*, 184113.
(24) Wang, K.; Li, W.; Li, S. *J. Chem. Theory Comput.* **2014**, *10*, 1546−1553.
(25) Sahu, N.; Yeole, S. D.; Gadre, S. R. *J. Chem. Phys.* **2013**, *138*, 104101.
(26) Huang, L.; Massa, L. *Future Med. Chem.* **2012**, *4*, 1479−1494.
(27) Ikegami, T.; Ishida, T.; Fedorov, D. G.; Kitaura, K.; Inadomi, Y.; Umeda, H.; Yokokawa, M.; Sekiguchi, S. 2005.
(28) Huang, L.; Massa, L.; Karle, J. *Proc. Nat. Acad. Sci. U.S.A.* **2009**, *106*, 1731−1736.
(29) Sawada, T.; Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. B* **2010**, *114*, 15700−15705.
(30) Ikegami, T.; Ishida, T.; Fedorov, D. G.; Kitaura, K.; Inadomi, Y.; Umeda, H.; Yokokawa, M.; Sekiguchi, S. *J. Comput. Chem.* **2010**, *31*, 447−454.
(31) Fedorov, D. G.; Alexeev, Y.; Kitaura, K. *J. Phys. Chem. Lett.* **2011**, *2*, 282−288.
(32) Mochizuki, Y.; Yamashita, K.; Fukuzawa, K.; Takematsu, K.; Watanabe, H.; Taguchi, N.; Okiyama, Y.; Tsuboi, M.; Nakano, T.; Tanaka, S. *Chem. Phys. Lett.* **2010**, *493*, 346−352.
(33) Canfield, P.; Dahlbom, M. G.; Hush, N. S.; Reimers, J. R. *J. Chem. Phys.* **2006**, *124*, 024301.
(34) Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2009**, *5*, 459−467.
(35) Sato, M.; Yamataka, H.; Komeiji, Y.; Mochizuki, Y.; Ishikawa, T.; Nakano, T. *J. Am. Chem. Soc.* **2008**, *130*, 2396−2397.
(36) Frauenheim, T.; Seifert, G.; Elstner, M.; Niehaus, T.; Kohler, C.; Amkreutz, M.; Sternberg, M.; Hajnal, Z.; Carlo, A. D.; Suhai, S. *J. Phys.: Cond. Matter* **2002**, *14*, 3015−3047.
(37) Gaus, M.; Cui, Q.; Elstner, M. *WIREs Comput. Mol. Sci.* **2014**, *4*, 49−61.
(38) Koskinen, P.; Mäkinen, V. *Comput. Mater. Sci.* **2009**, *47*, 237−253.
(39) Oliveira, A. F.; Seifert, G.; Heine, T.; Duarte, H. A. *J. Braz. Chem. Soc.* **2009**, *20*, 1193−1205.
(40) Seifert, G.; Joswig, J.-O. *WIREs Comput. Mol. Sci.* **2012**, *2*, 456−465.
(41) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864−B871.
(42) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133−A1138.
(43) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260−7268.
(44) Gaus, M.; Cui, Q.; Elstner, M. *J. Chem. Theory Comput.* **2011**, *7*, 931−948.
(45) Finnis, M. W.; Paxton, A. T.; Methfessel, M.; van Schilfgaarde, M. *Phys. Rev. Lett.* **1998**, *81*, 5149−5152.

(46) Sankey, O. F.; Demkov, A. A.; Windl, W.; Fritsch, J. H.; Lewis, J. P.; Fuentes-Cabrera, M. *Int. J. Quantum Chem.* **1998**, *69*, 327−340.
(47) Foulkes, W. M. C.; Haydock, R. *Phys. Rev. B* **1989**, *39*, 12520−12536.
(48) Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. *Phys. Rev. B* **1995**, *51*, 12947−12957.
(49) Hu, H.; Lu, Z.; Elstner, M.; Hermans, J.; Yang, W. *J. Phys. Chem. A* **2007**, *111*, 5685−5691.
(50) Giese, T. J.; Chen, H.; Dissanayake, T.; Giambau, G. M.; Heldenbrando, H.; Huang, M.; Kuechler, E. R.; Lee, T.-S.; Panteva, M. T.; Radak, B. K.; York, D. M. *J. Chem. Theory Comput.* **2013**, *9*, 1417−1427.
(51) Cawkwell, M. J.; Niklasson, A. M. N. *J. Chem. Phys.* **2012**, *137*, 134105.
(52) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149−5155.
(53) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787−1799.
(54) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.
(55) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701−706.
(56) *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*; Fedorov, D. G., Kitaura, K., Eds.; CRC Press: Boca Raton, FL, 2009.
(57) Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 6904−6914.
(58) Fedorov, D. G.; Nagata, T.; Kitaura, K. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7562−7577.
(59) Tanaka, S.; Mochizuki, Y.; Komeiji, Y.; Okiyama, Y.; Fukuzawa, K. *Phys. Chem. Chem. Phys.* **2014**, *16*, 10310−10344.
(60) Mazanetz, M. P.; Ichihara, O.; Law, R. J.; Whittaker, M. *J. Cheminf.* **2011**, *3*, 2.
(61) Sawada, T.; Fedorov, D. G.; Kitaura, K. *J. Am. Chem. Soc.* **2010**, *132*, 16862−16872.
(62) Watanabe, T.; Inadomi, Y.; Fukuzawa, K.; Nakano, T.; Tanaka, S.; Nilsson, L.; Nagashima, U. *J. Phys. Chem. B* **2007**, *111*, 9621−9627.
(63) Fukuzawa, K.; Komeiji, Y.; Mochizuki, Y.; Kato, A.; Nakano, T.; Tanaka, S. *J. Comput. Chem.* **2006**, *27*, 948−960.
(64) Fukuzawa, K.; Watanabe, C.; Kurisaki, I.; Taguchi, N.; Mochizuki, Y.; Nakano, T.; Tanaka, S.; Komeiji, Y. *Comp. Theor. Chem.* **2014**, *1034*, 7−16.
(65) Nakata, H.; Fedorov, D. G.; Yokojima, S.; Kitaura, K.; Nakamura, S. *Theor. Chem. Acc.* **2014**, *133*, 1477.
(66) Fedorov, D. G.; Avramov, P. V.; Jensen, J. H.; Kitaura, K. *Chem. Phys. Lett.* **2009**, *477*, 169−175.
(67) Carlson, P. J.; Bose, S.; Armstrong, D. W.; Hawkins, T.; Gordon, M. S.; Petrich, J. W. *J. Phys. Chem. B* **2012**, *116*, 503−512.
(68) Avramov, P. V.; Fedorov, D. G.; Sorokin, P. B.; Sakai, S.; Entani, S.; Ohtomo, M.; Y. Matsumoto, H. N. *J. Phys. Chem. Lett.* **2012**, *3*, 2003−2008.
(69) Okiyama, Y.; Tsukamoto, T.; Watanabe, C.; Fukuzawa, K.; Tanaka, S.; Mochizuki, Y. *Chem. Phys. Lett.* **2013**, *566*, 25−31.
(70) Roskop, L.; Fedorov, D. G.; Gordon, M. S. *Mol. Phys.* **2013**, *111*, 1622−1629.
(71) Sugiki, S.-I.; Kurita, N.; Sengoku, Y.; Sekino, H. *Chem. Phys. Lett.* **2003**, *382*, 611−617.
(72) Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2004**, *389*, 129−134.
(73) Chiba, M.; Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2007**, *444*, 346−350.
(74) Shimodo, Y.; Morihashi, K.; Nakano, T. *J. Mol. Struct.: THEOCHEM* **2006**, *770*, 163−168.
(75) Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2014**, *597*, 99−105.
(76) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *120*, 6832−6840.
(77) Nakano, T.; Mochizuki, Y.; Yamashita, K.; Watanabe, C.; Fukuzawa, K.; Segawa, K.; Okiyama, Y.; Tsukamoto, T.; Tanaka, S. *Chem. Phys. Lett.* **2012**, *523*, 128−133.

(78) Seifert, G.; Porezag, D.; Frauenheim, T. *Int. J. Quantum Chem.* **1996**, *58*, 185−192.

(79) Frauenheim, T.; Weich, F.; Köhler, T.; Uhlmann, S.; Porezag, D.; Seifert, G. *Phys. Rev. B* **1995**, *52*, 11492−11501.

(80) Haugk, M.; Elsner, J.; Frauenheim, T. *J. Phys.: Condens. Matter* **1997**, *9*, 7305−7315.

(81) Zheng, G.; Irle, S.; Morokuma, K. *Chem. Phys. Lett.* **2005**, *412*, 210−216.

(82) Irle, S.; Zheng, G.; Wang, Z.; Morokuma, K. *J. Phys. Chem. B* **2006**, *110*, 14531−14545.

(83) Nishimoto, Y.; Wang, Z.; Morokuma, K.; Irle, S. *Phys. Status Solidi B* **2012**, *249*, 324−334.

(84) Parr, R. G.; Pearson, R. G. *J. Am. Chem. Soc.* **1983**, *105*, 7512−7516.

(85) Kaminski, S.; Gaus, M.; Elstner, M. *J. Phys. Chem. A* **2012**, *116*, 11927−11937.

(86) Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2002**, *351*, 475−480.

(87) Nakano, T.; Kaminuma, T.; Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2000**, *318*, 614−618.

(88) Nagata, T.; Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2010**, *492*, 302−308.

(89) Zhechkov, L.; Heine, T.; Patchkovskii, S.; Seifert, G.; Duarte, H. A. *J. Chem. Theory Comput.* **2005**, *1*, 841−847.

(90) Schmidt, N. W.; Baldridge, K. K.; Baldridge, J. A.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. J.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347−1363.

(91) Johnson, D. D. *Phys. Rev. B* **1998**, *38*, 12807−12813.

(92) Fedorov, D. G.; Olson, R. M.; Kitaura, K.; Gordon, M. S.; Koseki, S. *J. Comput. Chem.* **2004**, *25*, 872−880.

(93) Frenzel, J.; Oliveira, A. F.; Jardillier, N.; Heine, T.; Seifert, G. Semi-relativistic, self-consistent charge Slater−Koster tables for density-functional based tight-binding (DFTB) for materials science simulations. TU-Dresden; 2004−2009.

(94) http://www.dftb.org/.

(95) Pipek, J.; Mezey, P. G. *J. Chem. Phys.* **1989**, *90*, 4916−4926.

(96) Dennington, R.; Keith, T.; Millam, J. *GaussView Version 5*; Semichem Inc.: Shawnee Mission, KS, 2009.

(97) Ponder, J. W. *TINKER: Software Tools for Molecular Design, 6.0*; Washington University School of Medicine: Saint Louis, MO, 2009.

(98) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049−1074.

(99) Fedorov, D. G.; Ishida, T.; Uebayasi, M.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 2722−2732.

(100) Brazhkin, V. V.; Lyapin, A. G.; Popova, S. V.; Voloshin, R. N.; Antonov, Y. V.; Lyapin, S. G.; Kluev, Y. A.; Naletov, A. M.; Mel'nik, N. N. *Phys. Rev. B* **1997**, *56*, 11465−11471.

(101) Fedorov, D. G.; Slipchenko, L. V.; Kitaura, K. *J. Phys. Chem. A* **2010**, *114*, 8742−8753.

(102) Li, H.; Fedorov, D. G.; Nagata, T.; Kitaura, K.; Jensen, J. H.; Gordon, M. S. *J. Comput. Chem.* **2010**, *31*, 778−790.

(103) Nakata, H.; Nagata, T.; Fedorov, D. G.; Yokojima, S.; Kitaura, K.; Nakamura, S. *J. Chem. Phys.* **2013**, *138*, 164103.

(104) Yang, Y.; Yu, H.; York, D.; Cui, Q.; Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 10861−10873.

(105) Alexeev, Y.; Mazanetz, M. P.; Ichihara, O.; Fedorov, D. G. *Curr. Top. Med. Chem.* **2012**, *12*, 2013−2033.

L

dx.doi.org/10.1021/ct500489d | *J. Chem. Theory Comput.* XXXX, XXX, XXX−XXX