Automatized Parametrization of SCC-DFTB Repulsive Potentials: Application to Hydrocarbons[†]

Michael Gaus,[§] Chien-Pin Chou,[‡] Henryk Witek,*[,‡] and Marcus Elstner*[,§]

Institute for Physical and Theoretical Chemistry, Technische Universität Braunschweig, Germany, and Institute of Molecular Science and Department of Applied Chemistry, National Chiao Tung University, Hsinchu, Taiwan

Received: April 1, 2009; Revised Manuscript Received: August 26, 2009

In this work, we derive and test a new automatized strategy to construct repulsive potentials for the self-consistent charge density functional tight-binding (SCC-DFTB) method. This approach allows one to explore the parameter space in a systematic fashion in order to find optimal solutions. We find that due to the limited flexibility of the SCC-DFTB electronic part, not all properties can be optimized simultaneously. For example, the optimization of heats of formation is in conflict with the optimization of vibrational frequencies. Therefore, a special parametrization for vibrational frequencies is derived. It is shown that the performance of SCC-DFTB can be significantly improved using a more elaborate fitting strategy. A new fit for C and H is presented, which results in an average error of 2.6 kcal/mol for heats of formations for a large set of hydrocarbons, indicating that the performance of SCC-DFTB can be systematically improved also for other elements.

Introduction

There exist two main approaches for theoretical determination of molecular properties. The first one is a collection of wave-function-based methods that originate from the Hartree−Fock theory. The second one is density functional theory (DFT). Both of these approaches constitute a field of intensive research activity aimed at improving the accuracy of the calculations and enhancing the computational efficiency. As an effect of this impressive development effort, various quantum chemical techniques emerged that were developed for different purposes. One of these techniques, derived as an approximation to DFT, is the self-consistent charge density functional tight-binding (SCC-DFTB) method. It was primarily developed in order to apply an approximated version of DFT to very large molecular systems, appearing in either physical or chemical considerations. Today, molecular systems of about 100 atoms can be handled by DFT on a standard desktop PC, while using SCC-DFTB, roughly 1000 atoms can be treated. SCC-DFTB is an approximate quantum chemical method that was derived from DFT by a second-order expansion of the total DFT energy with respect to density fluctuations around a suitable reference density.[1] It describes explicitly only valence electrons in minimal atomic basis sets; chemical cores are treated in an effective manner via additive two-center potentials. In addition, some of the approximations concerning the molecular matrix elements allowed for tabulating all of the necessary Hamiltonian and overlap matrix elements in so-called Slater−Koster parameter files, which eliminated the otherwise very costly integrations. The primary fields of potential applications of SCC-DFTB are the physics of solids (including crystals, amorphous materials, and semiconductors), the chemistry of biological molecules, surface chemistry and catalysis, molecular dynamic studies on a nanosecond time scale, and investigations of molecules with hundreds or thousands of conformations. While being quantitatively less accurate than the more sophisticated methods, SCC-DFTB makes such computationally demanding investigations possible, owing to its high efficiency.

Besides being an approximation to DFT, SCC-DFTB can be viewed as an extension of a tight-binding method, which includes charge self-consistency and is parametrized using DFT. The energy in tight-binding (TB) methods is composed of two parts, electronic and repulsive. The electronic part is described by a Hamiltonian, which is usually represented in a minimal basis of atom-centered basis functions. In DFTB, this Hamilton matrix is derived from DFT, using as the reference density the superposition of neutral atomic densities and a minimal basis of atomic wave functions, which is explicitly calculated.[2−4] The repulsive energy, which consists of the DFT double counting contributions and the core−core repulsion, can be approximated as a sum of atomic pair repulsion functions. The relation of DFT and TB methods has been discussed in detail by Foulkes and Haydock.[5] Standard tight-binding methods are usually based on the Harris functional approach,[6] that is, they diagonalize a suitable Hamiltonian once and use this non-self-consistent solution to derive further properties like forces, second derivatives, and so forth. For organic and biological molecules, however, an approximate inclusion of self-consistency has been found to be crucial. This has been done by a second-order expansion of the total DFT energy[1] followed by further approximations to derive a computationally efficient model. However, to account accurately for some molecular properties (e.g., proton affinities, hydrogen bonding), higher-order terms have to be included in the formalism,[7−9] and special care must be taken for a proper treatment of Coulomb interactions.

SCC-DFTB is parametrized using a generalized gradient approximation (GGA) functional. In the actual version, the electronic parameters are calculated using the PBE functional.[10] This means, however, that the well-known DFT-GGA deficiencies are inherited by SCC-DFTB. Of particular relevance is the DFT-GGA tendency to overpolarize extended π-conjugate systems,[11] the problems of ionic and charge-transfer excited states,[12] and the missing dispersion interactions, which have been included by augmenting SCC-DFTB using an empirical exten-

[†] Part of the "Walter Thiel Festschrift".

* To whom correspondence should be addressed. E-mail: hwitek@mail.nctu.edu.tw (H.W.); m.elstner@tu-bs.de (M.E.).

[§] Technische Universität Braunschweig.

[‡] National Chiao Tung University.

10.1021/jp902973m CCC: $40.75
Published on Web 09/24/2009

sion.[13] The performance and deficiencies of SCC-DFTB with respect to biological applications have been reviewed recently,[14,15] and methodological developments have been described in ref 16.

SCC-DFTB has been tested for various properties of small organic molecules like heats of formations, geometries, vibrational frequencies, dipole moments, and so forth in several recent publications. It should be noted that all of these test sets contain a large number of molecules representative of many chemical bonding situations. However, good performance for small molecules does not guarantee a good description of larger molecules. Good examples are the structures and relative energies of peptides, which pose significant problems for semiempirical models like AM1[17] and PM3[18] but are well described at the SCC-DFTB level[19,20] or with more elaborate NDDO methods like OM1[21] or OM2.[22,23] In general, SCC-DFTB is excellent in reproducing geometries. Also, reaction energies are reproduced reasonably well on average,[1,24] while heats of formation are overestimated, owing to the overbinding tendency of SCC-DFTB. Recently, the SCC-DFTB heats of formation have been systematically tested. It turns out that reparametrization of atomic contributions can improve the performance for heats of formation significantly; however, the refined NDDO methods like OM2 or PDDG-PM3[25] are still superior to SCC-DFTB in this respect.[26,27] The performance of SCC-DFTB for vibrational frequencies,[28−31] although reasonable on average, is less satisfactory than that for geometries. However, also vibrational frequencies could be improved significantly after reparametrization.[32]

These studies showed that the SCC-DFTB performance for distinct properties can be partially tweaked by more or less elaborate fitting procedures. A question arises whether the DFTB performance can be systematically improved for all of the considered properties using better fitting strategies for the repulsive potentials or whether there are optimization conflicts, in which one property is improved at the cost of others. In general, such a behavior is expected since the electronic part constitutes an approximation to full DFT, and its limited flexibility (minimal basis, charge self-consistency, fixed reference density, etc.) may lead to a limited transferability and conflicts in parameter optimization. In this work, we aim to explore this point in detail, using a new representation of repulsive potentials and an improved fitting strategy. The proposed methodology is applied to construct a set of repulsive potentials for C and H, which are subsequently tested in a systematic fashion using large sets of molecules.

Theory

The first subsection briefly introduces the SCC-DFTB method in a degree necessary for understanding the derivation and functioning of the algorithm for automatized construction of the repulsive SCC-DFTB potentials V^{rep}. A thorough discussion of the SCC-DFTB method was given elsewhere.[1,2] The following subsections describe the details of the proposed algorithm. Our procedure uses as an input a set of molecular equilibrium geometries together with the corresponding atomization energies. Because both the energies and forces associated with the employed molecular structures can be represented as linear functions of repulsive potentials, a solution of an inverse problem is capable, in principle, of yielding a set of V^{rep} that reproduces the input quantities.

SCC-DFTB. The self-consistent charge density functional tight-binding (SCC-DFTB) method[1] is an approximation to density functional theory (DFT). It originates from a second-order Taylor expansion of the total DFT energy

$$E[\rho] = \sum_i^{occ} \langle \psi_i | \hat{H}^0 | \psi_i \rangle - \frac{1}{2} \int\int \frac{\rho^0 \rho'^0}{|r - r'|} dr dr' - \int V^{xc}[\rho^0]\rho^0 dr + E^{xc}[\rho^0] + E^{core} + \frac{1}{2} \int\int \left[\frac{1}{|r - r'|} + \left(\frac{\delta^2 E^{xc}[\rho]}{\delta\rho\delta\rho'} \right)_{\rho=\rho^0} \right] \delta\rho\delta\rho' dr dr' \quad (1)$$

with respect to density fluctuations around some reference density ρ^0, where ρ^0 is usually chosen as a sum of neutral or confined atomic densities ρ_A^0. The last term in eq 1 describes the contribution to energy due to the density fluctuation $\delta\rho = \rho - \rho^0$. We approximate the density fluctuation as a sum of induced atomic charges, $\delta\rho = \sum_A \Delta q_A$, where $\Delta q_A = q_A - q_A^0$ and q_A are computed as atomic Mulliken charges. The last term in eq 1 can thus be rewritten as

$$E^{2nd} = \frac{1}{2} \sum_{AB} \Delta q_A \Delta q_B \gamma_{AB} \quad (2)$$

where the distance-dependent function γ_{AB} can be interpreted as a pure Coulomb interaction between the induced charges Δq_A and Δq_B at long distances and a Hubbard-type correlation at short distances. An explicit form of γ_{AB} can be found elsewhere.[1,33]

The energy contributions in the second line in eq 1 depend on the neutral atomic densities and interatomic distances. As usual in tight-binding theory, these terms are collected in a repulsive energy E^{rep} contribution.[5] This term can be approximated as a sum of short-range pair potentials

$$E^{rep} = \sum_{A>B} V_{AB}^{rep}(r_{AB}) \quad (3)$$

The repulsive potentials are determined by an appropriate fitting scheme, where a set of accurate energies and geometries, either theoretical or experimental, is used. In the following subsections, we propose a systematic and automatized method for constructing such potentials.

The first term of eq 1 involves a summation over a set of molecular Kohn−Sham orbitals ψ_i. The molecular orbitals (MOs) are built as linear combinations (LC) of atomic valence orbitals (AOs), $\psi_i = \sum_\nu c_{\nu i}\phi_\nu$. Note that the SCC-DFTB method employs only a minimal valence basis set for each atom. The valence AOs are determined by solving a confined radial pseudoatomic Kohn−Sham problem

$$[\hat{T} + \hat{V}^{eff} + \hat{V}^{conf}]\phi_i(r) = \varepsilon_i \phi_i(r) \quad (4)$$

where the explicit form of the kinetic $\hat{T}$ and potential $\hat{V}$ energy operators depend on the choice of either a one-component Schrödinger or a four-component Dirac equation. The additional confining potential $\hat{V}^{conf}$ is employed to compress the otherwise very diffused tail of valence orbitals. Note that in traditional quantum chemical calculations, very large basis sets are usually employed that give enough variational freedom to reoptimize the LC coefficients of valence AOs inside of a molecule or a solid. In the SCC-DFTB model, the limitation introduced by the choice of a minimal basis set forces us to mimic the presence of other atoms already at the atomic level via the additional confining potential. It is well-known that such a confinement improves the bonding description in cases where the minimal

basis is used.[34] At present, two explicit forms of the confining potentials are used, harmonic[1] and Woods−Saxon.[4]

In the atomic orbital basis, the overlap and Hamilton matrix elements are expressed as

$$S_{\mu\nu} = \langle\phi_\mu|\phi_\nu\rangle \qquad \text{and}$$

$$H^0_{\mu\nu} = \begin{cases} \varepsilon_\mu^{\text{neutral unconfined atom}} & \text{if } \mu = \nu \\ \langle\phi_\mu|\hat{H}(\rho_A^0 + \rho_B^0)|\phi_\nu\rangle & \text{if } A \neq B \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where A and B denote the atoms on which the atomic orbitals ϕ_μ and ϕ_ν are centered. Both overlap and Hamilton matrix elements are calculated and tabulated for a dense mesh of interatomic distances.

With the use of Mulliken charges, q_A can be expressed as

$$q_A = \frac{1}{2}\sum_i^{\text{occ}} n_i \sum_{\mu\in A}\sum_\nu (c^*_{\mu i}c_{\nu i}S_{\mu\nu} + c^*_{\nu i}c_{\mu i}S_{\mu\nu}) \tag{6}$$

Now, the generalized eigenvalue problem can be solved iteratively for the charge self-consistent Hamiltonian

$$\hat{H}_{\mu\nu} = \hat{H}^0_{\mu\nu} + \frac{1}{2}S_{\mu\nu}\sum_C \Delta q_C(\gamma_{AC} + \gamma_{BC}) \tag{7}$$

Thus, the final expression for the total SCC-DFTB energy reads

$$E^{\text{tot}} = \sum_{i\mu\nu} c_{\mu i}c_{\nu i}H^0_{\mu\nu} + \frac{1}{2}\sum_{AB}\Delta q_A\Delta q_B\gamma_{AB} + \sum_{A>B} V^{\text{rep}}_{AB}(r_{AB}) \tag{8}$$

Analytical Representation of Repulsive Potentials. In the actual version of SCC-DFTB, the repulsive potential V^{rep}_{AB} between atoms A and B is represented by cubic splines. Here, we choose to represent the repulsive potentials as a collection of fourth-order splines. This choice ensures that the first- and second-order derivatives of the repulsive potentials—necessary for geometry optimization and for computing harmonic vibrational frequencies—will be given as sufficiently smooth functions of interatomic distance. The previous choice of cubic splines yielded piecewise linear functions for the second-order derivatives that did not have enough flexibility to reproduce adequately the repulsive portion of the Hessian. As we see below, where the exact construction of our spline functions is given in more detail, using fourth-order splines does not introduce any additional complexity in comparison to the traditional third-order approach.

The definition of the repulsive potential V^{rep}_{AB} between atoms A and B starts with dividing the range of possible atomic separations $[0,\infty)$ into a set of intervals $I_0 = [0, r_1)$, $I_1 = [r_1, r_2)$, ..., $I_n = [r_n, r_{n+1})$, and $I_{n+1} = [r_{n+1},\infty)$ using a set of division points $r_1, r_2, ..., r_{n+1}$. Subsequently, for each of the intervals I_1, I_2, ..., I_n, we define a fourth-order polynomial given by

$$S_i(r) = \sum_{k=0}^{4} a_{k,i}(r - r_i)^k \tag{9}$$

where the five unknown parameters $a_{k,i}$ are to be determined. The total number of unknowns is thus $5n$. We require that S_{n+1} as well as its first three derivatives are identically equal to zero over the whole interval I_{n+1}. Therefore, the last division point, r_{n+1}, can be interpreted as a cutoff, beyond which the repulsive potential vanishes. To conclude this part of our presentation, we mention that S_0 is defined as

$$S_0(r) = \alpha\exp(\beta r + \gamma) \tag{10}$$

where the three parameters α, β, and γ are chosen to match the corresponding values of S_1 and its two lowest derivatives at $r = r_1$.

The number of division points $r_1, r_2, ..., r_{n+1}$ should be rather small. For pairs of elements not showing large diversity of bonding mechanisms (e.g., for $V^{\text{rep}}_{\text{CH}}$ or $V^{\text{rep}}_{\text{HH}}$), a choice of two to three division points seems to be sufficient. For more complicated cases (e.g., for $V^{\text{rep}}_{\text{CC}}$), a few additional division points may give a better description of the potential. We have found—for the reasons that we discuss later—that using a large number of division points may lead to unphysical shapes of the reproduced repulsive potentials.

Continuity Equations. In the following, we are referring to one specific repulsive potential V^{rep}_{XY}, X and Y being two different types of atoms. It is possible to eliminate $4n$ (out of $5n$) unknown parameters $a_{k,i}$ by imposing the spline continuity conditions given by

$$S_i(r_{i+1}) = S_{i+1}(r_{i+1}) \tag{11}$$

$$S'_i(r_{i+1}) = S'_{i+1}(r_{i+1}) \tag{12}$$

$$S''_i(r_{i+1}) = S''_{i+1}(r_{i+1}) \tag{13}$$

$$S'''_i(r_{i+1}) = S'''_{i+1}(r_{i+1}) \tag{14}$$

where the primes denote subsequent derivatives and $i = 1, 2, ..., n$.

The remaining n unknown parameters defining each of the repulsive potentials are fitted to reproduce a set of atomization energies and equilibrium structures for a chosen group of molecules. The atomization energies and equilibrium structures can be obtained either directly from experiment or from some accurate quantum chemical method. In the next two sections, we show how the SCC-DFTB atomization energies and equilibrium geometries can be used to obtain further conditions. While the continuity equations need to be exactly fullfilled, all other equations give an either over- or underdetermined system of linear equations. In the Appendix, we derive the system of linear equations arising from all mentioned conditions and show how it can be solved using the singular value decomposition.

Energy Equations. The total SCC-DFTB energy of a molecule M can be divided into two parts, the electronic part E^{el} given by the first two terms of eq 8 and the repulsive part E^{rep} given by the third term of eq 8

$$E^{\text{SCC-DFTB}} = E^{\text{el}} + E^{\text{rep}} \tag{15}$$

The atomization energy E^{at} of the molecule M is given by

$$E^{\text{at}} = E^{\text{SCC-DFTB}} - \sum_A E^{\text{el}}_A \tag{16}$$

where the summation runs over all atoms constituting the molecule M. A simple rearrangement of these equations gives

$$\sum_{A>B} V_{AB}^{\mathrm{rep}}(r_{AB}) = E^{\mathrm{at}} - E^{\mathrm{el}} + \sum_{A} E_A^{\mathrm{el}} \tag{17}$$

Note that the right-hand side (RHS) of eq 17 can be treated as a constant because it does not depend on the sought spline coefficients. The values of atomic electronic energies E_A^{el} on the RHS of eq 17 are usually computed as

$$E_A^{\mathrm{el}} = \sum_{i}^{\mathrm{occ}} n_i \varepsilon_i^A + E_A^{\mathrm{spin}} \tag{18}$$

where E_A^{spin} denotes the atomic spin polarization energy, n_i the occupation number and ε_i the atomic Kohn−Sham eigenvalues. However, it is possible to replace E_A^{el} with empirical atomic electronic energies, whose actual values will be determined via the fitting process described below. With this ansatz, eq 17 can be written as

$$\sum_{A>B} V_{AB}^{\mathrm{rep}}(r_{AB}) - \sum_{A} E_A^{\mathrm{el}} = E^{\mathrm{at}} - E^{\mathrm{el}} \tag{19}$$

where the quantities on the LHS depend on the unknown parameters and the quantities on RHS are constants. Both fitting strategies, that is, with using the atomic electronic energies given by eq 17 and with using optimized atomic electronic energies, will be employed in the remainder of this paper.

In the Appendix, a detailed derivation of the working energy equation is provided.

Force Equations. The force vector can be decomposed in a similar way as the SCC-DFTB total energy as

$$\boldsymbol{F}^{\mathrm{SCC\text{-}DFTB}} = \boldsymbol{F}^{\mathrm{el}} + \boldsymbol{F}^{\mathrm{rep}} \tag{20}$$

At the equilibrium geometry, this vector is identically equal to zero, and we can write

$$\boldsymbol{F}^{\mathrm{rep}} = -\boldsymbol{F}^{\mathrm{el}} \tag{21}$$

The repulsive part of the force vector on atom A can be expressed through the first derivatives of the repulsive potentials as

$$\boldsymbol{F}_A^{\mathrm{rep}} = -\sum_{B\neq A} \frac{\mathbf{r}_A - \mathbf{r}_B}{\mathbf{r}_{AB}} \frac{\mathrm{d}V_{AB}^{\mathrm{rep}}}{\mathrm{d}r}\bigg|_{r=r_{AB}} \tag{22}$$

In principle, it is also possible to employ nonequilibrium structures in our fitting procedure. The only complication is that for such a nonequilibrium structure, the total force vector in eq 20 is not equal to zero but equal to a reference total force vector which must be calculated by some quantum chemical method.

In the Appendix, the force working equations are derived in detail.

Additional Equations. It is sometimes convenient to impose some special conditions on the fitted repulsive potentials. We have found that the requirement that the repulsive potential V^{rep} has a particular value (say V) at some particular point (say r_z) can be easily expressed in a similar fashion as the previously discussed energy and force equations. Let us assume that the point r_z lies inside of the interval I_{p}, with r_{p} being the corresponding division point. Then, we can write

$$V^{\mathrm{rep}}(r_z) = V = S_{\mathrm{p}}(r_z) = \sum_{k=0}^{4} a_{k,\mathrm{p}}(r_z - r_{\mathrm{p}})^k \tag{23}$$

It is easy to see that similar additional conditions can be imposed not only on repulsive potentials but also on their derivatives. Note that particularly useful in practice are the additional conditions imposed on the second-order derivatives of the repulsive potentials at some particular points because they help to adjust the values of harmonic vibrational frequencies of a given type of vibration (e.g., the C−H stretch).

Reaction Equations. A simple modification of the presented energy equations enables fitting the repulsive potentials also to energies of chemical reactions. Let us consider for example a simple reaction between molecules M_1 and M_2 given as

$$n_1 M_1 + n_2 M_2 \rightarrow n_3 M_3 + n_4 M_4 \tag{24}$$

We would like to determine the shape of the repulsive SCC-DFTB potentials that yields the correct value of the reaction energy E^{rea} for this reaction. For this purpose, we construct four energy equations for molecules M_1, M_2, M_3, and M_4 as described in eq 19. Adding these equations with the appropriate reaction coefficients yields a working equation for reaction energies. Details can be found in the Appendix.

Results

In this section, we describe in detail the fitting process used to find optimal repulsive potentials. In the first step, the electronic parameters are defined. Second, the fit and test sets are specified, and finally, the additional parameters of the repulsive potentials are determined.

The Electronic Parameters. We discuss the construction of repulsive potentials with an electronic part constructed with the following form of the harmonic confining potential

$$\hat{V}^{\mathrm{conf}} = \left(\frac{r}{r_0}\right)^2 \tag{25}$$

Two different confining radii are used for each atom, r_0^{w} and r_0^{d}, corresponding to the confinement of atomic orbitals and atomic densities, respectively. A good choice for r_0^{w} is found to be twice the covalent radius of the respective element; this value can be further optimized.[2] As discussed in refs 15 and 5, the determination of r_0^{w} can be compared with the optimization of a basis set, whereas r_0^{d} is an empirical parameter. In the present paper, the following initial confining radii are chosen: $r_0^{\mathrm{w}}(\mathrm{H}) = 3.0$, $r_0^{\mathrm{d}}(\mathrm{H}) = 2.5$, $r_0^{\mathrm{w}}(\mathrm{C}) = 2.7$, and $r_0^{\mathrm{d}}(\mathrm{C}) = 7.0$. These values of confining radii for hydrogen and carbon will be further optimized in the present study. As mentioned in eq 18, the electronic energy of an atom A is defined as

$$E_A^{\mathrm{el}} = \sum_{i}^{\mathrm{occ}} n_i \varepsilon_i^A + E_A^{\mathrm{spin}} \tag{26}$$

TABLE 1: Spin Polarization Energies E_A^{spin}, Atomic Orbital Energies ε_p^A and ε_s^A, and Total Atomic Electronic Energies E_A^{el} (in Hartree) Obtained from Atomic PBE Calculations

element	E_A^{spin}	ε_p^A	ε_s^A	E_A^{el}
H	−0.0411		−0.2386	−0.2797
C	−0.0455	−0.1944	−0.5049	−1.4440

TABLE 2: Summary of Parameter Sets Developed in This Study

	original confining radii	optimized confining radii	frequency optimized confining radii
approach 1	opt1[a]	opt3[b]	opt5[c]
approach 2	opt2[a]	opt4[d]	opt6[d]

[a] Confining radii $r_0^w(C) = 2.7$, $r_0^d(C) = 7.0$, $r_0^w(H) = 3.0$, $r_0^d(H) = 2.5$. [b] Confining radii $r_0^w(C) = 2.7$, $r_0^d(C) = 5.0$, $r_0^w(H) = 3.3$, $r_0^d(H) = 3.0$. [c] Confining radii $r_0^w(C) = 3.0$, $r_0^d(C) = 7.0$, $r_0^w(H) = 2.7$, $r_0^d(H) = 2.5$. [d] Confining radii $r_0^w(C) = 3.0$, $r_0^d(C) = 7.0$, $r_0^w(H) = 2.5$, $r_0^d(H) = 2.5$.

TABLE 3: Optimized Atomic Electronic Energies (in Hartree)

	opt1	opt3	opt5
E_H^{el}	−0.2555	−0.2514	−0.2524
E_C^{el}	−1.4527	−1.4176	−1.4472

The atomic Kohn−Sham eigenvalues ε_i are calculated within the DFT framework using the PBE functional[10] and uncompressed atomic orbitals and densities ($r_0^w \rightarrow \infty$ and $r_0^d \rightarrow \infty$). The last term E_A^{spin} describes the atomic spin polarization energy, which can easily be calculated from DFT.[35] The actual values of ε_i^A, E_A^{spin}, and E_A^{el} for H and C are given in Table 1.

Note that the original fitting procedure[1] used the LDA spin polarization energies along with the repulsive potentials determined by taking the difference of the reference energy and the SCC-DFTB electronic energy

$$E^{rep}(R) = \{E^{ref}(R) - E^{el}\}_{\text{reference system}} \quad (27)$$

for a predefined set of reference molecules. The LDA spin polarization energies for hydrogen and carbon are $E_H^{spin} = -0.0330$ and $E_C^{spin} = -0.0439$.

In this work, we use the fitting procedure as described in the previous section and apply it using the following two approaches. In approach 1, the atomic electronic energies E_A^{el} are treated as free parameters and are optimized together with the parameters of the repulsive potentials. In approach 2, the calculated atomic electronic energies from Table 1 are used, and only the parameters of the repulsive potentials are optimized. Therefore, in approach 2, the atomic energies are not optimized.

In total, we optimize six different sets of repulsive potentials, three using approach 1 and three using approach 2. Within each approach, one set of repulsive potentials is determined for the original confining radii, one for optimized confining radii and one for confining radii optimized, to give optimal performance for the vibrational frequencies. The considered parameter sets are listed in Table 2. The optimal values of the atomic electronic energies E_A^{el} for the three sets of repulsive potentials, opt1, opt3, and opt5, are given in Table 3.

Training Sets. The actual fitting procedure of E^{rep} is done in two consecutive steps. In the first step, a training set 1 is used to find the parameters defining each E^{rep}. The training set 1 contains experimental structures of hydrogen, methane, ethyne, ethene, and ethane (for references, see Supporting Information). These molecules are chosen because they represent the most important bonding situations occurring in molecules consisting of carbon and hydrogen. Unfortunately, the repulsive potentials obtained using such a training set produce large errors for the equilibrium geometries of cyclopropene and 2-butyne in the case of approach 1. To remedy this problem, the experimental equilibrium geometries of these two molecules have been included in the training set 1 for opt1, opt3, and opt5. The reference experimental atomization energies are taken from the CCCBDB database.[36] The vibrational zero-point energies have to be excluded from these values because SCC-DFTB is not parametrized to enthalpies at 0 K or to heats of formation at 298 K but to energies excluding the vibrational and thermal contribution. The zero-point energy contribution has been approximated as a half of the sum of the experimental fundamental frequencies. For hydrogen, methane, ethyne, ethene, and ethane, these energies are 109.6, 419.5, 405.1, 562.7, and 711.4 kcal/mol, respectively. For cyclopropene and 2-butyne, only the force equations are used in the parametrization process.

In the second step, the performance of the fit is evaluated using a training set 2, which contains atomization energies, equilibrium geometries, and vibrational frequencies of 15 molecules together with 32 selected reaction energies. Details are given in the Supporting Information. The reason for employing a second, larger training set is the following. The division points and additional equations for each repulsive potential have to be specified as an initial step of our fitting procedure. This is a nontrivial problem and will be discussed in more detail below. Dependent on the performance of the fit for the training set 2, the division points are changed, either moved, deleted, or added. Then, the repulsive potentials are refitted according to the first step and again tested using the training set 2. This procedure is iterated until a satisfactory performance is found.

Parameters Defining the Repulsive Potentials. In principle, the division points r_p and the cutoff radii r_{n+1} defining the spline functions of the repulsive H−H, C−H, and C−C pair potentials can be freely chosen. Clearly, the shortest bond length between atom types X and Y should lie in the first interval I_1^{XY}, that is, it should be larger than the first division point r_1^{XY}.

As discussed in detail in ref 2, a cutoff radius is introduced for the repulsive potential beyond which the potential and its derivatives are zero. The cutoff radius in the standard parametrization of SCC-DFTB was chosen to be smaller than second neighbor distances, resulting in cutoff radii of 4.3, 3.5, and 2.64 atomic units (au) for C−C, C−H, and H−H, respectively. Here, the same cutoff radii are used, except for C−C, where a cutoff of 4.8 au is applied. This is larger than second neighbor distances in some molecules (e.g., benzene, propene, propane, cyclohexane); however, the potential decays rapidly, being negligible at the second neighbor distance.

The number of intervals is another free parameter which is related to the number of fitting objectives. For example, in the parametrization of the H−H potential, we have only three objectives, the atomization energy, the equilibrium bond length, and the stretch frequency of the hydrogen molecule, which can directly be connected to the repulsive potential and its first and second derivatives. As described in the methods section, for each interval, there is only one free parameter to be determined; all others are defined by the continuity equations. Thus, three intervals for the H−H potential are necessary to fulfill the three objectives. The same holds for the C−H potential. The different C−H bonds of the molecules in training set 1 are similar such

that we have again only three objectives as above, and only three intervals are necessary for the repulsive potential. For C−C, the situation is different since the single, double, and triple bonds have substantially different characteristics. In principle, one could fit the zeroth, first, and second derivatives to reference data for ethane, ethene, and ethyne using nine intervals (three intervals for each type of bond as above). However, for systems with intermediate bond lengths (e.g., benzene), such a fit results in large errors. This is due to the fact that the slope and curvature of the intervals of the single, double, and triple bonds do not tend to match at the division points. A pragmatic solution to this problem is to use intervals covering more than one bond type, which leads to an interpolation between these two regions. This of course means that choosing the optimal interval division for the C−C potential is an empirical procedure of trial and error. As a result, an extensive scan for different numbers and position of intervals had to be performed because the properties of training set 2 depend sensitively on the choice of the intervals. The division points found this way define the intervals and are listed for each parameter set in the Supporting Information. (Another possible way to avoid this empirical search protocol for suitable intervals could be using equidistant intervals as a first fit. In a second step, the resulting spline function is interpolated to a single polynomial. Preliminary tests show that the artifacts in the first and second derivatives are eliminated by this technique.)

For the fitting approach 1 (i.e., including the fit of atomic electronic energies), the situation becomes even more complicated since two more parameters, E_C^{el} and E_H^{el}, have to be determined. Both parameters appear in all energy equations (except for hydrogen, where only E_H^{el} appears) and thus affect all repulsive potentials. That means, for example, that for opt1, we use only two intervals for the C−H potential in order to leave one degree of freedom for the determination of the atomic electronic energy. We have tested in a "brute force" manner several numbers and positions of the intervals; details can be found in the Supporting Information. Note that the search for the best possible intervals depends on the choice of additional equations (see below) if good performance of vibrational frequencies is also desired.

So far, only fitting schemes using equilibrium geometries and atomization energies have been discussed. In order to achieve a good performance also for vibrational frequencies, additional equations have to be included in the fitting process. We choose the form $V''^{rep}_{XY}(r_z) = V$, where the second derivative of a pair potential $X-Y$ at the distance r_z is set to a certain value V. For H_2, r_z is set to the equilibrium distance, and V is chosen appropriately in order to reproduce the experimental fundamental frequency of H_2 with good accuracy. Note that, in general, the additional predefined conditions are not exactly fulfilled in the optimization scheme due to the limitations resulting from the applied least-squares fit.

As a first guess for V, one can take the value of the second derivative of the H−H potential at a given point, which is then varied by hand. Similarly, for C−H, V is chosen to minimize the error of the C−H stretch frequencies in the training set 2. In the more complicated case of the C−C potential, we have considered for simplicity only the C−C stretch modes of ethane, ethene, and ethyne. Unfortunately, inclusion of additional equations improves the C−C stretch frequencies at the cost of geometries and atomization energies, which shows the limits of optimization within the SCC-DFTB framework. For a further improvement, modifications of the electronic part seem to be necessary. By choosing a set of additional equations, the ratio of errors for frequencies and for geometries or atomization energies can be controlled. For future work, an appropriately weighted objective function combined with an optimization algorithm could automatically find a desired ratio. Such algorithms for similar problems have recently been reported, for example, in refs 25 and 37.

TABLE 4: Additional Equations of the Form $V''^{rep}_{XY}(r_z) = V$

$X-Y$	r_z	V(opt1)	V(opt2)	V(opt3)	V(opt4)	V(opt5)	V(opt6)
H−H	0.743	0.413	0.420	0.416	0.415	0.4075	0.415
C−H	1.081	0.400	0.400	0.400	0.380	0.400	0.380
C−C	1.203						1.300
C−C	1.300	0.950		0.980		0.950	
C−C	1.330						0.800
C−C	1.505	0.340		0.320		0.310	0.370

TABLE 5: Experimental Fundamental CC Stretch Frequencies and Deviation of Different Parameter Sets (in cm^{-1})

molecule	exp[a]	opt1	opt2	opt3	opt4	opt5	opt6	mio[b]	M2005[c]
ethyne	1974	+13	+50	−30	+33	+62	−35	+146	+3
ethene	1623	+138	+191	+132	+142	+129	+32	+200	+67
ethane	995	+62	+43	+16	+30	+15	+29	+137	−26

[a] Experimental values from ref 32. [b] Parameter set from ref 1. [c] Parameter set from ref 32.

Neglecting the additional equations in the fitting scheme leads to substantial errors in the vibrational frequencies as large as 2000 (H_2) and 500 cm^{-1} (C−H stretch in hydrocarbons). For all of the parameter sets shown in Table 2, one additional equation for the H−H potential and one for the C−H potential is needed. Due to limits of optimization, we decided to create additional sets of repulsive potentials (opt5 and opt6), which show improved performance for vibrational frequencies. All additional equations are listed in Table 4.

Determination of E^{rep} Using the Original Confining Radii. In the first step, we keep the original electronic confining radii and optimize only the repulsive potentials. Set opt1 is created using fitting approach 1, that is, fitting also atomic electronic energies. For set opt2, we apply fitting approach 2 using the calculated atomic electronic energies from Table 1. Interestingly, set opt1 shows large errors in the C−C stretch frequencies if constructed without additional equations for C−C. Including two additional equations for C−C reduces these errors significantly. For set opt2, the errors in the frequencies are reasonably small already without additional equations and even lower in comparison to the original mio parameter set (see Table 5). Including additional equations for the C−C potential reduces errors for frequencies but worsens geometries and atomization energies; therefore, they were not applied.

Table 6 summarizes the performance of opt1 and opt2 in comparison to the original mio parameter set and with the parameter set M2005 optimized by Małolepsza et al. for vibrational frequencies.[32] The main advantage of the new sets of repulsive potentials is a significant improvement in the computed atomization energies; the mean absolute error for 14 atomization energies is reduced from 36.5 (mio) to 4.1 and 4.0 kcal/mol (opt1 and opt2, respectively). It is worth mentioning here that in the original mio parameter set, the atomization energies were consistently overestimated by roughly 5% and that the error of 36.5 kcal/mol could be reduced to only 8.8 kcal/mol just by employing the spin polarization energies derived from PBE (instead of LDA). Since the LDA values are consistently larger than the PBE ones, it can be concluded that the mio parameters lead to a consistent overbinding for all

TABLE 6: Mean and Maximum Absolute Deviation of Several Properties of the Training Set 2

property[a]	N^b	opt1	opt2	opt3	opt4	opt5	opt6	mio[c]	M2005[c,d]
E^{at} (kcal/mol)	14	4.1	4.0	1.4	4.0	8.3	5.0	36.5	40.6
E^{at}_{max} (kcal/mol)		14.6	19.1	4.0	16.2	27.8	23.3	87.1	88.3
r (Å)	41	0.007	0.006	0.008	0.006	0.007	0.014	0.010	0.016
r_{max} (Å)		0.023	0.030	0.025	0.029	0.022	0.041	0.026	0.040
a (deg)	25	0.6	0.6	0.6	0.5	0.5	0.7	0.5	1.2
a_{max}		1.8	1.7	1.8	1.9	2.0	2.1	1.7	2.7
ν (cm $^{-1}$)	305	42	46	48	39	38	32	61	31
ν_{max} (cm $^{-1}$)		161	229	250	185	175	163	352	123
E^{rea} (kcal/mol)	32	5.1	4.2	2.4	4.7	6.6	7.2	5.2	8.3
E^{rea}_{max} (kcal/mol)		14.8	14.3	6.9	15.5	32.4	23.9	14.9	46.4

[a] Atomization energies E^{at}, bond lengths r, bond angles a, harmonic vibrational wavenumbers ν, and reaction energies E^{rea}; max stands for maximum absolute deviation. The reaction energy is compared to G3B3[38] results at 0 K. The zero-point energies are calculated using harmonic vibrational frequencies calculated analytically with SCC-DFTB.[39] All other properties are compared to expermiental values. For details, see Supporting Information. [b] Number of comparisons. [c] For the calculation of E^{at}, the originally used LDA spin polarization energies were taken; for PBE spin polarization energies, the MAD for mio is 8.8 and the maximal absolute deviation is 30.5 kcal/mol, and that for M2005 is 72.4 and 155.0 kcal/mol. [d] Hydrogen was excluded since these parameters were not optimized for this molecule.

molecules. It is interesting to mention that this error can be further reduced to 3.8 kcal/mol if the original electronic atomic energies defined by eq 18 are replaced by empirical values obtained by fitting the total SCC-DFTB atomization energies to the experimental atomization energies of the training set 2. Geometries are well reproduced by all sets, while vibrational frequencies are more accurate with opt1 and opt2 in comparison to the original mio set but do not achieve the accuracy of M2005. For reaction energies, no major improvement can be achieved, showing again that the original mio fit leads to a set of quite consistent repulsive potentials, however, including a large overbinding per bond.

Figure 1 shows the shape of repulsive potentials for mio, opt1, and opt2. Set opt1 is more repulsive than opt2 for the C−H and H−H potentials, whereas for the C−C potential, this tendency is reversed. One main difference between opt2 and mio arises from the different spin polarization energies used in the fitting (PBE calculated values versus LDA ones). This is reflected in the relatively small shift of the curves between mio and opt2. The eminent shift of the repulsive potentials C−H and H−H of opt1 shows that the optimization of the atomic parameters E^{el}_A leads to a difference in the potential depth. This difference is effectively incorporated into the optimized atomic electronic energies. Our tests so far indicate that this shift can be different for different atom pairs. This can become problematic since the effective binding energy is no longer represented by the "depth" of the total energy (electronic plus repulsive) but is transferred to the atomic values. When comparing the opt1 binding energy curves with those from some high-level reference quantum calculation, the opt1 potentials can be more shallow; the difference is captured by the atomic values. This may lead to problems when new atom types are included without refitting the parameters of the old atom types, that is, consistent fits can only be achieved when all desired atom types are fitted at once.

This fact can be illustrated with the example reaction $C_2H_6 + H_2O \rightarrow CH_3OH + CH_4$. The reaction enthalpy at 0 K of the mio set is 12.1 kcal/mol and reproduces the G3B3[38] result (11.7 kcal/mol). Replacement of the C−C, C−H, and H−H repulsive potentials with the opt1 potentials gives 24.7 kcal/mol, and it is 5.4 kcal/mol for opt2. Set opt1 yields a large error, enforcing the assumption that the opt1 potentials C−C and C−H are not transferable to the mio potentials C−O and O−H. In opt2, the difference to the mio result is much smaller and mainly due to the shift of the C−C repulsive potential (see Figure 1). While the mio parameter set overbinds C−C and C−O, this overbinding is reduced for C−C in opt2. We expect similar reaction enthalpies for mio and opt2 when also reducing the C−O overbinding with the presented fitting procedure. Note that for reactions, the atomic electronic energies cancel out. (Another example reaction is $C_2H_6 + NH_3 \rightarrow CH_3NH_2 + CH_4$, where the same arguments hold. The mio parameter set gives a reaction enthalpy at 0 K of 8.1 kcal/mol, very similar to the G3B3 result of 7.9 kcal/mol. Using the C−C, C−H, and H−H potentials of opt1 and all further pair potentials from the mio set yields 20.7 kcal/mol; for opt2, it is 1.4 kcal/mol.)

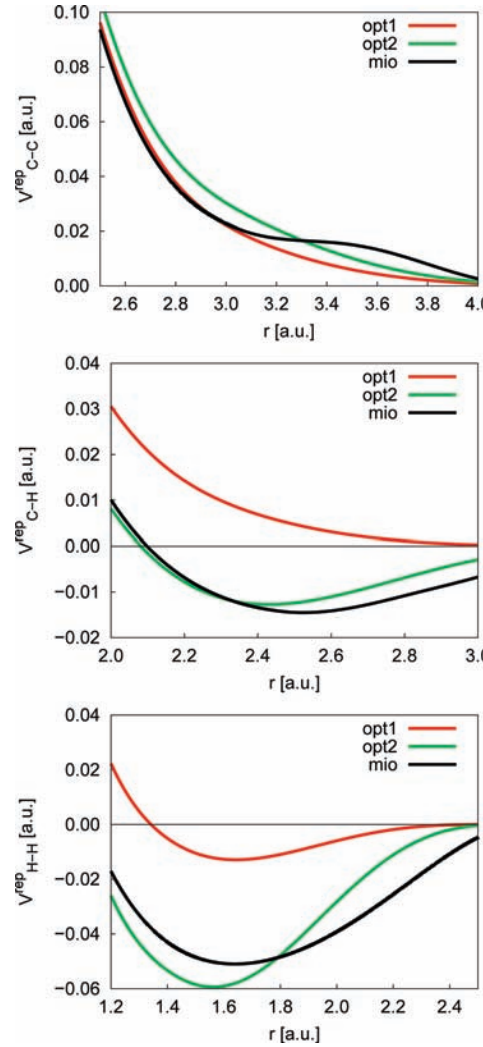

Figure 1. Original and optimized repulsive potentials.

The mio repulsive potential corresponding to the CC pair displays a characteristic hump at around 3.6 au. The origin of this hump is closely related to the previous paradigm of determination of the repulsive potentials. The short-range part of the CC mio was obtained from auxiliary DFT calculations for a set of small molecules (C_2H_2, C_2H_4, and C_2H_6). This short-range potential was subsequently down-shifted to ensure correct

energetics of hydrocarbons. On the other hand, the long-range part of the potential was required to vanish beyond 4.3 au. Combining the two segments of the mio CC potential into a single curve resulted in producing the unphysical hump mentioned earlier. Clearly, this superfluous behavior of the potential is one of the drawbacks of constructing the repulsive potentials in a semimanual manner. As can be seen from the curves presented in Figure 1, this behavior can be totally eliminated while using the automatized approach.

The fact that the H−H repulsive potential becomes negative (i.e., attractive) may lead in practice to problems in molecular dynamics simulations when the interatomic distance between two nonbonded hydrogen atoms becomes too small. This potential difficulty can be alleviated using modified electronic confining radii, as discussed below.

Optimization of E^{rep} and the Confining Radii. As the next step, we have tested different values of confining radii used for obtaining the atomic wave functions (2.4, 2.7, and 3.0 for carbon and 2.5, 2.7, 3.0, and 3.3 for hydrogen) and the atomic input densities (5.0, 7.0, and 9.0 for carbon and 2.5, 3.0, 3.5, and 5.0 for hydrogen). The repulsive potentials V_{AB}^{rep} have been determined for every combination of these confining radii. We again used the two approaches mentioned above, including (opt3 and opt5) and excluding (opt4 and opt6) the fit of the electronic atomic energies E_A^{el}. The parameter sets opt3 and opt4 are optimized to give accurate atomization energies and geometries, while opt5 and opt6 are designed to improve vibrational frequencies. The additional conditions are found to be similar for different confining radii. They are further optimized once the "best" confining radii have been found and are shown in Table 4. The confining radii are determined by testing all parameter sets on training set 2. Table 6 gives an overview over all constructed repulsive potentials indicating that various properties may require different electronic parameters, that is, this demonstrates the limits of the transferability of the electronic part of DFTB.

The repulsive potential sets opt3 and opt4 are fitted to give the most accurate atomization and reaction energies with acceptable errors for equilibrium geometries and frequencies. While geometries are described satisfactorily, opt3 shows much smaller errors for atomization and reaction energies. In parameter set opt4 (as well as in opt1 and opt2), large atomization energy errors are found for small cyclic structures such as cyclopropane and cyclopropene. The superb performance of opt3 for the atomization and reaction energies is further confirmed on a larger testing set in the following section. A slight shadow on the future performance of opt3 is cast by the fact that adding additional elements to the fitting scheme may severely perturb the delicate balance between E_C^{el} and E_H^{el} necessary for obtaining the reported very good agreement between the SCC-DFTB and experimental atomization and reaction energies. The perturbation is inevitably associated with the fitting procedure since the parameters for all elements are linked via the atomic electronic energies as discussed above.

The repulsive potential sets opt5 and opt6 are constructed to obtain small errors for vibrational frequencies. Geometries are still accurate; however, the errors of atomization and reaction energies are much larger now. The results for the training set 2 (see Table 6) show improved vibrational frequencies with errors comparable to those of the M2005 parameter set. In comparison to the original mio parameter set, we want to point out the improvement of the CC stretch frequencies as shown in Table 5.

TABLE 7: Mean and Maximum Absolute Deviation of Several Molecular Properties of the G3/99 Test Set

property[a]	N[b]	opt1	opt2	opt3	opt4	mio	PBE[c]	B3LYP[c]
ΔH_f (kcal/mol)	39	4.8	5.8	2.2	7.1	55.4[d]	26.0	7.4
ΔH_f^{max} (kcal/mol)		19.2	29.0	16.3	25.2	114.8[d]	81.3	17.3
ΔH_f (kcal/mol)[e]	39	4.3	4.0	2.1	3.7	4.1	1.8	3.4
ΔH_f^{max} (kcal/mol)[e]		19.7	25.1	16.2	20.5	21.7	7.4	9.6
r (Å)	196	0.008	0.008	0.008	0.007	0.011	0.008	0.003
r_{max} (Å)		0.155	0.267	0.038	0.142	0.035	0.020	0.016
a (deg)	177	0.6	0.7	0.6	0.6	0.5	0.4	0.4
a_{max} (deg)		14.4	19.9	8.2	13.4	4.8	1.9	1.9
d (deg)	5	1.5	1.5	1.3	2.1	1.5	0.7	1.0
d_{max} (deg)		6.3	6.3	5.5	8.8	6.7	2.0	2.6

[a] Heats of Formation ΔH_f at 298 K, bond lengths r, bond angles a, and dihedral angles d; max stands for maximum absolute deviation. Geometric data is compared to the MP2/cc-pVTZ calculations. For details, see Supporting Information. [b] Number of comparisons. [c] Basis set 6-311G(2d,2p). [d] ΔH_f calculated using LDA spin polarization energies; taking PBE values instead gives a MAD of 13.2 kcal/mol and a maximal absolute deviation of 34.2 kcal/mol. [e] Calculated using atomic electronic energies fitted to 14 experimental atomization energies of training set 2.

The optimization of the confining radii significantly improves the atomization and reaction energies for opt3. For the other parameter sets, only small improvement is achieved. This shows that there is only a limited influence of the confining radii on the performance of SCC-DFTB. The difference between opt5 and opt3 stems from a different choice of the confining radii and the values for the additional equations. The input for the fitting procedure of opt6 differs from that of opt4 only by three additional equations for the C−C potential, as shown in Table 4. It is clear from the presented data that the quality of the C−C stretch frequencies can only be improved at the cost of deteriorating the atomization and reaction energies. As discussed above, we believe that this is an effect of the approximations inherently present in the current SCC-DFTB framework.

Test of the Parameters on Larger Molecule Sets

The G3/99 Molecule Set. In this subsection, we perform a test on the G3/99[40] molecule set containing H_2 and 38 hydrocarbons. The tested properties include heats of formation at 298.15 K and equilibrium geometries (see Table 7). The heats of formation are calculated as described in ref 41. The enthalpies of formation for gaseous atoms at 0 K and the $(H^{298} - H^0)$ values for hydrogen and carbon in their standard states are taken from experiment (ΔH_f^0(0 K) for hydrogen and carbon is 51.63 and 169.98 kcal/mol, respectively, and the corresponding values of $(H^{298} - H^0)$ are 1.01 and 0.25 kcal/mol).[41] The heat capacity corrections for molecular vibrations are estimated as

$$E^{vib} = R\sum_k \theta_k\left(\frac{1}{2} + \frac{1}{e^{\theta_k/T} - 1}\right) \qquad \text{where} \quad \theta_k = \frac{h\nu_k}{k_B} \tag{28}$$

using harmonic vibrational frequencies ν calculated analytically with SCC-DFTB.[39] R is the molar gas constant, θ is the vibrational temperature, T is the temperature, and h and k_B are the Planck and the Boltzmann constants, respectively. Further thermal corrections are included within the classical approximation for translations ((3/2)RT), rotations ((3/2)RT for nonlinear and RT for linear molecules), and the PV term (RT).

TABLE 8: Atomic Electronic Energies Fitted to 14 Experimental Atomization Energies of Training Set 2

element	opt1	opt2	opt3	opt4	opt5
H	−0.2564	−0.2802	−0.2514	−0.2808	−0.2544
C	−1.4516	−1.4416	−1.4176	−1.4404	−1.4392

	opt6	mio	M2005	PBE[a]	B3LYP[a]
H	−0.2819	−0.2762	−0.2603	−0.4965	−0.5026
C	−1.4422	−1.4545	−1.4409	−37.8066	−37.8538

[a] Fitted for basis set 6-311G(2d,2p).

Further, two approaches for the calculation of atomization energies have been chosen. In the first approach, we have used the atomic electronic energies as discussed in the Theory section. The results are displayed in the first two lines of Table 7. In the second approach, the atomic electronic energies are fitted to yield least-squares errors for the 14 experimental atomization energies of training set 2. These results are displayed in the third and fourth lines of Table 7. It is not surprising that the latter approach reveals smaller errors for the heats of formation. The fitted atomic electronic energies are listed for each parameter set in Table 8. Heats of formation are also calculated for the widely used density functionals PBE and B3LYP. For PBE calculations, the heats of formation are largely underestimated, whereas for B3LYP, they are overestimated. The errors are readily reduced when fitting the atomic energies. The corresponding values can be found in Table 8. Substantial improvements using optimized atomic energies have been noted in other studies.[27,42−45]

Set opt3 shows an excellent performance with a MAD of 2.2 kcal/mol for heats of formation; the largest deviations are found for azulene and methylene (respectively, −7.3 and −16.3 kcal/mol). The SCC-DFTB bond lengths and angles compared to MP2/cc-pVTZ geometries give an overall MAD of 0.008/0.007 Å and 0.6/0.6° for the opt3/opt4 parameter sets, showing good agreement with the reference data. Similar accuracy is yielded for the original mio, opt1, and opt2 parameter sets and both tested DFT methods. The largest discrepancies in the opt3/opt4 equilibrium geometries are observed for bicyclobutane, 0.038/0.142 Å for bond distances and 8.2/13.4° for bond angles. The largest MAD for opt3 is observed for the C═C and C−H bonds (0.008 Å), and the largest MAD for opt4 is for the C−C bond (0.013 Å). Keeping in mind that for bicyclobutane an exceptional large deviation is found, all other bond lengths are reproduced with good accuracy. The (signed) mean deviation shows that the C≡C bonds tend to be shorter, whereas all other bond lengths tend to be larger than the MP2/cc-pVTZ reference. More details can be found in the Supporting Information.

Larger Molecule Sets. Recently, Jorgensen and co-workers published a collection of experimental heats of formation, isomerization enthalpies, conformational energetics, and MP2/6-31 g(d) geometries,[46] which are very suitable to benchmark approximate methods. (While the comparisons on the G3/99 test set were made with MP2/cc-pVTZ geometries, MP2/6-31G(d) geometries for the Jorgensen test set were used. This facilitates comparison to the data compiled in refs 27 and 46. The mean absolute deviation (MAD) of MP2/6-31G(d) compared to MP2/cc-pVTZ bond lengths for the G3/99 test set is found to be only 0.004 Å. All data are listed in the Supporting Information.) We used this set to test our new parametrization, as shown in Tables 9−11.

Heats of Formation. The best performance is found for the parameter set opt3; with a MAD of 2.6 kcal/mol, it is comparable to the results of PDDG-PM3. The largest deviations are found for cubane (−35.5 kcal/mol) and diamantane (−13.0 kcal/mol). In general, large deviations are found for large aromatic systems such as anthracene, azulene, and biphenylene, bicyclic structures, and highly substituted cyclic compounds. Similar problems have been detected for opt4. As already mentioned for the G3/99 test set, a refitting of the atomic electronic energies on the 14 experimental atomization energies of training set 2 generally reduces this error (third and fourth lines of Table 9). This effect is significant, for example, for the mio set, where the mean absolute deviation is reduced to only 3.9 kcal/mol. Note that this refitting of atomic electronic energies is only done for hydrocarbons. When including further elements, a refit also influences hydrogen and carbon; thus, also the heats of formation for hydrocarbons are expected to be less accurate.

Geometries. Generally, SCC-DFTB describes geometries very well, being slightly more accurate than PDDG-PM3 (with the exception of the frequency-optimized sets opt6 and M2005). Again, the best performance is observed for the opt3 and mio sets with maximal absolute deviations for bond lengths smaller than 0.04 Å. The largest bond length error is consistently found for the bridge CC bond in bicyclobutane (up to 0.27 Å with opt2 and M2005). The second largest deviation is already much smaller; the error for the C−C bond in 1,3-butadiyne with opt2, opt3, and opt4 is smaller than 0.024 Å, and the error for the C≡C bond in acetylene with opt1 and opt5 is smaller than 0.031 Å. Only for the frequency-optimized set opt6 is the second largest error 0.046 Å for the C−C bond of 1,3-butadiyne. Similar observations are found for bond angles and dihedrals. While bicyclobutane gives exceptionally large errors, deviations for all other bond angles are within 3.4°, and for dihedrals they are within 8°.

Isomerization Enthalpies. Table 10 compares calculated enthalpies (including corrections for the zero-point energy and temperature, as described above) for a selected set of isomerization reactions.[27] The averaged deviation for the original mio parameter set is 4.5 kcal/mol. The performance of SCC-DFTB can be improved with the parameter set opt3, giving a MAD of only 2.2 kcal/mol, which is of comparable accuracy as that for PDDG-PM3 (2.4 kcal/mol). The main reason for the better performance of the opt3 set over the mio set is a better description of small cyclic hydrocarbons (cyclopropene and cyclopropane) by the former set. The parameter set opt4 on the other hand does not show any improvement over the original mio set. Probably the most serious problem of opt4 is the wrong sign of the isomerization enthalpies for the isomerization of sterically crowded alkanes to the corresponding linear isomers (2,2-dimethylpropane → *n*-pentane and 2,2,3,3-tetramethylbutane → *n*-octane), leading to qualitatively wrong information about the relative stability of linear alkanes. A dispersion correction as proposed in ref 13 improves these results only marginally, yielding isomerization enthalpies of 0.0 and −2.0 kcal/mol. For the sets opt1, opt2, opt5, opt6, and M2005, the mean absolute deviations are 4.6, 4.9, 5.9, 5.7, and 8.1 kcal/mol, respectively.

Conformational Energetics. The comparison of results shown in Table 11 is based on a small compilation of conformational energies given in ref 27. All of the tested parameter sets give performance similar or slightly better than PDDG-PM3. The largest error is found for *cis*-1,3-dimethylcyclohexane. Qualitative errors are found only with the M2005 parameter set for butane, methylcyclohexane, and *cis*-1,3-dimethylcyclohexane, for which the stability pattern is reversed with respect to experiment. This behavior is probably related to significantly larger cutoff radii used to construct the M2005 repulsive

TABLE 9: Mean and Maximum Absolute Deviations of Several Molecular Properties of the Jorgensen Test Set

property[a]	N^b	opt1	opt2	opt3	opt4	opt5	opt6	mio	M2005	PDDG[c]
ΔH_f	254	6.4	6.1	2.6	9.2	24.3	15.7	87.9[d]	159.0	2.6
ΔH_f^{max}		32.8	29.1	35.5	29.4	76.3	66.9	184.6[d]	337.5	39.1
ΔH_f^{e}	254	5.1	3.2	2.4	3.8	6.8	7.7	3.9	7.2	
ΔH_f^{maxe}		31.6	25.2	35.2	20.5	27.8	56.6	21.7	30.5	
r	111	0.007	0.008	0.005	0.007	0.008	0.018	0.009	0.020	0.011
r_{max}		0.154	0.266	0.037	0.141	0.196	0.178	0.034	0.266	0.057
a	57	0.8	1.0	0.7	0.7	0.8	0.9	0.6	1.3	1.3
a_{max}		16.4	21.9	10.2	15.4	18.6	17.5	6.8	22.1	11.8
d	20	1.6	1.9	1.0	1.7	1.9	1.7	1.3	2.7	2.9
d_{max}		8.2	9.6	5.9	9.3	8.9	7.8	9.4	13.4	17.1

[a] Heats of formation ΔH_f at 298 K in kcal/mol, bond lengths r in Å, bond angles a, and dihedral angles d in degrees; max stands for maximum absolut deviation. Geometric data is compared to MP2/6-31 g(d) calculations. For details, see ref 46. [b] Number of comparisons. [c] PDDG-PM3 values from ref 27. [d] ΔH_f calculated using LDA spin polarization energies; taking PBE values instead gives a MAD of 19.9 kcal/mol and a maximal absolute deviation of 88.9 kcal/mol. [e] Calculated using atomic electronic energies fitted to 14 experimental atomization energies of training set 2.

TABLE 10: Deviation from Experiment for Selected Isomerization Enthalpies (kcal/mol) at 298.15 K

			exp[a]	mio	opt3	opt4	PDDG[a]
propyne	→	allene	1.2	+4.0	+2.7	+4.8	+4.63
propyne	→	cyclopropene	21.8	+16.5	+6.5	+18.0	−0.23
propene	→	cyclopropane	7.9	+6.3	+0.5	+6.8	+0.29
trans-2-butene	→	*cis*-2-butene	1.1	−0.0	+0.1	−0.5	+0.53
2-methylpropene	→	*trans*-2-butene	1.3	−0.1	+0.6	−1.1	−1.91
trans-2-butene	→	1-butene	2.8	+1.3	+0.8	+0.6	+2.18
1,3-butadiene	→	cyclobutene	11.3	+2.2	−3.8	+1.7	−3.51
cyclopentene	→	vinylcyclopropane	22.2	+11.5	+3.2	+9.1	+2.73
1-*trans*-3-pentadiene	→	1,4-pentadiene	7.1	+1.0	+0.6	+0.2	−1.77
2,2-dimethylpropane	→	*n*-pentane	5.1	−2.4	−0.1	−5.3	+2.09
2,2,3,3-tetramethylbutane	→	*n* -octane	4.1	−3.4	+0.8	−7.4	+3.14
toluene	→	norbornadiene	46.7	+2.6	−4.2	+4.9	−2.72
styrene	→	cyclooctatetraene	35.8	+7.7	+5.3	+4.8	−4.92
mean absolute deviation				4.5	2.2	5.0	2.4

[a] Experimental and PDDG-PM3 values from ref 27.

TABLE 11: Deviation for Conformational Energetics (kcal/mol)

		ref ΔE^a	mio	opt3	opt4	PDDG[a]
butane	anti versus gauche	0.7	−0.2	−0.2	−0.4	−0.4
ethane	anti versus eclipsed	2.8	−0.5	−0.7	−0.8	−1.7
methylcyclohexane	eq versus ax	1.8	−0.9	−0.8	−1.3	−0.9
cis-1,3-dimethylcyclohexane	eq,eq versus ax,ax	5.5	−2.1	−1.9	−3.3	−2.3
propene	eclipsed versus anti	2.0	−0.9	−0.9	−1.0	−1.3
1,3-butadiene	trans versus skew	2.5	−1.4	−0.9	−1.2	−1.8
mean absolute deviation			1.0	0.9	1.3	1.4

[a] Reference and PDDG-PM3 values from ref 27.

potential. For M2005, the cutoff radii are set to 7 atomic units, whereas for mio, the cutoff radii are chosen in a way to ensure that the second neighbor interaction vanishes. This results in cutoff radii of 2.64 au for the H−H, 3.5 au for the C−H, and 4.3 au for the C−C repulsive pair potential. A geometrical analysis of the structures of *anti*- and *gauche*-butane shows that the distances between the first and third carbon atoms is 4.916 and 4.939 au, respectively. Because for shorter distance the repulsive energy is higher, the anti conformer is artificially destabilized by the M2005 parameter set. Clearly, for the mio set, the analogous repulsive energy contribution is zero for both conformers.

Cations, Radicals, and Anions. The heats of formation for 19 cations, radicals, and anions (species containing only carbon and hydrogen) compiled by Jorgensen and co-workers[27] show a mean absolute deviation to experimental data of 9.4 kcal/mol for PDDG/PM3. For the original mio parameter set, which uses the PBE atomic spin polarization energies, a mean absolute deviation of 10.8 kcal/mol is found. This deviation is reduced to 9.3 kcal/mol only for the opt3 set; this value is similar to the value of PDDG/PM3. For opt1, opt2, and opt4, the respective errors are 11.0, 11.6, and 13.2 kcal/mol. When fitting the atomic electronic energies to the atomization energies of training set 2, the mean absolute deviations do not change significantly. More details can be found in the Supporting Information.

Vibrational Frequencies. The implementation of analytical second derivatives[39] greatly simplified the calculation of harmonic vibrational frequencies within the SCC-DFTB formalism. We have used this new functionality to determine the SCC-DFTB frequencies for a group of 14 hydrocarbons representing a variety of typical bonding situations. Mean and maximal absolute deviations with respect to experimental fundamental vibrational frequencies are given in Tables 12 and 13, respectively. The contributions for doubly degenerate vibrations have been included twice in the averaging, and those for triply degenerate vibrations have been included thrice. The total number of computed vibrational frequencies is 349. The number of available experimental modes[36,47] is only 346. Therefore, for

TABLE 12: Mean Absolute Deviation of Calculated Harmonic Vibrational Frequencies from Experimental Fundamental Frequencies for a Group of 14 Hydrocarbons

molecule	BLYP[a]	opt1	opt2	opt3	opt4	opt5	opt6	mio	M2005
methane	23	55	30	47	34	41	30	65	44
ethyne	32	30	83	101	57	47	54	70	28
ethene	19	40	60	50	39	36	36	49	34
ethane	20	40	34	28	29	26	24	52	28
allene	13	53	74	73	54	52	45	58	39
cyclopropane	17	43	56	52	42	35	37	55	33
propene	21	27	32	29	26	21	19	42	20
1,3-butadiene	18	35	48	49	33	31	27	46	21
bicyclo[2,1,0]pentane	50	40	46	45	40	41	34	52	42
spiropentane	32	52	58	59	47	41	40	68	52
cyclohexane	27	31	25	23	28	21	24	49	23
benzene	17	53	67	70	46	42	34	58	34
1,3,5-hexatriene	19	39	49	51	37	36	33	46	24
cubane	22	52	47	63	52	64	43	96	35
MAD (total)[b]	25	42	48	50	40	38	33	58	32

[a] Calculated with the cc-pVTZ basis set. [b] Mean absolute deviation of all 349 vibrational frequencies of the 14 molecules shown.

TABLE 13: Maximal Absolute Deviation of Calculated Harmonic Vibrational Frequencies versus Experimental Fundamental Frequencies for a Group of 14 Hydrocarbons

molecule	BLYP[a]	opt1	opt2	opt3	opt4	opt5	opt6	mio	M2005
methane	48	113	79	102	50	82	37	137	51
ethyne	70	69	159	250	107	81	103	140	72
ethene	64	145	189	158	152	143	102	201	82
ethane	58	64	63	54	53	45	45	136	48
allene	36	159	229	181	185	175	110	245	89
cyclopropane	40	81	133	111	77	78	87	149	75
propene	70	150	192	138	156	140	83	219	54
1,3-butadiene	72	149	191	142	153	143	94	214	64
bicyclo[2,1,0]pentane	158	125	136	137	145	163	127	154	140
spiropentane	110	283	274	306	259	194	180	435	137
cyclohexane	117	64	63	103	72	74	95	116	103
benzene	52	161	199	144	156	127	99	230	100
1,3,5-hexatriene	88	147	189	169	152	140	122	203	89
cubane	51	112	115	188	102	135	143	352	83
MAX (total)[b]	158	283	274	306	259	194	180	435	140

[a] Calculated with the cc-pVTZ basis set. [b] Maximal absolute deviation of all 349 vibrational frequencies of the 14 molecules shown.

two modes of 1,3,5-hexatriene and one mode of bicyclo[2,1,0]pentane, the DFT harmonic vibrational frequencies (B3LYP[48−50]/cc-pVTZ[51] with a scaling factor of 0.965[36]) have been used instead. The unscaled SCC-DFTB frequencies are compared with the DFT results (BLYP[48,52]/cc-pVTZ,[51] unscaled). In order to compare the intrisic accuracy, we decided to use unscaled SCC-DFTB frequencies. Further, it has been shown that for the mio set, a uniform scaling factor is very close to one.[28] All of the DFT calculations were performed using the GAUSSIAN 03 program.[53]

The mean absolute deviations of the 349 analyzed vibrational modes are noticeably reduced for all of the optimized parameter sets in comparison to the original mio parameter set. The results are most significantly improved for the frequency-optimized sets opt5 and opt6, with deviations of 38 and 33 cm^{-1}, respectively. The latter set of repulsive potentials gives similar accuracy as the M2005 parameter set (32 cm^{-1}), but both of them give approximately 15% larger error than the BLYP/cc-pVTZ computational scheme (25 cm^{-1}). The largest deviations to experiment are observed for the SCC-DFTB frequencies of allene, bicyclo[2,1,0]pentane, and spiropentane. A complete list of all of the computed and experimental frequencies can be found in the Supporting Information.

Linear Alkanes. Recent comparisons of heats of formation for linear alkanes using G3[54] and B3LYP[42] showed that both computational schemes display a cummulative error, that is, that the total error in the computed heats of formation grows with the length of the chain. This is a systematic error which we consider worth reducing or even eliminating. Using a dispersion correction[55] and fitting the atomic energies for DFT improves the situation but does not solve this problem completely.[42] Similar results are found for SCC-DFTB, as shown in Figure 2. For the mio set, the overbinding is obvious. For opt4, a small underbinding is detected; it can be partially eliminated by including a dispersion correction.[13] No cumulative error is found

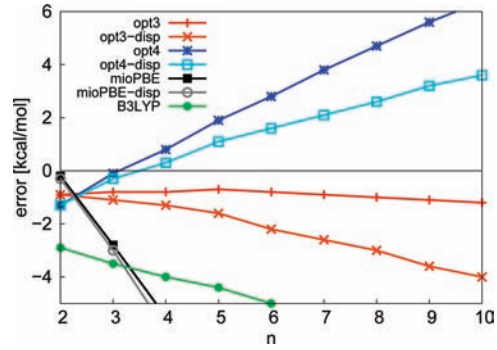

Figure 2. Errors in computed heats of formation for linear alkanes C_nH_{2n+2} for opt3, opt4, and the mio(PBE) parameter sets and also including disperion corrections. The results are compared with B3LYP/6-31G(d) results, including a dispersion correction from ref 42. The line connecting the points is plot to facilitate the comparisons.

for any of these sets if the atomic energies are fitted to minimize the errors of heats of formation of these linear alkanes. With this result, it is not too much of a surprise that for opt3, no error accumulation is found. As discussed above, the atomic energies of opt3 are already fitted to training set 1, which partially accounts for the accumulation problem.

Conclusion

We have presented a new methodology which greatly simplifies the generation of the SCC-DFTB repulsive potentials and considerably improves the accuracy, in particular, for heats of formation and vibrational frequencies. The method is based on a solution of a linear inverse problem for a set of repulsive potentials for a given group of elements. Since in most cases, the effective linear problem is not directly invertible (being either over- or underdetermined), a singular value decomposition approach is used to extract the meaningful portion of information from the effective linear problem. The application of fourth-order splines results in smooth second derivatives, leading to an improved description of vibrational frequencies.

In this work, the formalism of the electronic part of SCC-DFTB has not been modified; however, we have slightly adjusted the values of atomic electronic energies and confining radii for atomic orbitals and densities entering the Hamilton matrix elements to improve the overall performance. The present work clearly shows that the approximations inherent in SCC-DFTB limit its overall performance. Heats of formation and vibrational frequencies cannot be simultaneously optimized to an accuracy comparable with that of full DFT methods. Therefore, we suggest using a special parametrization for vibrational frequencies, when needed (sets opt5 and opt6).

The SCC-DFTB repulsive potentials have up to now been derived for a variety of elements, including C, H, N, O, Mg, P, S, and so forth. The repulsive potentials depend on the electronic parameters for which they have been fitted, in particular, the confining radii. In a first step, we have derived optimized repulsive parameters using the confining radii of the actual parameter set. Therefore, the sets opt1 and opt2, which show improved properties for the repulsive potentials for C and H, can be used in conjunction with the parameters with the other elements parametrized so far.

On the other hand, we have optimized the confining radii as well (opt3, opt4, opt5, opt6). These sets are no longer consistent with the parameters of the other elements, that is, the aim of ongoing and future work will be to also reoptimize the parameters of other elements. The variation of electronic parameters improves all properties, most significantly, the heats of formation, however within a limited range. This means that the electronic parameters cannot be used for a significant improvement of performance, for example, within a brute force fitting approach. On the other hand, this shows the robustness of SCC-DFTB, that is, a variation of electronic parameters will also not lead to significant failures, of course within the limits of DFT-GGA. With opt3, SCC-DFTB can successfully compete with the special parametrization PDDG-PM3 and also with B3LYP, being much better than PBE at least for hydrocarbons. This shows that with a suitable choice of parameter systematic deficiencies of the underlying PBE functional can be improved. Whether this favorable behavior will remain for other atom types is the subject of our current investigation.

A fundamental decision in the parametrization procedure is whether the atomic parameters (atomic electronic energies) and repulsive potentials are optimized in one step. This is done for the sets opt1, opt3, and opt5 and is a general strategy for the parametrization of NDDO-type semiempirical methods. This however, influences the repulsive potential itself, that is, the repulsive potentials are arbitrarily shifted to optimize the performance. This means that part of the energy of a bond, which is described by parameters depending on two centers (pairwise potentials), is shifted to atomic parameters; in the case of SCC-DFTB, it is shifted to the atomic electronic energies. So far, this seems to pose no problem as long as parameters for all atom types are determined in one optimization step. This means, however, that the parametrization for new atom types will lead to an adjustment of the already parametrized ones, which will lead to an increasing complexity of the problem when more elements are to be included. Therefore, we consider the approach 2, where only repulsive potentials are optimized (opt2, opt4, opt6), to be more practical in this respect. The atomic values can still be reoptimized after determining E^{rep}, leading still to a slight improvement, although the performance of approach 1 cannot be completely matched. Ongoing work will clarify this point in more detail.

Acknowledgment. This work was supported in part by the Deutsche Forschungsgemeinschaft through Forschergruppe 490 (M.E.), by the National Science Council of Taiwan (NSC96-2113-M-009-022), Ministry of Education MOE-ATU project, and by the National Center for High-Performance Computing NCHC, Hsinchu, Taiwan (H.W.).

Appendix

Continuity Equations. In the following, we are referring to one specific repulsive potential V_{XY}^{rep}, X and Y being two different types of atoms as mentioned for eqs 11−14. These conditions can then be conveniently written in matrix language after introducing a column vector $\boldsymbol{a}_3$ containing the unknown parameters $a_{0,i}$, $a_{1,i}$, $a_{2,i}$, and $a_{3,i}$

$$\boldsymbol{a}_3^{\mathrm{T}} = [a_{0,1}a_{1,1}a_{2,1}a_{3,1}...a_{0,n}a_{1,n}a_{2,n}a_{3,n}] \tag{29}$$

and a column vector $\boldsymbol{a}_4$ containing the unknown parameters $a_{4,i}$

$$\boldsymbol{a}_4^{\mathrm{T}} = [a_{4,1}a_{4,2}...a_{4,n}] \tag{30}$$

and three auxiliary matrices Q_i, R_i, and S given by

$$Q_i = \begin{bmatrix} 1 & (r_{i+1} - r_i) & (r_{i+1} - r_i)^2 & (r_{i+1} - r_i)^3 \\ 0 & 1 & 2(r_{i+1} - r_i) & 3(r_{i+1} - r_i)^2 \\ 0 & 0 & 2 & 6(r_{i+1} - r_i) \\ 0 & 0 & 0 & 6 \end{bmatrix} \tag{31}$$

$$R_i = \begin{bmatrix} (r_{i+1} - r_i)^4 \\ 4(r_{i+1} - r_i)^3 \\ 12(r_{i+1} - r_i)^2 \\ 24(r_{i+1} - r_i) \end{bmatrix} \quad \text{and} \quad S = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & -6 \end{bmatrix} \tag{32}$$

Then, eqs 11−14 can be expressed concisely using block matrices T and W as

$$\boldsymbol{T}\boldsymbol{a}_3 + \boldsymbol{W}\boldsymbol{a}_4 = \begin{bmatrix} Q_1 & S & 0 & \cdots & 0 \\ 0 & Q_2 & S & \ddots & \vdots \\ 0 & 0 & Q_3 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & S \\ 0 & 0 & \cdots & 0 & Q_n \end{bmatrix} \boldsymbol{a}_3 + \begin{bmatrix} R_1 & 0 & 0 & \cdots & 0 \\ 0 & R_2 & 0 & \ddots & \vdots \\ 0 & 0 & R_3 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & R_n \end{bmatrix} \boldsymbol{a}_4 = 0 \quad (33)$$

Clearly, the matrix T, being an upper triangular matrix with nonzero diagonal elements, is nonsingular and can be inverted, yielding

$$\boldsymbol{a}_3 = -T^{-1}W\boldsymbol{a}_4 \quad (34)$$

As mentioned earlier, the previously used third-order spline does not provide an accurate representation of the repulsive portion of the Hessian. Equation 34 shows that using fourth-order splines does not increase the computational complexity of the problem in comparison to the standard third-order spline, owing to the existence of an additional continuity condition. This reasoning shows also that V_{AB}^{rep} is completely determined if n parameters $a_{4,1}$, $a_{4,2}$, ..., $a_{4,n}$ are specified.

The remaining unknown parameters $a_{4,1}$, $a_{4,2}$, ..., $a_{4,n}$ defining each of the repulsive potentials are fitted to reproduce a set of atomization energies and equilibrium structures for a chosen group of molecules. In the next two subsections, we show how the SCC-DFTB atomization energies and equilibrium geometries can be represented in terms of the unknown parameters $a_{4,1}$, $a_{4,2}$, ..., $a_{4,n}$.

Energy Equations. To derive the working energy equation for molecule M, eq 19 is cast in a matrix form. For the clarity of presentation, it is convenient to assume that the molecule M is built only of two different types of atoms, say X and Y (similarly like C_6H_6 is built only of carbon and hydrogen). It is easy to see that the discussion of a general case does not involve any further complications but would lead to more cumbersome notation. Then, the LHS of eq 19 depends only on three different repulsive potentials, V_{XX}^{rep}, V_{XY}^{rep}, and V_{YY}^{rep}, and two unknown atomic electronic energies, E_X^{el} and E_Y^{el}. Let us examine in detail a single component of the LHS of eq 17, $V_{AB}^{\mathrm{rep}}(r_{AB})$. We may further assume without loss of generality that $V_{AB}^{\mathrm{rep}} = V_{XX}^{\mathrm{rep}}$ and that r_{AB} lies inside of some interval I_{p}, r_{p} being the corresponding division point. This allows for writing

$$V_{AB}^{\mathrm{rep}}(r_{AB}) = \sum_{k=0}^{4} a_{k,\mathrm{p}}^{XX}(r_{AB} - r_{\mathrm{p}})^k \quad (35)$$

which can be rewritten in a matrix form as

$$V_{AB}^{\mathrm{rep}}(r_{AB}) = Z_3\boldsymbol{a}_3^{XX} + Z_4\boldsymbol{a}_4^{XX} \quad (36)$$

Here, the row vectors Z_3 and Z_4 are given explicitly by

$$Z_3 = [0 \ \cdots \ 0 \ \ 1 \ \ (r_{AB} - r_{\mathrm{p}}) \ \ (r_{AB} - r_{\mathrm{p}})^2 \ \ (r_{AB} - r_{\mathrm{p}})^3 \ \ 0 \ \cdots \ 0] \quad (37)$$

and

$$Z_4 = [0 \ \cdots \ 0 \ \ (r_{AB} - r_{\mathrm{p}})^4 \ \ 0 \ \cdots \ 0] \quad (38)$$

The position of the first nonzero component of Z_3 is $4p - 3$. The position of the only nonzero component of Z_4 is p. Using eq 31 allows for writing eq 33 only as a function of $\mathbf{a}_4^{XX}$ as

$$V_{AB}^{\mathrm{rep}}(r_{AB}) = (Z_4 - Z_3T^{-1}W)\boldsymbol{a}_4^{XX} \quad (39)$$

Let us denote by $\boldsymbol{a}$ the vector of all of the unknown parameters; it can be represented as a block vector

$$\boldsymbol{a} = \begin{bmatrix} \boldsymbol{a}_4^{XX} \\ \boldsymbol{a}_4^{XY} \\ \boldsymbol{a}_4^{YY} \\ E_X^{\mathrm{el}} \\ E_Y^{\mathrm{el}} \end{bmatrix} \quad (40)$$

Then, $V_{AB}^{\mathrm{rep}}(r_{AB})$ can be written as

$$V_{AB}^{\mathrm{rep}}(r_{AB}) = [(Z_4 - Z_3T^{-1}W) \ \ \mathbf{0} \ \ \mathbf{0} \ \ 0 \ \ 0] \begin{bmatrix} \boldsymbol{a}_4^{XX} \\ \boldsymbol{a}_4^{XY} \\ \boldsymbol{a}_4^{YY} \\ E_X^{\mathrm{el}} \\ E_Y^{\mathrm{el}} \end{bmatrix} = Z_{AB}\boldsymbol{a} \quad (41)$$

Similarly, the sum of all atomic electronic energies can be written in an analogous way

$$\sum_A E_A^{\mathrm{el}} = n_X E_X^{\mathrm{el}} + n_Y E_Y^{\mathrm{el}} = [\mathbf{0} \ \ \mathbf{0} \ \ \mathbf{0} \ \ n_X \ \ n_Y] \begin{bmatrix} \boldsymbol{a}_4^{XX} \\ \boldsymbol{a}_4^{XY} \\ \boldsymbol{a}_4^{YY} \\ E_X^{\mathrm{el}} \\ E_Y^{\mathrm{el}} \end{bmatrix} = G\boldsymbol{a} \quad (42)$$

where n_X and n_Y are the number of atoms X and Y, respectively, in the molecule M and G is a shortcut for the row vector on the RHS containing n_X and n_Y. Substituting these equations into eq 19 yields the final working energy equation for molecule M in a matrix form

$$Z_M\boldsymbol{a} = (\sum_{A>B} Z_{AB} - G)\boldsymbol{a} = E^{\mathrm{at}} - E^{\mathrm{el}} \quad (43)$$

where Z_M is a shortcut for the term in parentheses. These equations can be readily extended to account for a larger number of atom types.

Force Equations. To derive the force working equations, the first task is to cast eq 21 in a matrix form. To simplify the discussion, we again assume that the system of interest consists only of two atom types, X and Y. As above, we choose $V_{AB}^{\mathrm{rep}} = V_{XX}^{\mathrm{rep}}$ and assume that r_{AB} lies inside of the interval I_{p}. Then, we can easily evaluate the derivative in eq 22 as

$$\left.\frac{\mathrm{d}V_{AB}^{\mathrm{rep}}}{\mathrm{d}r}\right|_{r=r_{AB}} = \sum_{k=1}^{4} a_{k,\mathrm{p}}^{XX} \cdot k \cdot (r_{AB} - r_{\mathrm{p}})^{k-1} \tag{44}$$

which can be readily rewritten in a matrix form as

$$\left.\frac{\mathrm{d}V_{AB}^{\mathrm{rep}}}{\mathrm{d}r}\right|_{r=r_{AB}} = P_3 \boldsymbol{a}_3^{XX} + P_4 \boldsymbol{a}_4^{XX} \tag{45}$$

Here, the row vectors P_3 and P_4 are given explicitly by

$$P_3 = [0 \quad \cdots \quad 0 \quad 1 \quad 2 \quad (r_{AB} - r_{\mathrm{p}}) \quad 3(r_{AB} - r_{\mathrm{p}})^2 \quad 0 \quad \cdots \quad 0] \tag{46}$$

and

$$P_4 = [0 \quad \cdots \quad 0 \quad 4(r_{AB} - r_{\mathrm{p}})^3 \quad 0 \quad \cdots \quad 0] \tag{47}$$

The position of the first nonzero component of P_3 is $4p - 2$. The position of the only nonzero component of P_4 is p. Using eq 34 allows for writing eq 45 as a function of $\boldsymbol{a}_4^{XX}$ only as

$$\left.\frac{\mathrm{d}V_{AB}^{\mathrm{rep}}}{\mathrm{d}r}\right|_{r=r_{AB}} = (P_4 - P_3 T^{-1} W)\boldsymbol{a}_4^{XX} \tag{48}$$

Using the vector $\mathbf{a}$ of all of the unknowns defined in eq 40 enables expression of the derivative of V_{AB}^{rep} as

$$\left.\frac{\mathrm{d}V_{AB}^{\mathrm{rep}}}{\mathrm{d}r}\right|_{r=r_{AB}} = [(P_4 - P_3 T^{-1} W) \quad \mathbf{0} \quad \mathbf{0} \quad 0 \quad 0] \begin{bmatrix} \boldsymbol{a}_4^{XX} \\ \boldsymbol{a}_4^{XY} \\ \boldsymbol{a}_4^{YY} \\ E_X^{\mathrm{el}} \\ E_Y^{\mathrm{el}} \end{bmatrix} = P_{AB}\boldsymbol{a} \tag{49}$$

Substituting this equation into eq 22 and combining with eq 21 yields the final working equation for the force on atom A in a matrix form

$$P_{M,A}\boldsymbol{a} = -\left(\sum_{B \neq A} \frac{\boldsymbol{r}_A - \boldsymbol{r}_B}{r_{AB}} P_{AB}\right)\boldsymbol{a} = -\boldsymbol{F}_A^{\mathrm{el}} \tag{50}$$

Note that matrix $P_{M,A}$ consists of three rows. Therefore, the total number of force equations for the molecule M is $3n$, where n is the total number of atoms constituting the molecule M.

As mentioned above, it is also possible to employ nonequilibrium structures in our fitting procedure. The only complication is that for such a nonequilibrium structure, the total force vector must be calculated using some quantum chemical method and be combined with the SCC-DFTB electronic force vector to yield the RHS of eq 50.

Additional Equations. Equation 23 can be expressed in a matrix as

$$S_{\mathrm{p}}(r_z) = V = U_3 \boldsymbol{a}_3 + U_4 \boldsymbol{a}_4 \tag{51}$$

where U_3 and U_4 are explicitly given by

$$U_3 = [0 \quad \cdots \quad 0 \quad 1 \quad (r_z - r_{\mathrm{p}}) \quad (r_z - r_{\mathrm{p}})^2 \quad (r_z - r_{\mathrm{p}})^3 \quad 0 \quad \cdots \quad 0] \tag{52}$$

and

$$U_4 = [0 \quad \cdots \quad 0 \quad (r_z - r_{\mathrm{p}})^4 \quad 0 \quad \cdots \quad 0] \tag{53}$$

The position of the first nonzero component of U_3 is $4p - 3$. The position of the only nonzero component of U_4 is p. Using eq 34 allows for writing eq 51 only as a function of $\mathbf{a}_4$ as

$$S_{\mathrm{p}}(r_z) = V = (U_4 - U_3 T^{-1} W)\boldsymbol{a}_4 \tag{54}$$

Now let us again assume for simplicity that our task is to determine the repulsive potentials for molecules built only from two different types of atoms, X and Y, and that the additional condition concerns the repulsive potential between atoms of type X. Using the vector $\mathbf{a}$ of all of the unknowns defined in eq 37 enables expression of the condition $S_{\mathrm{p}}^{XX}(r_z) = V$ as

$$S_{\mathrm{p}}^{XX}(r_z) = V = [(U_4 - U_3 T^{-1} W) \quad \mathbf{0} \quad \mathbf{0} \quad 0 \quad 0] \begin{bmatrix} \boldsymbol{a}_4^{XX} \\ \boldsymbol{a}_4^{XY} \\ \boldsymbol{a}_4^{YY} \\ E_X^{\mathrm{el}} \\ E_Y^{\mathrm{el}} \end{bmatrix} = U\boldsymbol{a} \tag{55}$$

As mentioned above, similar additional conditions can be imposed not only on repulsive potentials but also on their derivatives. The only modification required for this purpose is using appropriate derivatives of vectors U_3 and U_4 in eqs 52 and 53. It is probably already obvious for a careful reader that the previously discussed vectors P_3 and P_4 in eq 49 are the first derivatives of the vectors Z_3 and Z_4 in eq 41.

Reaction Equations. Starting from eq 24, we construct four energy equations for molecules M_1, M_2, M_3, and M_4 as described in eq 43. Adding these equations with the appropriate reaction coefficients yields a working equation for reaction energies in the form

$$J^{\mathrm{rea}}\boldsymbol{a} = (n_3 Z_{M_3} + n_4 Z_{M_4} - n_1 Z_{M_1} - n_2 Z_{M_2})\boldsymbol{a} = E^{\mathrm{rea}} - (n_3 E_{M_3}^{\mathrm{el}} + n_4 \boldsymbol{E}_{M_4}^{\mathrm{el}} - n_1 E_{M_1}^{\mathrm{el}} - n_2 E_{M_2}^{\mathrm{el}}) = L^{\mathrm{rea}} \tag{56}$$

which is in close analogy to the equations discussed previously for energies and forces. Note that the LHS of this equation does not depend explicitly on the atomic electronic energies E_X^{el} and E_Y^{el} since they effectively cancel out upon the addition.

Solving the System of Linear Equations. The equations for atomization energies, forces, reaction energies, and the additional conditions have the same structural form. Therefore, it is possible to write them as a single set of linear equations. After introducing auxiliary matrices and vectors

$$Z^{\mathrm{tot}} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_m \end{bmatrix} \qquad E^{\mathrm{ref}} = \begin{bmatrix} E_1^{\mathrm{at}} - E_1^{\mathrm{el}} \\ E_2^{\mathrm{at}} - E_2^{\mathrm{el}} \\ \vdots \\ E_m^{\mathrm{at}} - E_m^{\mathrm{el}} \end{bmatrix} \tag{57}$$

$$P^{\mathrm{tot}} = \begin{bmatrix} P_{1,1} \\ P_{1,2} \\ \vdots \\ P_{1,n_1} \\ P_{2,1} \\ P_{2,2} \\ \vdots \\ P_{2,n_2} \\ \vdots \\ P_{m,n_m} \end{bmatrix} \qquad F^{\mathrm{ref}} = \begin{bmatrix} -F_{1,1}^{\mathrm{el}} \\ -F_{1,2}^{\mathrm{el}} \\ \vdots \\ -F_{1,n_1}^{\mathrm{el}} \\ -F_{2,1}^{\mathrm{el}} \\ -F_{2,2}^{\mathrm{el}} \\ \vdots \\ -F_{2,n_2}^{\mathrm{el}} \\ \vdots \\ -F_{m,n_m}^{\mathrm{el}} \end{bmatrix} \tag{58}$$

$$U^{\mathrm{tot}} = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_o \end{bmatrix} \qquad U^{\mathrm{ref}} = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_o \end{bmatrix} \tag{59}$$

and

$$J^{\mathrm{tot}} = \begin{bmatrix} J_1 \\ J_2 \\ \vdots \\ J_s \end{bmatrix} \qquad L^{\mathrm{ref}} = \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_s \end{bmatrix} \tag{60}$$

the final system of equations can then be written as

$$\begin{bmatrix} Z^{\mathrm{tot}} \\ P^{\mathrm{tot}} \\ U^{\mathrm{tot}} \\ J^{\mathrm{tot}} \end{bmatrix} \boldsymbol{a} = \begin{bmatrix} E^{\mathrm{ref}} \\ F^{\mathrm{ref}} \\ U^{\mathrm{ref}} \\ L^{\mathrm{ref}} \end{bmatrix} \Leftrightarrow M\boldsymbol{a} = R \tag{61}$$

where M and R are shortcuts for the matrix on the LHS and the vector on the RHS, respectively. The symbols appearing in eqs 57−60 have the following meanings: m in eq 57 is the number of atomization energies used in the fitting procedure, the indices indicating specific molecules m and n_k in eq 58 are the number of equilibrium structures used in the fitting procedure and the number of atoms in the kth structure, respectively, o in eq 59 is the number of additional equations used, and s in eq 60 is the number of reaction energies used in the fitting procedure. Note that in eqs 50 and 58, $P_{i,j}$ is a matrix consisting of three rows, and $F_{i,j}^{\mathrm{el}}$ is a vector containing three entries.

The set of linear equations in eq 61 is usually either over- or underdetermined. We have found that its solution can be found in the most numerically stable fashion via the singular value decomposition[56] (SVD) of the matrix M. Thus, we can rewrite eq 61 as

$$M\boldsymbol{a} = U\Sigma V^{\mathrm{T}}\boldsymbol{a} = R \tag{62}$$

where U and V are orthogonal square matrices and Σ is a diagonal rectangular matrix containing the singular values sorted from the largest to the smallest. Assuming that all of the singular values are different than zero, the solution $\boldsymbol{a}$ can be expressed as

$$\boldsymbol{a} = V\Sigma^{-1}U^{\mathrm{T}}R \tag{63}$$

However, in most cases, some of the singular values will be very close to zero. In these cases, to prevent numerical artifacts, we truncate the singular values to the first t nonzero numbers. This procedure can be considered as using the concept of a generalized inverse to solve eq 61. Then, the solution **a** is given by the same formula as that in eq 63, with V and U truncated to their first t columns and Σ^{-1} being the inverse of the upper-left $t \times t$ segment of Σ.

The solution given by eq 63 determines all of the unknown electronic energies and the coefficients $a_{4,k}$ for all of the involved spline segments. Subsequently, the remaining spline coefficients are computed from eq 34.

Supporting Information Available: Excel file for computed and experimental data. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268.
(2) Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. *Phys. Rev. B* **1995**, *51*, 12947–12957.
(3) Widany, J.; Frauenheim, T.; Köhler, T.; Sternberg, M.; Porezag, D.; Jungnickel, G.; Seifert, G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *53*, 4443–4452.
(4) Witek, H. A.; Köhler, C.; Frauenheim, T.; Morokuma, K.; Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 5712–5719.
(5) Foulkes, W. M. C.; Haydock, R. *Phys. Rev. B* **1989**, *39*, 12520–12536.
(6) Harris, J. *Phys. Rev. B* **1985**, *31*, 1770–1779.
(7) Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 5614–5621.
(8) Yang, Y.; Yu, H.; York, D.; Cui, Q.; Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 10861–10873.
(9) Yang, Y.; Yu, H.; York, D.; Elstner, M.; Cui, Q. *J. Chem. Theory Comput.* **2008**, *4*, 2067–2084.
(10) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
(11) Wanko, M.; Hoffmann, M.; Frauenheim, T.; Elstner, M. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 511–518.
(12) Wanko, M.; Hoffmann, M.; Strodel, P.; Koslowski, A.; Thiel, W.; Neese, F.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2005**, *109*, 3606–3615.
(13) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149–5155.
(14) Elstner, M.; Frauenheim, T.; Suhai, S. *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 29–41.
(15) Elstner, M. *Theor. Chem. Acc.* **2006**, *116*, 316–325.
(16) Frauenheim, T.; Seifert, G.; Elstner, M.; Niehaus, T.; Köhler, C.; Amkreutz, M.; Sternberg, M.; Hajnal, Z.; Di Carlo, A.; Suhai, S. *J. Phys.: Condens. Matter* **2002**, *14*, 3015–3047.
(17) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
(18) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209–220.
(19) Elstner, M.; Jalkanen, K. J.; Knapp-Mohammady, M.; Frauenheim, T.; Suhai, S. *Chem. Phys.* **2000**, *256*, 15–27.
(20) Elstner, M.; Jalkanen, K. J.; Knapp-Mohammady, M.; Frauenheim, T.; Suhai, S. *Chem. Phys.* **2001**, *263*, 203–219.
(21) Kolb, M.; Thiel, W. *J. Comput. Chem.* **1993**, *14*, 775–789.
(22) Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495–506.

(23) Möhle, K.; Hofmann, H.; Thiel, W. *J. Comput. Chem.* **2001**, *22*, 509–520.
(24) Krüger, T.; Elstner, M.; Schiffels, P.; Frauenheim, T. *J. Chem. Phys.* **2005**, *122*, 1–5.
(25) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601–1622.
(26) Otte, N.; Scholten, M.; Thiel, W. *J. Phys. Chem. A* **2007**, *111*, 5751–5755.
(27) Sattelmeyer, K. W.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. A* **2006**, *110*, 13551–13559.
(28) Witek, H. A.; Morokuma, K. *J. Comput. Chem.* **2004**, *25*, 1858–1864.
(29) Witek, H. A.; Morokuma, K.; Stradomska, A. *J. Chem. Phys.* **2004**, *121*, 5171–5178.
(30) Witek, H. A.; Morokuma, K.; Stradomska, A. *J. Theor. Comput. Chem.* **2005**, *4*, 639–655.
(31) Witek, H. A.; Irle, S.; Zheng, G.; De Jong, W. A.; Morokuma, K. *J. Chem. Phys.* **2006**, *125*, 214706.
(32) Małolepsza, E.; Witek, H. A.; Morokuma, K. *Chem. Phys. Lett.* **2005**, *412*, 237–243.
(33) Elstner, M.; Cui, Q.; Munih, P.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Comput. Chem.* **2003**, *24*, 565–581.
(34) Eschrig, H. *Optimized LCAO Method and Electronic Structure of Extended Systems*; Springer: Berlin, Germany, 1989.
(35) Köhler, C.; Seifert, G.; Gerstmann, U.; Elstner, M.; Overhof, H.; Frauenheim, T. *Phys. Chem. Chem. Phys.* **2001**, *3*, 5109–5114.
(36) NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101, Release 14; Johnson, R. D., III, Ed.; http://srdata.nist.gov/cccbdb (Sept 2006).
(37) Knaup, J. M.; Hourahine, B.; Frauenheim, T. *J. Phys. Chem. A* **2007**, *111*, 5637–5641.
(38) Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **1999**, *110*, 7650–7657.
(39) Witek, H. A.; Irle, S.; Morokuma, K. *J. Chem. Phys.* **2004**, *121*, 5163–5170.
(40) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, *112*, 7374–7383.
(41) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063–1079.
(42) Tirado-Rives, J.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2008**, *4*, 297–306.
(43) Stewart, J. J. P. *J. Mol. Model.* **2004**, *10*, 6–12.
(44) Csonka, G. I.; Ruzsinszky, A.; Tao, J.; Perdew, J. P. *Int. J. Quantum Chem.* **2005**, *101*, 505–511.
(45) Brothers, E. N.; Scuseria, G. E. *J. Chem. Theory Comput.* **2006**, *2*, 1045–1049.
(46) Sattelmeyer, K. W.; Tubert-Brohman, I.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2006**, *2*, 413–419.
(47) Miaskiewicz, K.; Smith, D. A. *Chem. Phys. Lett.* **1997**, *270*, 376–381.
(48) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
(49) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
(50) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
(51) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
(52) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
(53) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
(54) Redfern, P. C.; Zapol, P.; Curtiss, L. A.; Raghavachari, K. *J. Phys. Chem. A* **2000**, *104*, 5850–5854.
(55) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787–1799.
(56) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes—The Art of Scientific Computing*, 3rd ed.; Cambridge University Press: New York, 2007.

JP902973M