JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Parametrization and Benchmark of DFTB3 for Organic Molecules

Michael Gaus,[†,‡] Albrecht Goez,[§] and Marcus Elstner*[,†]

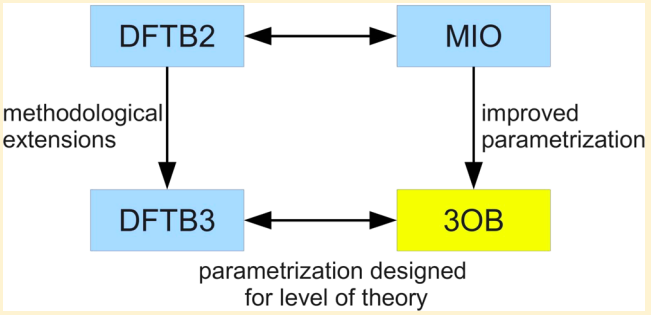
[†]Institute of Physical Chemistry, Karlsruhe Institute of Technology, Kaiserstr. 12, 76131 Karlsruhe, Germany

[‡]Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin—Madison, 1101 University Avenue, Madison, Wisconsin 53706, United States

[§]Institute for Physical and Theoretical Chemistry, Braunschweig Institute of Technology, Hans-Sommer-Str. 10, 38106 Braunschweig, Germany

Supporting Information

ABSTRACT: DFTB3 is a recent extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB) and derived from a third order expansion of the density functional theory (DFT) total energy around a given reference density. Being applied in combination with the parametrization of its predecessor (MIO), DFTB3 improves for hydrogen binding energies, proton affinities, and hydrogen transfer barriers. In the present study, parameters especially designed for DFTB3 are presented, and its performance is evaluated for small organic molecules focusing on thermochemistry, geometries, and vibrational frequencies from our own and several databases from literature. The new parameters remove significant overbinding errors, reduce errors for geometries of noncovalent interactions, and improve the overall performance.



1. INTRODUCTION

Fast semiempirical quantum mechanical methods are frequently used to resolve dynamical (bio-)molecular properties, compare large amounts of different conformations, or investigate properties of large systems. Major demands for these methods are their speed, robustness, accuracy, and transferability to different chemical environments. One of these methods that has frequently been used over the last years is the self-consistent-charge density-functional tight-binding method (DFTB2 or SCC-DFTB).[1] It is derived from a second order Taylor series expansion of the density functional theory (DFT) total energy expression around a reference density. DFTB2 estimates the zeroth and first order terms using a linear combination of atomic orbitals of a minimal basis, a two-center integral approximation, and neglects crystal-field and three-center integrals.[2] The density fluctuation in second order is approximated via a charge redistribution in a self-consistent manner, where the electron−electron interaction of excess charge on one atom is approximated by the Hubbard parameter. As the Coulomb interaction changes with respect to the charge on that atom, we have recently introduced an extension to the third order term of the Taylor series which accounts for that effect in a linear fashion.[3] Another modification of DFTB2 addressed the interpolation of the Coulomb interaction between large and zero interatomic distances improving specifically atom pair interactions including hydrogen. The method which results from combining these changes is called DFTB3 and broadens the applicability, especially in terms of accuracy for hydrogen binding energies, proton affinities, and proton transfer barriers without losing its speed and robustness.[3]

Another way to improve the accuracy of semiempirical methods apart from revising the formalism is to refine its parametrization; however, severe methodological approximations can be at most partially compensated for by a more sophisticated parametrization. DFTB3 contains several parameters. Along the original idea of developing an approximate DFT method,[4,5] most of them are determined from atomic DFT calculations. Other parameters are of a technical nature, e.g., the exponents for the basis set used for the atomic DFT calculations. Since a very large basis set is applied, the specific choice of the exponents within a physically reasonable range has only a minor influence on the results. The remaining parameters are physically motivated but are empirical and fitted to DFT, high-level ab initio, or experimental data. In the present study, we focus only on the latter type of parameters.

Another classification of the parameters distinguishes electronic and repulsive parameters. Besides the atomic DFT calculated and associated technical ones, some empirical parameters belong to the electronic part. These are responsible for confining atomic orbitals and atomic reference densities around the nuclei. Orbitals and densities of free atoms are usually too diffuse and are not suitable for precalculating the DFTB two-center integrals.[1,6] As a result of this confinement, the Pauli

Received: October 1, 2012

   dx.doi.org/10.1021/ct300849w | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

repulsion between closed shell fragments, which is in particular important for nonbonding interactions, is underestimated. In summary, there are two empirical electronic parameters per element, the so-called wave function and density compression radii. Another empirical electronic parameter enters to account for the peculiarity of the Coulomb interaction for the hydrogen atom and was discussed in detail earlier.[3,7,8]

The repulsive parameters are distance-dependent, short-ranged, atomic pair potentials that are fitted to atomization or reaction energies, atomic forces, and/or second derivatives of the total energy with respect to nuclear coordinates. Besides the original procedure of optimizing repulsive potentials,[6,9] there have been recent developments focusing on a more automatized routine. In particular, we have developed a scheme that represents the repulsive potentials as spline functions.[10,11] A set of conditions is composed for carefully chosen reference molecules and reference values. The coefficients of the spline functions are then found by minimizing the squared errors. The parametrization engine developed later by Bodrog et al. includes choices for a variety of functional representations of the repulsive potential and involves options for automatically generating reference data.[12] In either case, the limitation of the automatization lies in the selection and weighting of the reference data combined with the technical parameters for interpolating between them. Our strategy is to maintain some portion of chemical intuition within the parametrization process in order to yield smooth potentials that are as accurate as possible for energetics and derived properties for a wide range of equilibrium and nonequilibrium molecular systems.

Till now, the so-called MIO parameters [www.dftb.org] as developed for DFTB2[1,9] were used in combination with DFTB3. After adding DFTB3-specific parameters, that is a reasonable choice because the new methodological developments only influence systems with significant atomic net charges.[3] Nevertheless, two main shortcomings of MIO are well-known, one of which is a strong overbinding for almost all covalent bonds. While this overbinding is balanced between different bond types resulting in error cancellation for reaction energies, it caused an overestimation of atomization energies and proton affinities [the latter property was then improved by fitting DFTB3-specific parameters instead of using the physically motivated path of calculating them from atomic DFT calculations]. A second drawback in the use of the MIO parameters are too short-ranged noncovalent interactions as, e.g., for hydrogen bonds. During the MIO parametrization, that resulted from a compromise between hydrogen binding energies and hydrogen bond lengths. Certain constraints leading to this optimization conflict are removed within DFTB3, which is a more flexible method in this respect. Therefore, a new parameter set based on DFTB3 is developed. Following our parametrization scheme,[10,11] we revise the wave function compression radii in order to increase the Pauli repulsion. This leads to a new balance between hydrogen bond lengths and hydrogen binding energies. In addition, we adapt the repulsive potentials and density compression radii to remove the overbinding which is apparent within the MIO parametrization.

While the name MIO stands for parameters for **m**aterials and b**io**logical systems, we call the new parameters 3OB, highlighting its design for DFTB**3** as well as **o**rganic and **b**iological applications. Starting with the most important elements for such applications, parameters are presented for carbon, hydrogen, nitrogen, and oxygen.

In the Theory section, we describe the DFTB3 methodology and explain where each parameter enters the formalism. Also, the fitting procedure for the repulsive parameters is summarized. Our new parameters are then tested using extensive benchmark sets of molecules where we focus on a comparison between the new 3OB and the older MIO parameters. Also, some data from PBE,[13] B3LYP,[14−16] BLYP,[14,17] and semiemipirical PDDG[18] and OM3[19] calculations are included. Remaining challenges are discussed and conclusions are drawn.

2. THEORY

In this section, we briefly introduce the working equations for the DFTB3 method that has been described in detail in ref 3 and point out where the electronic and repulsive parameters enter the formalism. A short summary is given of our recently introduced fitting procedure for the repulsive potentials.[10,11]

2.1. DFTB3. DFTB3 is derived from density functional theory (DFT) by a third-order expansion of the DFT total energy around an initial density ρ^0. After several approximations, the total energy can be written as

$$\begin{aligned} E^{\mathrm{DFTB3}} &= E^{\mathrm{H0}} + E^{\gamma} + E^{\Gamma} + E^{\mathrm{rep}} \\ &= \sum_{iab} \sum_{\mu\in a} \sum_{\nu\in b} n_i c_{\mu i} c_{\nu i} H^0_{\mu\nu} + \frac{1}{2}\sum_{ab} \Delta q_a \Delta q_b \gamma_{ab} \\ &\quad + \frac{1}{3}\sum_{ab} \Delta q_a^2 \Delta q_b \Gamma_{ab} + \frac{1}{2}\sum_{ab} V^{\mathrm{rep}}_{ab} \end{aligned} \tag{1}$$

n_i denotes the occupation number of molecular orbital ψ_i which is expanded in a minimal basis of confined pseudoatomic orbitals ϕ_μ as

$$\psi_i = \sum_\mu c_{\mu i}\phi_\mu \tag{2}$$

The ϕ_μ are expressed in terms of Slater-type orbitals and spherical harmonics (for details, see ref 6) as

$$\phi_\mu(\vec{r}) = \sum_{n a l_\mu m_\mu} a_{n\alpha} r^{l_\mu + n}\, \mathrm{e}^{-\alpha r}\, Y_{l_\mu l_\nu}(\vec{r}/r) \tag{3}$$

and are determined from a modified atomic DFT calculation

$$\left[\hat{T} + V^{\mathrm{nucleus}} + V^{\mathrm{Hartree}} + V^{\mathrm{xc}} + \left(\frac{r}{r^{\mathrm{wf}}}\right)^2\right]\phi_\mu = \varepsilon_i \phi_\mu \tag{4}$$

V^{xc} is expressed in terms of the PBE exchange correlation approximation,[13] and r^{wf} is a parameter called the wave function compression radius. The Hamilton matrix elements are written in a two-center approximation as

$$H^0_{\mu\nu} = \begin{cases} \varepsilon^{\mathrm{free\ atom}} & \text{if } \mu = \nu \\ \left\langle \phi_\mu \middle| \hat{T} + V[\rho_a^0 + \rho_b^0] \middle| \phi_\nu \right\rangle & \text{if } a \neq b \\ 0 & \text{if } a = b,\ \mu \neq \nu \end{cases} \tag{5}$$

The diagonal elements $\varepsilon^{\mathrm{free\ atom}}$ are computed neglecting the extra term $(r/(r^{\mathrm{wf}}))^2$ in eq 4, while the reference densities ρ_b^0 and ρ_a^0 of atoms a and b are calculated replacing $r^{\mathrm{wf}}_{a/b}$ with $r^{\mathrm{dens}}_{a/b}$ (called density compression radii). That is, in practice we use different compression radii for the atomic wave functions and atomic densities; for more details, see ref 20.

In eq 1, $\Delta q_a = q_a - q_a^0$ denotes the atomic net charge of atom a (negative Mulliken charge). γ_{ab} is a function describing the Coulomb interaction between two spherical charge distributions located at atoms a and b and has the form

$$\gamma_{ab} = 1/r_{ab} - S(U_a, U_b, r_{ab})\cdot h(U_a, U_b, r_{ab}) \quad (6)$$

For large interatomic distances r_{ab}, this formula basically reduces to the Coulomb interaction of two point charges. The short-ranged function $S\cdot h$ interpolates that long-ranged part with the proper on-site electron−electron interaction of the excess charge; i.e., at $r_{ab} = r_{aa} = 0$ the electron−electron interaction is determined by the atomic Hubbard parameter $\gamma_{aa} = U_a$. The function h is given by

$$h(U_a, U_b, r_{ab}) = \begin{cases} 1 & \text{if neither } a \text{ nor } b \text{ are hydrogen atoms} \\ \exp\left[-\left(\frac{U_a + U_b}{2}\right)^{\zeta} r_{ab}^2\right] & \text{if } a \text{ or } b \text{ or both are hydrogen atoms} \end{cases} \quad (7)$$

where a new parameter ζ is introduced to account for the peculiarity of hydrogen (details see refs 3, 7, and 8). The Hubbard parameters U_a are determined from an atomic DFT calculation as the first derivative of the highest occupied atomic orbital with respect to its occupation number, whereas ζ is fitted to reproduce the binding energy of the water dimer. The derivative of γ_{ab} with respect to the atomic charge q_a is given by Γ_{ab}, which is an analytical function that contains the Hubbard derivative with respect to the atomic charge as a parameter. The Hubbard derivative U^d is calculated as the derivative of the Hubbard parameter with respect to charge around the neutral atom.

Finally, the so-called repulsive energy E^{rep} is approximated as a sum of two-center potentials V_{ab}^{rep}, which are fitted to experimental or ab initio data as described below.

The basis set coefficients $c_{\mu i}$ can then be obtained by applying the variational principle to the total energy expression in eq 1 to yield the Kohn−Sham equations ($S_{\mu\nu} = \langle\phi_\mu|\phi_\nu\rangle$):

$$\sum_b \sum_{\nu\in b} c_{\nu i}(H_{\mu\nu} - \varepsilon_i S_{\mu\nu}) = 0 \quad \forall\, a, \mu \in a, i \quad (8)$$

$$H_{\mu\nu} = H^0_{\mu\nu} + S_{\mu\nu}\sum_c \Delta q_c\left(\frac{1}{2}(\gamma_{ac} + \gamma_{bc}) + \frac{1}{3}(\Delta q_a \Gamma_{ac} + \Delta q_b \Gamma_{bc}) + \frac{\Delta q_c}{6}(\Gamma_{ca} + \Gamma_{cb})\right) \quad \forall\, a, b, \mu \in a, \nu \in b \quad (9)$$

2.2. Fitting of Repulsive Potentials. To determine the repulsive potentials V_{ab}^{rep} in eq 1, a recently developed search protocol (eqs 10, 11) is applied, where the repulsive potential between atom types A and B is described by a spline function in the covalent binding region, i.e., fourth order polynomials which are defined to be continuously differentiable up to its third derivative

$$S_{ABi} = \sum_{k=0}^{4} s_{ABik}(r - r_{ABi})^k \quad (10)$$

where r_{ABi} is the division point of inverval i and s_{ABik} represents the unknowns. Note that the repulsive energy is a sum over all atom pairs and usually written as $(1/2)\Sigma_{ab}V_{ab}^{rep}$. A cutoff is introduced at which the function and its first three derivatives reach zero. The cutoff is typically chosen to lie between first and second neighbor distances of "common" molecules (e.g., the two hydrogens in methane are second neighbors; i.e., there is no covalent bond between the hydrogens) and greatly facilitates the fitting procedure. The remaining degrees of freedom for the spline function are then fitted to experimental or high-level ab initio values of atomization energies, geometries, reaction energetics, and vibrational stretching frequencies.

All conditions are formulated in terms of a linear equation system consisting of energy, force, reaction, and additional equations. For example, the atomization energy of a molecule is given by

$$E^{at} = -E^{tot} + \sum_a E_a^{atom} \quad (11)$$

where E_a^{atom} is the energy of atom a (its calculation is described in section 3.1) and can be rearranged after inserting eq 1 for the total energy E^{tot} as

$$E^{rep} = \frac{1}{2}\sum_{ab} V_{ab}^{rep}(r_{ab}) = -E^{at} - E^{el} + \sum_a E_a^{atom} \quad (12)$$

where $E^{el} = E^{H0} + E^{\gamma} + E^{\Gamma}$ (that is, the difference between the DFTB3 total energy and E^{rep}, see eq 1). Providing a reference atomization energy and a corresponding fixed molecular geometry, the right-hand side of that equation can be calculated, while the left-hand side can be expressed in terms of the unknowns s_{ABik} via eq 10. Similar equations can be derived for reaction energies and atomic forces. For the former, the geometry of the participating molecules as well as a reference energy must be provided. For the latter, a molecule at equilibrium is necessary where the total force acting on each atom in each spatial direction equals zero. Alternatively, a nonequilibrium structure can be chosen including corresponding reference forces (e.g., from high-level calculations). Additional equations are usually defined in the form $S''_{ABz}(r_z) = V_z$, the second derivative of the spline function at a particular interatomic distance r_z being assigned the value V_z. This value is chosen to improve the vibrational stretch frequencies.

Subsequently, the equation system is solved for the unknown spline coefficients s_{ABik} using a singular value decomposition which gives a least-squares "solution" for an overdetermined equation system and a numerically stable solution in case the equation system is underdetermined. Nevertheless, the overall quality of the potentials is sensitive to the choice of the division points for the spline function. Therefore, finding good division points is the most time-consuming step of the fitting procedure. Several attempts of automatizing this step have not provided results that improved over the empirical choice of division points; however, empirical search protocols quickly converge to good results.[10,11]

3. PARAMETRIZATION OF C, H, N, AND O

For the parametrization of DFTB, two distinct groups of parameters have to be optimized: the electronic parameters appearing in the first three contributions of the DFTB3 total energy (eq 1), which are two parameters r_a^{wf} and r_a^{dens} for each atom type [please note that the other two parameters, the Hubbard parameter U_a and its derivative U_a^d are calculated, i.e., they are not subject to optimization in this work] and the repulsive parameters, which are basically the two-center potentials V_{ab}^{rep}. In the current work, we have optimized both sets of parameters for generating the new DFTB3 parametrization 3OB.

The general procedure was to first create a preliminary set, starting with the electronic parameters used in the MIO set

dx.doi.org/10.1021/ct300849w | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

[which are $r_C^{wf} = 2.7$, $r_C^{wdens} = 7.0$, $r_H^{wf} = 3.0$, $r_H^{dens} = 2.5$, $r_N^{wf} = 2.2$, $r_N^{dens} = 11.0$, $r_O^{wf} = 2.3$, $r_O^{dens} = 9.0$] and fitting repulsive potentials that predict atomization energies, bond lengths, and vibrational stretching frequencies for the most simple systems H_2, CH_4, H_2O, and NH_3 as accurately as possible.

Afterward followed the optimization of the electronic parameters, i.e. changing the wave function and density compression (r_a^{wf} and r_a^{dens}), for all elements as well as ζ (see eq 7) to reproduce the following criteria until satifactory accuracy was reached: H−N−H and H−O−H bond angles in ammonia and water; vibrational bending frequencies of ammonia, water, and methane; heavy atom distances within the dimers $(H_2O)_2$, $(NH_3)_2$, and $H_2O{-}NH_3$; and the binding energy of the water dimer. Note that these criteria are not directly affected by the short-ranged repulsive potentials V_{ab}^{rep}; i.e., the optimization of the electronic parameters is independent from the optimization of the repulsive potentials. The main idea of reparametrizing the electronic parameters was 2-fold: first, we choose larger r_a^{wf} such that an increased Pauli repulsion causes larger noncovalent bond distances (e.g., hydrogen bond lengths), as mentioned in the Introduction. Second, larger r_a^{wf} and smaller r_a^{dens} lead to more transferable repulsive potentials, which allow the elimination of the overbinding which is present when the MIO parameter set is used.

Finally, the repulsive potentials were optimized to reproduce atomization energies, geometrical properties, and vibrational frequencies for the G2/97[21] test set. During the last step, further small changes within the electronic parameters were made to find repulsive potentials that vanish at the cutoff as smoothly as possible without generating large gradients close to the cutoff. We found that especially the density compression radii r_a^{dens} of C, N, and O were ideal to change for this purpose because they have a rather small influence on all other properties. In the following, the newly obtained parameters are presented. For a detailed description of every parameter, see ref 11.

3.1. Electronic Parameters. The electronic parameters are summarized in Table 1. l_{max} is the maximal angular momentum of an atom, and n_{max} is the maximal integer index that n runs to in eq 3. E^{spin} is the spin-polarization energy of one atom, i.e. the energy difference between the atomic ground state and a hypothetical spin-unpolarized atom (singlet state). It is used to calculate atomization energies as

Table 1. Overview of the New Electronic Parameters (in Atomic Units if Not Unitless), the Global Parameter $\zeta = 4.00$

parameter	H	C	N	O
l_{max}	0	1	1	1
n_{max}	2	2	2	2
α_0	0.50	0.50	0.50	0.50
α_1	1.00	1.14	1.21	1.26
α_2	2.00	2.62	2.90	3.17
α_3		6.00	7.00	8.00
r^{wf}(s,p)	3.0	3.3	2.8	2.5
r^{dens}	2.5	6.5	10.0	6.0
ε_s	−0.2386004	−0.5048917	−0.66	−0.8788325
ε_p		−0.1943551	−0.2607280	−0.3321317
E^{spin}	−0.0410614	−0.0454791	−0.1147656	−0.0557761
U	0.4195	0.3647	0.4309	0.4954
U^d	−0.1857	−0.1492	−0.1535	−0.1575

$$E^{at} = -E^{tot} + \sum_a E_a^{atom} \tag{13}$$

where a runs over all atoms of the molecule, and the atomic energies E_a^{atom} are determined by the atomic eigenvalues ε_i^a, their corresponding orbital occupation n_i, and the spin-polarization energy as

$$E_a^{atom} = \sum_i n_i \varepsilon_i^a + E_a^{spin} \tag{14}$$

All other parameters shown in Table 1 are discussed in the Theory section. The on-site energies $\varepsilon_{s/p}$, E^{spin}, the Hubbard U, and Hubbard derivative parameters U^d are calculated using the PBE exchange-correlation functional.[13] One exception is the ε_s of nitrogen where the calculated value has been shifted by about +0.022 hartree in order to balance the sp^2 and sp^3 hybridizations of nitrogen for the inversion barrier of ammonia and the dihedral angle of formamide and acetamide. For a more detailed discussion, see ref 11. Note that an orbital-resolved DFTB method has been discussed where angular-momentum-dependent Hubbard parameters become necessary (e.g., ref 22). Because the presented electronic and repulsive parameters are optimized using atomic Hubbard parameters, we recommend using only the non-orbital-resolved methodology. If a future extension of the 3OB parameter set to further elements adopts the orbital-resolved method, we recommend using the atomic Hubbard parameter for all angular momenta instead of distinct Hubbard parameters.

3.2. Repulsive Parameters for General Use. The repulsive potentials are fitted to several properties that are summarized in Table 2. Reference values are calculated from methods that are known to produce results close to experimental ones. These are B3LYP[14−16]/cc-pVTZ[23] optimized geometries of small neutral closed-shell molecules, atomization energies as calculated from the composite G3B3 method[24] (with exception of H_2, see below), and proton transfer barriers from MP2[25]/G3large[26] (structures and energetics). Furthermore, Table 2 shows some so-called additional equations. These are chosen in order to reproduce experimental vibrational frequencies of small molecules, as will be shown in section 4.

In our previous studies, it was found that a small number of spline division points leads to robust results.[10,11] For each spline interval, there is one effective degree of freedom; all other degrees of freedom are determined by the conditions that the spline function and its rst three derivatives are continuous. Thus, for fitting the H−H repulsive potential to the hydrogen molecule, we use three types of reference data, atomization energy, bond distance, and the vibrational frequency. It follows that three intervals (i.e., four division points) for the H−H repulsive potential are sufficient to exactly describe these data points.

We apply the same ideas for other potentials, e.g., the H−C potential. The difference now is that we have several H−C bonds with different chemical environments in our training set; thus the resulting potential interpolates between those. For most heavy element pairings, the number and position of division points is further refined to improve performance for different bond types such as single, double, triple, and aromatic bonds. In comparison to H−C, the interatomic distances are distributed over a much broader range. The refinement is an empirical process of moving or adding a spline interval to modify the interpolation for different interatomic distances and add more flexibility. More details on this empirical search are given in refs 10 and 11.

dx.doi.org/10.1021/ct300849w | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 2. Parameters Defining the Repulsive Potentials

Molecules of Training Set 1[a] (w^{eeq}, w^{feq} if Different from 1.0) and E^{at}					
dihydrogen	69.8	methanimine	438.5	methyl formate	785.9
methane	419.7	methylamine	580.5	dinitrogen	227.8
ammonia	296.8	trimethylamine	1159.2	hydrazine	436.1
water	231.9	carbonmonoxide (0.1, 1.0)	259.8	nitrous acid (0.1, 0.1)	311.3
ethyne (2.0, 1.0)	404.9	carbondioxide (0.1, 1.0)	390.6	nitromethane	601.1
ethene (2.0, 2.0)	563.0	formaldehyde	374.4	dioxygen (singlet)	91.4
ethane (2.0, 4.0)	711.4	methanol	512.1	hydrogen peroxide	267.5
benzene (1.5, 1.0)	1365.9	formic acid	500.9		
hydrogen cyanide	313.3	acetic acid	802.2		

proton transfer reactions[b]	r^{XX}	E^{bar}
$[H_2O{-}H{-}OH_2]^+$	2.5, 2.6, 2.7, 2.8	0.61, 2.36, 5.20, 8.85
$[HO{-}H{-}OH]^-$	2.5, 2.6, 2.7, 2.8	0.50, 2.30, 5.16, 8.81
$[H_3N{-}H{-}NH_3]^+$	2.6, 2.7, 2.8, 2.9	0.40, 1.94, 4.44, 7.70
$[H_2N{-}H{-}NH_2]^-$	2.5, 2.6, 2.7, 2.8	0.05, 1.44, 3.49, 6.26

potential	division points (au)	additional equations (au)
H−H	1.4, 1.6, 1.8, 2.0	$V''(1.404) = 0.423$
H−C	2.0, 2.5, 3.4, 3.5	$V''(2.043) = 0.375$
H−N	1.8, 2.5, 3.0, 3.5, 4.4	$V''(1.916) = 0.500$, $V''(2.362) = 0.150$, $V''(2.778) = 0.020$
H−O	1.7, 2.2, 2.7, 3.4, 3.7	$V''(1.816) = 0.620$
C−C	2.0, 2.8, 3.0, 4.0, 4.8	$V''(2.513) = 0.920$
C−N	2.0, 2.4, 2.6, 3.1, 3.8, 4.2	$V''(2.767) = 0.350$, $V''(2.166) = 1.700$
C−O	2.0, 2.9, 3.5, 4.6	$V''(2.683) = 0.370$
N−N	2.0, 3.3, 3.4, 4.8	$V''(2.062) = 2.400$, $V''(2.714) = 0.350$
N−O	1.9, 2.4, 2.6, 3.5, 3.6, 3.9	$V''(2.268) = 1.300$, $V''(2.589) = 0.900$
O−O	2.0, 2.8, 3.4, 4.4, 4.5	$V''(2.279) = 0.750$, $V''(2.744) = 0.195$

[a] E^{at} is the reference atomization energy in kcal/mol, w^{eeq} and w^{feq} are the weighting factors for energy and force equations in 1/au; reaction equations and additional equations are weighted with 1/au. [b] E^{bar} is the reference proton transfer barrier in kcal/mol for a heavy atom distance r^{XX} in Å.

A particular problem is the interpolation of the repulsive potential towards the cutoff. While the potential is often substantially repulsive in the binding region, we require it to be zero at the cutoff. This can cause large gradients. Interestingly, the density compression radii r_a^{dens} have a large influence on the electronic energy contributions of eq 1 (first three terms on the right-hand side). Varying them can make these terms less negative, which requires a less positive repulsive potential for which to compensate. An important effect of this is that the repulsive potentials become less steep and much more short ranged, which allows for a shorter cutoff and therefore for an elimination of the overbinding present in the MIO parameters. Interestingly, r_a^{dens} seems to have only a small influence on the other properties that are considered. That is, after fitting the repulsive potentials, a similar accuracy can be achieved for bond lengths or vibrational frequencies.

For fitting H−C, no information of stretched H−C bonds was used. At longer distances, the potential is just an interpolation between equilibrium values and the cutoff. However, for the H−N and H−O potential, we have explicitely included information of stretched systems via proton transfer energetics because of their great importance in biological systems; i.e., the energy-barrier for proton transfer model systems at different distances, as indicated in Table 2, were included in the fit.

In the particular case of the N−O potential, inclusion of nitrous acid in the fit set is necessary due to its two different N−O bond lengths of 1.178 and 1.391 Å (B3LYP/cc-pVTZ), which are also different from the one of nitromethane with 1.218 Å. Nitrous acid contains also a H−O bond. During the fit, this affects the H−O potential in an unfavorable manner, and therefore, small weighting factors for nitrous acid are introduced to diminish this effect. However, the impact on the N−O potential remains almost the same in comparison to using weighting factors of 1.0 due to the unique N−O distances in the fit set and an appropriate choice of division points.

There are two well-known deficiencies within DFTB2 and the MIO parametrization which led to work-arounds using special parametrizations of the repulsive potentials. These problems are not resolved within the new DFTB3 parametrization, and thus, these work-arounds have to be applied here as well:

- H−H-mod: When fitting the H−H repulsive potential to the G3B3 atomization energy of the H_2 molecule (109.8 kcal/mol), it becomes strongly attractive. This has been found to be problematic in molecular dynamics simulations because artificial H−H bonds form. Trying to adjust the density compression radius as mentioned above, however, cannot shift the potential to be repulsive without deteriorating the description of bond angles, vibrational bending frequencies, hydrogen binding, and structure and energetics of proton transfer barriers. To prevent this problem, a 50 kcal/mol underbinding is introduced for the H_2 molecule, and the repulsive potential is fitted to E^{at} = 69.8 kcal/mol as shown in Table 2.

  To be able to properly describe covalent H−H bonds, an alternative fit for the H−H repulsive potential has been carried out. As reference served the proper H_2 atomization energy of 109.8 kcal/mol as calculated by G3B3, the B3LYP/cc-pVTZ geometry of H_2, and the H−H additional equation from Table 2. The division points 1.4, 1.8, 2.5, and 2.6 bohr were used. Since in most applications the H−H covalent binding energy is not important but a

dx.doi.org/10.1021/ct300849w | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

reasonable description in a molecular dynamics simulation is, we use the shifted potential in the standard parameter set and provide the unshifted potential under the name H−H-mod. Thus, for describing the H_2 molecule, the H−H-mod potential should be used.

- H−N-mod: Nitrogen hybridization seems to pose a problem for minimal basis set methods like DFTB as well as for NDDO type semiempirical methods.[27] For deprotonation energies of sp^3 hybridized nitrogen containing systems, we find consistent errors of about 14 kcal/mol, whereas sp^1 and sp^2 systems are described with much smaller errors. Specifically for situations where proton affinities of sp^3 nitrogen are important, we provide a special parametrization H−N-mod, where we shift the repulsive potential by these 14 kcal/mol. This of course limits the transferability of DFTB because we cannot change the hybridization state during a reaction. Thus, H−N-mod should only be used in specific situations but is not an overall solution to the underlying problem. For a detailed discussion, see ref 3 and references therein. The H−N-mod potential is fitted solely to the geometry of the B3LYP/cc-pVTZ optimized ammonia molecule (force equation weight = 2.0), an atomization energy (with a weight of 1.0) of 338.8 kcal/mol (instead of the G3B3 value of 296.8), the additional equation $V^{rep\prime\prime}(r = 1.43) = -0.03$, and all the nitrogen related proton transfer barriers as listed in Table 2 using the spline division points 1.8, 2.5, 3.1, 3.5, and 4.4.

3.3. Repulsive Parameters for Improved Vibrational Stretch Frequencies. For many applications, vibrational frequencies are more important than energetics. While the overall performance on vibrational stretch frequencies seems reasonably good for the MIO parameters, there are some specific frequencies where they show large deviations. This has been discussed in refs 10, 28−32. Similar problems are found for the 3OB parameters. Trading in some accuracy for energetics, we have developed a set of repulsive potentials, called 3OB-f in the following, which improve specifically the C═C, C═N, and C═O stretch frequencies.

Starting from the 3OB parameter set, the C−C, C−N, and C−O potentials have been refitted separately using the following data; the geometries are again optimized using B3LYP/cc-pVTZ:

- C−C: ethyne (1, 5, 404.9), ethene (1, 5, 563.0), ethane (1, 10, 711.4), benzene (1, 5, 1365.9), acetic acid (1, 5, 802.2), methyl formate (1, 5, 785.9), where in parentheses the weight of the energy and force equations is given, respectively, as well as the reference atomization energy calculated from G3B3 in kcal/mol. Three additional equations are applied (notation see Table 2), $V''(2.260) = 1.30$, $V''(2.502) = 0.62$, and $V''(2.886) = 0.25$, and the division points are 2.0, 2.6, 3.0, 3.6, and 4.8.
- C−N: Hydrogen cyanide (1, 5, 313.3), methanimine (1, 5, 438.5), methylamine (1, 5, 580.5), nitromethane (1, 5, 601.1), trimethylamine (1, 5, 1159.2). Three additional equations are applied, $V''(2.166) = 1.55$, $V''(2.387) = 0.70$, and $V''(2.767) = 0.30$, and the division points are 2.0, 2.6, 3.0, 3.6, and 4.2.
- C−O: Carbon monoxide (0.1, 1, 259.8), carbon dioxide (0.1, 1, 390.6), methanol (1, 1, 512.1), formaldehyde (1, 1, 374.4), formic acid (1, 1, 500.9), acetic acid (1, 1, 802.2), methyl formate (1, 1, 785.9). Two additional equations are applied, $V''(2.266) = 1.20$ and $V''(2.683) = 0.37$, and the division points are 2.0, 2.9, 3.5, and 4.6.

4. BENCHMARKS AND DISCUSSION

In this section, we benchmark our new parameters 3OB and 3OB-f with respect to experimental and high level ab initio data. We further compare the new parameters to our old MIO parametrization as well as to standard generalized gradient approximated (GGA) and hybrid DFT and selected semiempirical methods (PDDG/PM3, OM3). Unless noted otherwise, energetics are given for geometries that are optimized at the respective level of theory. The different DFTB parameter sets are summarized in Table 3. The hydrogen molecule has always been calculated using the H−H-mod repulsive potential.

Table 3. Parameter Sets for DFTB2 and DFTB3 for the Following Benchmark Tests

name	description
DFTB2/MIO	DFTB2 and MIO parameters
DFTB3/MIO/calc	DFTB3, MIO parameters, $\zeta = 4.05$, U^d for C, H, N, O as in Table 1
DFTB3/MIO/fit	DFTB3, MIO parameters, $\zeta = 4.20$, U^d for C, H, N, O are −0.23, −0.16, −0.13, −0.19
DFTB3/3OB	DFTB3, 3OB parameters as developed in this work, new recommeded parameters for general use
DFTB3/3OB-f	DFTB3, 3OB-f parameters as developed in this work, new recommeded parameters for vibrational frequency studies

All DFTB calculations have been carried out using our in-house DFTB code; for G3B3, MP2, B3LYP, BLYP, and PBE calculations as well as results from semiempirical methods that were not reported in the literature, the Gaussian 09 software package[33] has been used.

4.1. Main Differences Between the MIO and 3OB Parametrizations. 3OB has been developed particularly for DFTB3. It surpasses the MIO parametrization in two ways: first, it improves noncovalent bonds by improving both bond distances and energies, and second, it drastically reduces the overbinding apparent in DFTB3/MIO. [Please note that the newly chosen electronic parameters improve total energy properties, i.e. properties related to total energies and their derivatives. How it performs for purely electronic properties which depend on the electronic structure (KS energies) like excitation energies has to be tested separately.]

For a short demonstration of the performance for nonbonding interactions, we chose the binding energies and geometries of the water dimer, the ammonia dimer, and the benzene dimer in a T-shaped conformation as presented in Table 4. DFTB3 has been augmented with the London-type dispersion correction for DFTB as developed earlier.[34] [In comparison to calculations without dispersion corrections, the distance between the centers of mass of the benzene rings decreases by about 0.3 Å, and the interaction energy increases by about 2 kcal/mol for DFTB3. For the water and ammonia dimers, the dispersion effect is much smaller: the binding interaction increases by 0.1 and 0.3 kcal/mol, respectively; the heavy atom distances differ by less than 0.003 Å, and therefore, dipersion correction is not included for these cases in Table 4.] While for DFTB2/MIO the binding energy of water and ammonia dimer is underestimated, it is improved applying DFTB3/MIO mainly due to the modified Coulomb interaction as introduced with DFTB3 (eqs 6 and 7). However, this causes the hydrogen bond length to shorten to

Table 4. Deviation of DFTB Variants to High-Level References for Intermolecular Distances r (Å) and Binding Energies E^{bdg} (kcal/mol)

system	property[a]	ref[b]	DFTB2/MIO	MIO/calc	MIO/fit	3OB
water dimer	r	2.912	−0.048	−0.106	−0.108	−0.039
	E^{bdg}	−5.0	+1.7	+0.1	+0.1	+0.4
ammonia dimer	r	3.213	−0.145	−0.156	−0.154	−0.006
(assymmetric)	E^{bdg}	−3.1	+1.2	+0.3	+0.4	+1.3
benzene dimer[c]	r	4.948	−0.148	−0.159	−0.158	−0.012
(T-shape)	E^{bdg}	−2.7	0.0	−0.1	−0.1	+0.4

[a] r corresponds to the heavy atom distance for the water and ammonia dimer and the distance between the centers of mass of the benzene rings for the benzene dimer. [b] Reference values are CCSD(T)/extrapolated to nearly complete basis for the water dimer,[35] MP2/aug-cc-pV5Z for the ammonia dimer,[36] DFT-D for geometry, and CCSD(T)/extrapolated to nearly complete basis set for binding energy of the benzene dimer.[37] [c] Dispersion correction as developed in ref 34 is included.

Table 5. Mean and Maximum Absolute Deviations for Our Modified G2/97 CHNO Test Set

property[a]	N^b	DFTB2/MIO	MIO/calc	3OB	3OB-f	PBE[c]	B3LYP[c]
E^{at} (kcal/mol)	65	45.6	46.9	5.2	10.2	21.4	3.6
E^{at}_{max} (kcal/mol)		134.2	129.5	35.1	53.5	52.0	13.7
E^{at} (kcal/mol)[d]	65	7.4	7.1	4.8	8.3	4.2	2.4
E^{at}_{max} (kcal/mol)[d]		77.7	66.6	27.6	53.7	22.4	11.0
r (Å)	236	0.014	0.013	0.008	0.010	0.013	0.006
r_{max} (Å)		0.055	0.066	0.037	0.059	0.036	0.015
a (deg)	196	0.9	1.0	0.8	0.9	0.4	0.2
a_{max} (deg)		6.9	7.0	6.0	6.8	2.5	1.3

[a] Atomization energies are compared to G3B3 results. Bond lengths r and bond angles a are compared to B3LYP/cc-pVTZ calculations; max stands for maximum absolute deviation. For details, see the Supporting Information. [b] Number of comparisons. [c] Basis set 6-31G(d). [d] Calculated using atomic electronic energies fitted to the 25 atomization energies of training set 1.

about 0.1 Å smaller than the high-level ab initio reference. By increasing the Pauli repulsion via the wave function compression radius of oxygen and slightly modifying the parameter ζ for the Coulomb interaction, the new 3OB parameters reveal both properties, heavy atom distance and interaction energy, to be close to the reference. For the ammonia dimer, we trade in some accuracy for the interaction energy in order to improve the intermolecular distance. In the case of the benzene dimer, the more diffuse basis set on carbon (increase of the wave function compression radius) results in a larger intermolecular distance, without losing much of the good performance for the interaction energy.

The overbinding in the MIO parametrization causes an overestimation of atomization energies as will be discussed in detail in the following subsection. However, the overbinding of different bond types, e.g., the C−C and the C−O bonds, is balanced, such that reaction energies are in good agreement with high-level ab initio calculations. Nevertheless, it has been shown that this balance is not achieved for the MIO sulfur[38] and phosphorus parameters[3] leading to errors on the order of 50 kcal/mol.[3,39] [The overbinding increases the parametrization effort in the sense that reactions have to be considered that compare all combinations of different bond types. Even more severe, it might happen that the repulsive potentials are too repulsive in the binding region such that they can only be interpolated to zero at the cutoff by introducing an undesirable large gradient. As a consequence, the development of the MIO phosphorus parameters focused on geometries and specific reaction types (phosphate hydrolysis reactions or proton affinities) rather than nonisodesmic reactions.]

In contrast, the 3OB parameters are directly fitted to reproduce atomization energies reducing the effort of the parametrization process and easing the prevention of over-binding. Therefore, we expect this strategy to be helpful also for the parametrization of further elements.

4.2. Atomization Energies and Geometries. Atomization energies and geometrical properties for a slightly modified G2/97[21] molecule set are compared to reference values in Table 5. Specifically, only neutral, closed-shell, C, H, N, or O containing molecules were taken from that test set. In addition, the following few molecules were added for the purpose of including further binding situations at distinctly different interatomic distances from what already appear in the standard set: singlet oxygen, methanimine, *trans*-diazene, and nitrous and nitric acid. More details are given in the Supporting Information.

As a reference for atomization energies, the G3B3 composite method was used excluding zero-point and thermal corrections. G3B3 has been shown to perform excellently for 299 energies (enthalpies of formation, ionization potentials, electron affinities, proton affinities) of an assessed G2/97 test set, showing a mean absolute deviation (MAD) from experiments of only 0.99 kcal/mol.[40]

The MIO parameters show large MADs for atomization energies resulting from the large overbinding; the mean signed error (MSE) for, e.g., DFTB3/MIO/calc is almost as large as its MAD. The overbinding is substantially reduced for the 3OB set, which reveals a MAD of 5.2 kcal/mol and a MSE of +3.2 kcal/mol. The largest deviations for 3OB are 35.1, 25.3, 18.0, and 16.8 kcal/mol, which correspond to the molecules cyanogen, nitrous oxide, carbon monoxide, and carbon dioxide, respectively. More systematically, we find that molecules that include NO bonds or C≡N triple bonds show the largest errors, an observation that has already been recognized for the MIO parametrization.[41,39] 3OB-f performs slightly worse for atomization energies, as this property was not the focus of its optimization.

dx.doi.org/10.1021/ct300849w | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 6. C−C, C−N, and C−O Vibrational Stretching Frequencies in cm^{-1}[a]

molecule	exptl[b]	BLYP[c]	B3LYP[c]	DFTB2/MIO	MIO/calc	MIO/fit	3OB	3OB-f
H_3C-CH_3	995	958	996	1131	1125	1124	1015	1041
$H_2C=CH_2$	1623	1639	1693	1824	1819	1818	1776	1636
HC≡CH	1974	2009	2072	2114	2114	2113	2018	1978
H_3C-OH	1033	1051	1046	1046	1046	1046	1044	1032
$H_2C=O$	1746	1748	1824	1876	1869	1867	1840	1756
C≡O	2170	2114	2212	2394	2395	2395	2167	2072
H_3C-NH_2	1130	1144	1053	1204	1192	1194	1084	1129
$H_2C=NH$	1638	1646	1713	1904	1902	1902	1834	1649
HC≡N	2089	2113	2201	2253	2254	2254	2065	2098
benzene (CC1,A_{1g})	992	983	1015	1157	1156	1156	1090	1010
benzene (CC2,E_{2g})	1596	1575	1637	1825	1823	1832	1751	1621
furane (CO1,A_1)	1066	1046	1087	1249	1248	1248	1197	1145
furane (CO2,B_2)	1180	1140	1199	1139	1136	1137	1309	1186
pyrrol (CN1,A_1)	1148	1135	1175	1383	1380	1381	1262	1163
pyrrol (CN2,B_2)	1425	1393	1450	1646	1638	1637	1616	1464
MAD		23	48	161	159	159	94	25

[a]CX1 for X = C, N, O are breathing modes. CX2 are stretching modes. For pyrrol, the NH bending mode contributes considerably. [b]Experimental values taken from ref 50. [c]Basis set cc-pVTZ.

Full density functional calculations have been carried out using PBE and B3LYP. We compare to small basis set 6-31G(d)[42] calculations, since they are often used for larger molecular systems, where DFTB finds its main application area. PBE reveals a large overbinding; the MSE with 20.6 kcal/mol is almost as high as the MAD of 21.4 kcal/mol. In contrast, B3LYP gives by far the best result with a MAD of only 3.6 kcal/mol and the largest deviations being 13.7 and 9.3 kcal/mol for water and hydrogen peroxide. Note that the MAD for B3LYP with a cc-pVTZ basis set further reduces to 2.5 kcal/mol while for PBE/cc-pVTZ the MAD is almost unchanged with 20.8 kcal/mol.

To improve accuracy for atomization energies, the atomic contributions in eq 11 can be fitted to experimental or high-level ab initio data. Similarly, this has been done in previous studies for DFT and semiempirical methods[43−45] as well as for SCC-DFTB[39] using experimental heats of formations. Here, we fit the atomic contributions E_a^{atom} (eq 11) to minimize the weighted MAD of the 25 atomization energies that have been used for fitting the repulsive potentials (see Table 2 for molecules and corresponding weights). Details can be found in the Supporting Information.

The results for the modified G2/97 test set using the fitted atomic contributions are presented in the third and fourth lines of Table 5. For 3OB, the results improve only marginally due to the fact that the atomization energies have already been included in the parameter fitting by adjusting the density compression radii and the repulsive potential. However, DFTB3/3OB still shows an overall improvement over DFTB3/MIO.

Bond lengths and angles are compared to B3LYP/cc-pVTZ optimized geometries which reproduce experimental geometries very well.[46−48] The description of geometries for 3OB slightly improves over the already excellent performance of the older parametrization. With the largest error of only 0.037 Å for bond lengths, it reaches the PBE/6-31G(d) performance and is only outperformed by B3LYP/6-31G(d). The trend for the MIO parametrization to overestimate CH bond lengths in aldehydes[39] persists also for 3OB, even though it is smaller in magnitude. The errors for the aldehyde C−H bonds are in the range of 0.035−0.040 Å for MIO/calc and lower than 0.025 Å for 3OB.

Compared to DFT, the bond angles show larger deviations; here, the largest errors are found for water and ammonia, where the bond angles are overestimated by 6.0° and 3.3°. For the $C_{bridge}C_{bridge}H$ angle in bicyclobutane and the HON angle in nitric acid, the errors are −5.5° and +5.1°. All further deviations are below 3°.

A well-known failure of the MIO parametrization is the planar description of H_2O_2,[28] which is unchanged for 3OB. In contrast, a CCSD(T)/aug-cc-pVQZ calculation reveals the dihedral angle to be 112.4°.[49]

4.3. Vibrational Stretch Frequencies. Table 6 shows the performance of DFTB for selected vibrational stretch frequencies in comparison with BLYP and B3LYP. The computed vibrational frequencies are harmonic and are not scaled, while the experimental data refer to fundamental frequencies. 3OB improves over MIO for almost all frequencies but still shows large deviations, e.g., for the stretching mode of the C=C and C=N double bond. As has been shown above, a careful DFTB parametrization can reach the accuracy of (small basis set) DFT methods for energies and geometries. However, we were not able to fit energies and vibrational frequencies to the same accuracy. Therefore, we decided to make a special parametrization, 3OB-f, which improves vibrational frequencies but loses accuracy for energetics.

4.4. Reaction Energies. In addition to G3B3 atomization energies, a comparison of reaction energies excluding zero-point vibrational and thermal corrections has been compiled in Table 7 consisting of hydrogenations and reactions with CH_4, NH_3, and H_2O. In all 49 reactions, only 32 molecules are involved.

The atomization energies are consistently overestimated by MIO, leading to MADs of over 40 kcal/mol, but the reaction energies show much lower overall deviations with only 8−9 kcal/mol. This shows that the large overbinding as apparent in the MIO parametrization is balanced between different bond types. The largest errors of 20 kcal/mol and more are found when NO bonds are involved. Furthermore, large errors are produced for reactions where large stochiometric coefficients appear.

The 3OB parameters were fitted to reproduce atomization energies; thus, if the errors for the atomization energies of all participating reactants are small, also the errors for the reaction energy must be small. For reactions that include molecules with large errors in atomization energies, we still find large errors in the reaction energies (if not canceled). The problematic cases are

Table 7. Deviation of DFTB for 49 Reaction Energies Compared to G3B3[a]

reaction	G3B3	DFTB2/ MIO	MIO/ calc	MIO/ fit	3OB	3OB-f	PBE[b]	B3LYP[b]	PBE[c]	B3LYP[c]
$H_3C-CH_3 + H_2 \rightarrow 2CH_4$	−18.2	+1.8	+1.0	+1.1	+1.5	−2.0	+0.3	−1.1	−0.4	−1.9
$H_2C=CH_2 + 2H_2 \rightarrow 2CH_4$	−56.8	−5.7	−6.6	−6.6	−1.6	−6.4	−5.5	−4.9	−2.8	−1.8
$HC\equiv CH + 3H_2 \rightarrow 2CH_4$	−105.1	−4.7	−5.7	−5.6	+2.7	+2.9	−13.4	−10.8	−6.8	−3.4
$C_6H_6 + 9H_2 \rightarrow 6CH_4$	−163.9	−7.0	−12.5	−12.4	−2.1	−31.7	−8.5	−13.2	−5.7	−9.3
$H_2N-NH_2 + H_2 \rightarrow 2NH_3$	−47.7	+4.0	−0.4	+0.3	−11.9	−11.9	+9.4	+7.6	+3.1	+1.5
$HN=NH + 2H_2 \rightarrow 2NH_3$	−78.6	−2.6	−10.1	−8.9	−13.8	−13.8	+12.5	+11.2	+1.2	+0.8
$N_2 + 3H_2 \rightarrow 2NH_3$	−36.3	+1.1	−7.5	−6.1	−0.4	−0.4	+5.1	+7.9	−6.5	−2.0
$HO-OH + H_2 \rightarrow 2H_2O$	−86.5	+0.6	−8.6	−8.6	−2.9	−2.9	+21.5	+17.9	+9.4	+6.1
$^1O_2 + 2H_2 \rightarrow 2H_2O$	−152.8	+7.4	−6.7	−6.5	+11.0	+11.0	+25.2	+21.3	+5.4	+1.9
$H_3C-NH_2 + H_2 \rightarrow CH_4 + NH_3$	−26.2	+1.1	−1.5	−1.2	+0.2	+9.3	+3.9	+2.6	+0.4	−0.8
$H_2C=NH + 2H_2 \rightarrow CH_4 + NH_3$	−58.4	−0.1	−4.2	−3.7	+1.9	+8.2	+2.9	+3.1	−1.2	−0.3
$HCN + 3H_2 \rightarrow CH_4 + NH_3$	−73.7	+5.6	+0.4	+1.0	+8.9	+21.0	−2.6	+0.1	−6.4	−2.6
pyridine + $9H_2 \rightarrow 5CH_4 + NH_3$	−170.2	−3.8	−12.6	−11.9	−0.1	−6.8	+2.1	−2.7	−3.8	−7.2
$H_3C-OH + H_2 \rightarrow CH_4 + H_2O$	−29.7	+0.4	−4.3	−4.1	+1.5	+1.6	+7.2	+6.2	+1.7	+1.0
$H_2C=O + 2H_2 \rightarrow CH_4 + H_2O$	−57.6	+0.3	−7.5	−7.4	+6.8	+5.5	+10.0	+10.0	+1.1	+2.0
$C\equiv O + 3H_2 \rightarrow CH_4 + H_2O$	−62.3	+10.3	+2.3	+2.4	+17.6	+18.0	+0.8	+4.9	−6.3	−1.0
$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O$	−53.6	+3.4	−12.8	−13.2	+16.0	+13.8	+23.4	+20.1	+6.3	+5.1
$HC(=O)OH + 3H_2 \rightarrow CH_4 + 2H_2O$	−53.1	−7.1	−19.4	−19.2	+2.8	+1.1	+22.3	+19.0	+6.7	+4.7
$N_2O + 4H_2 \rightarrow 2NH_3 + H_2O$	−115.5	+45.9	+28.3	+30.1	+16.0	+16.0	+36.6	+26.0	+14.2	+6.2
$HNO_2 + 3H_2 \rightarrow NH_3 + 2H_2O$	−119.9	+32.7	+16.5	+17.4	−14.1	−14.1	+40.8	+30.4	+16.5	+7.8
$HNO_3 + 4H_2 \rightarrow NH_3 + 3H_2O$	−165.7	+79.4	+49.3	+51.6	−0.9	−0.9	+61.6	+43.8	+28.4	+12.9
$H_2C=CH_2 + H_2 \rightarrow H_3C-CH_3$	−38.5	−7.5	−7.7	−7.7	−3.1	−4.4	−5.8	−3.8	−2.4	+0.1
$HC\equiv CH + H_2 \rightarrow H_2C=CH_2$	−48.3	+1.0	+0.9	+0.9	+4.3	+9.3	−7.9	−5.9	−4.0	−1.7
$C_6H_6 + 6H_2 \rightarrow 3H_3C-CH_3$	−109.2	−12.5	−15.5	−15.7	−6.7	−25.6	−9.5	−9.8	−4.4	−3.5
$HN=NH + H_2 \rightarrow H_2N-NH_2$	−30.9	−6.6	−9.7	−9.2	−2.0	−2.0	+3.1	+3.6	−1.9	−0.8
$N_2 + H_2 \rightarrow HN=NH$	42.3	+3.7	+2.6	+2.8	+13.4	+13.4	−7.4	−3.3	−7.7	−2.8
$^1O_2 + H_2 \rightarrow HO-OH$	−66.3	+6.8	+2.0	+2.1	+13.9	+13.9	+3.8	+3.4	−4.0	−4.2
$H_2C=NH + H_2 \rightarrow H_3C-NH_2$	−32.1	−1.2	−2.7	−2.4	+1.7	−1.0	−1.0	+0.5	−1.6	+0.4
$HCN + H_2 \rightarrow H_2C=NH$	−15.3	+5.7	+4.6	+4.7	+6.9	+12.7	−5.5	−2.9	−5.2	−2.3
$H_2C=O + H_2 \rightarrow H_3C-OH$	−27.9	−0.1	−3.3	−3.3	+5.3	+3.9	+2.8	+3.8	−0.6	+1.0
$HC(=O)OH + H_2 \rightarrow H_2C=O + H_2O$	4.5	−7.4	−11.9	−11.8	−4.0	−4.4	+12.3	+9.1	+5.5	+2.7
$HC(=O)OH + 2H_2 \rightarrow H_2C-OH + H_2O$	−23.4	−7.5	−15.1	−15.1	+1.3	−0.4	+15.2	+12.8	+5.0	+3.7
$HNO_3 + H_2 \rightarrow HNO_2 + H_2O$	−45.8	+46.6	+32.8	+34.2	+13.2	+13.2	+20.8	+13.4	+11.9	+5.1
$HN(CH_3)_2 + H_2 \rightarrow H_3C-NH_2 + CH_4$	−22.1	−1.5	−3.7	−3.5	−2.8	+6.4	+2.8	+1.2	−0.4	−1.9
$N(CH_3)_3 + H_2 \rightarrow HN(CH_3)_2 + CH_4$	−19.0	−3.2	−5.0	−4.9	−5.2	+4.1	+1.2	−0.7	−1.8	−3.6
$O(CH_3)_2 + H_2 \rightarrow H_3C-OH + CH_4$	−24.6	−1.4	−5.2	−5.1	−1.7	−1.5	+5.5	+4.4	+0.2	−0.9
$HC(=O)OCH_3 + H_2 \rightarrow HC(=O)OH + CH_4$	−24.9	−2.4	−5.2	−5.1	−1.2	−0.9	+3.4	+2.7	−1.3	−1.9
$HC(=O)OCH_3 + H_2 \rightarrow H_2C=O + H_3C-OH$	9.3	−10.2	−12.8	−12.7	−6.7	−6.9	+8.6	+5.6	+2.6	−0.2
$H_2C=CH_2 + CH_4 \rightarrow CH_3-CH_2-CH_3$	−22.6	−7.7	−7.0	−7.1	−2.3	−0.1	−5.2	−1.5	−1.1	+3.2
$HC\equiv CH + CH_4 \rightarrow CH_2=CH-CH_3$	−35.0	−0.9	−0.1	−0.2	+3.6	+12.5	−8.9	−4.9	−4.0	+0.4
$H_3C-NH_2 + CH_4 \rightarrow H_3C-CH_3 + NH_3$	−8.0	−0.7	−2.6	−2.4	−1.3	+11.3	+3.5	+3.7	+0.8	+1.2
$H_2C=NH + CH_4 \rightarrow H_2C=CH_2 + NH_3$	−1.6	+5.6	+2.4	+2.9	+3.5	+14.7	+8.4	+8.0	+1.6	+1.4
$HCN + CH_4 \rightarrow HC\equiv CH + NH_3$	31.4	+10.3	+6.1	+6.7	+6.1	+18.1	+10.8	+10.9	+0.4	+0.8
$H_3C-OH + CH_4 \rightarrow H_3C-CH_3 + H_2O$	−11.5	−1.4	−5.3	−5.2	−0.1	+3.6	+6.8	+7.4	+2.1	+2.9
$H_2C=O + CH_4 \rightarrow H_2C=CH_2 + H_2O$	−0.8	+6.0	−0.9	−0.8	+8.4	+12.0	+15.5	+14.9	+4.0	+3.7
$HC(=O)OH + CH_4 \rightarrow H_2C=O + H_2O$	13.2	−5.8	−9.8	−9.7	−2.3	+0.9	+11.0	+9.7	+4.8	+4.0
$H_3C-NH_2 + H_2O \rightarrow H_3C-OH + NH_3$	3.5	+0.7	+2.7	+2.9	−1.3	+7.7	−3.3	−3.6	−1.3	−1.7
$H_2C=NH + H_2O \rightarrow H_2C=O + NH_3$	−0.8	−0.4	+3.3	+3.7	−4.9	+2.7	−7.1	−6.9	−2.3	−2.3
$HCN + 2H_2O \rightarrow HC(=O)OH + NH_3$	−20.6	+12.7	+19.8	+20.2	+6.1	+19.8	−24.9	−18.9	−13.1	−7.3
MAD		8.2	8.3	8.6	5.5	8.7	11.0	9.0	4.7	3.0
MAX		79.4	49.5	51.6	17.6	31.7	61.6	43.8	28.4	12.9

[a]Energies are calculated at 0 K excluding zero point energy and thermal corrections. All numbers are given in kcal/mol. [b]Basis set 6-31G(d). [c]Basis set cc-pVTZ.

carbon monoxide, carbon dioxide, hydrazine, and transdiazene, as well as molecules with C≡N triple bonds or NO bonds. The latter, however, improve significantly in comparison to MIO.

The reaction set is also quite a challenge for PBE and B3LYP if the basis set 6-31G(d) is used but substantially improves when the larger cc-pVTZ basis is applied. Note that, different than for the atomization energies, this is also true for PBE, showing that a larger basis set results in a more consistent overbinding. It is interesting to note that for those reactions, where DFTB shows large errors, also the small basis set DFT calculations show large

dx.doi.org/10.1021/ct300849w | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

errors. On the other hand, these errors decrease substantially when going to a larger basis set. This may point to a crucial limit of DFTB's accuracy, which could be given by its minimal basis set. Reaction energies could therefore be properties which cannot be improved without extending the theoretical model.

4.5. Proton Affinities, Hydrogen Binding Energies, and Proton Transfer Barriers. Recently, we compiled tests for proton affinities, hydrogen binding energies,[8] and proton transfer barriers[3] and tested several DFTB3 variants. The results for the new parametrization are briefly summarized and presented in detail in the Supporting Information.

As reported in the earlier studies, the proton affinities of 23 oxygen containing molecules show quite different errors depending on the charge state at the DFTB2/MIO level, while DFTB3 gives more consistent results. The proton affinities are overestimated with DFTB3/MIO/calc, which we related to the overbinding in the O−H bond [since the MSE with +5.2 kcal/mol is similarly as large as the MAD of 5.5 kcal/mol. As a temporary solution, we used fitted Hubbard derivatives DFTB3/MIO/fit (as opposed to the methodological, more justified solution of calculating them from PBE, DFTB3/MIO/calc) yielding improved results and a MSE of close to zero]. We expected the DFTB3/3OB parametrization to remedy this problem, which is indeed the case as reflected in the small error of 3.7 kcal/mol. [The 3OB results are slightly worse than MIO/fit (MAD of 2.9 kcal/mol); therefore we also fitted the Hubbard derivatives for 3OB and came to similar results as with MIO/fit (MAD of 3.0 kcal/mol). However, we decided to stay with the calculated values to be methodologically more consistent. Furthermore, during the fit of the Hubbard derivatives, the U^d of carbon becomes very small for both, MIO and 3OB (−0.23 in comparison to the calculated value of −0.1492 au) which has been found to cause SCC convergence problems in a few specific cases of large systems.]

As mentioned in section 3.2, the standard H−N parameters show large deviations for proton affinities of sp^3-hybridized nitrogen systems. When using the H−N-mod parameters for these systems and the standard ones for sp^1 and sp^2, calling this combination NHmix, we find overall good results for the small compilation of nine proton affinities, the MADs being 2.9 kcal/mol for 3OB.

The hydrogen binding energies greatly improve when using DFTB3 instead of DFTB2 mainly due to the modification of the γ function (eq 6). However, DFTB3 shows a systematic error pattern: the binding energy for protonated water clusters is overestimated, while it is underestimated for the deprotonated water clusters. This trend is somewhat more pronounced for the new parametrization leading to slightly larger errors and is part of the renewed balance between hydrogen bond length and binding energy (see section 4.1). The average errors for 22 hydrogen binding energies are 2.7 and 3.8 kcal/mol for MIO/fit and 3OB.

The proton transfer barriers involve the following simple model systems: $[H_2O{-}H{-}OH_2]^+$, $[HO{-}H{-}OH]^-$, $[H_3N{-}H{-}NH_3]^+$, $[H_2N{-}H{-}NH_2]^-$, $[H_3N{-}H{-}OH_2]^+$, and $[H_2O{-}H{-}NH_3]^+$. Each system was tested for four different fixed heavy atom distances mimicking situations in larger systems such as proteins. We estimated the barrier geometry by moving the shared hydrogen atom on a straight line between the heavy atoms and calculating the energy for a large number of points. The geometry corresponding to the highest energy was used as a barrier structure, whereas for the minimum energy geometries, all atoms besides the two heavy atoms were allowed to move. It has been shown that DFTB3/MIO systematically improves over DFTB2/MIO for these tests.[3] DFTB3/3OB maintains this good performance, as shown for a few examples in Table 8. More details can be found in the Supporting Information.

Table 8. Proton Transfer Barrier in kcal/mol for a Fixed Distance (rXY) between the Heavy Atoms (X,Y ∈ {O,N}): Deviation of DFTB and DFT in Comparison to MP2/G3large[a]

barrier	rXY	MP2	DFTB2/MIO	MIO/calc	3OB	PBE[b]	B3LYP[b]
$[H_2O{-}H{-}H_2O]^+$	2.7	5.2	−1.1	−0.7	−0.8	−3.2	−1.5
$[OH{-}H{-}OH]^-$	2.7	5.2	−4.6	−0.2	−0.1	−2.7	−0.9
$[NH_3{-}H{-}NH_3]^+$	2.8	4.4	−2.4	−2.0	−2.0	−3.0	−1.6
$[NH_2{-}H{-}NH_2]^-$	2.7	3.5	−3.5	+1.5	+0.5	−2.1	−0.8
$[NH_3{-}H{-}H_2O]^{+c}$	3.1	35.1	−11.0	−8.9	−5.9	−7.8	−3.8
$[H_2O{-}H{-}NH_3]^{+c}$	3.1	6.7	+0.3	−0.6	−2.3	−4.5	−2.3
$[NH_2{-}H{-}OH]^{-c}$	3.0	18.6	−11.1	−3.9	−3.1	−5.1	−1.4
$[OH{-}H{-}NH_2]^{-c}$	3.0	11.6	−6.6	+8.4	+5.8	−4.9	−2.0

[a]Barriers are computed as described in the text at 0 K, and no zero-point energy correction has been included. [b]Basis set 6-31+G(d,p). [c]Barrier in comparison to the relaxed structure with the proton binding to the heavy atom that is written on the left-hand side of that proton.

4.6. Jorgensen Test Sets. Jorgensen and co-workers published a collection of experimental heats of formation, isomerization enthalpies, conformational energetics, and MP2/6-31G(d) geometries[39,51] showing that PDDG/PM3 is the most accurate semiempirical method in terms of energetics when compared to AM1 and PM3 and also more accurate than DFTB2. The test set is more comprehensive than the G2/97 set, so we used it to evaluate our new parametrization 3OB. [Note that small differences between DFTB2/MIO shown below and the results in ref 39 may stem from the use of slightly different convergence criteria. However, these differences do not change the overall picture.]

4.6.1. Heats of Formation. The heats of formation are calculated as described in ref 21. The enthalpies of formation for gaseous atoms at 0 K, $\Delta H_f^0(0K)$, and the heat capacity corrections $(H^{298} - H^0)$ for the atoms in their standard states are taken from the Active Thermochemical Tables.[52] The $\Delta H_f^0(0K) - (H^{298} - H^0)$ values for C, H, N, and O are 169.73, 50.62, 111.49, and 57.95 kcal/mol, respectively. The heat capacity corrections for molecular vibrations are estimated as

$$E^{vib} = R \sum_k \theta_k \left(\frac{1}{2} + \frac{1}{e^{\theta_k/T} - 1} \right) \text{ where } \theta_k = \frac{h\nu_k}{k_B} \quad (15)$$

using harmonic vibrational frequencies ν at the respective DFTB level. R is the molar gas constant, θ the vibrational temperature, T the temperature, h and k_B the Planck and the Boltzmann constants. 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).

In summary, the heat of formation at 298.15 K can be written as

dx.doi.org/10.1021/ct300849w | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 9. Mean and Maximum Absolute Deviations for the Jorgensen Test Set

property[a]	N^b	AM1[c]	PM3[c]	PDDG[c]	DFTB2/MIO	MIO/calc	MIO/fit	3OB
ΔH_f (kcal/mol)	622	6.8	4.4	3.2	89.9	92.2	92.3	6.1
ΔH_f^{max} (kcal/mol)		38.1	39.6	39.1	195.2	197.5	197.8	53.3
ΔH_f (kcal/mol)[d]	622				5.4	6.3	6.3	5.1
ΔH_f^{max} (kcal/mol)[d]					81.0	70.0	72.2	46.5
ΔH_f (kcal/mol)[e]	622			2.8[f]	5.7	5.5	5.5	3.7
ΔH_f^{max} (kcal/mol)[e]				33.1[f]	86.0	77.4	79.2	40.7
r (Å)	218	0.017	0.012	0.013	0.012	0.011	0.011	0.008
r_{max} (Å)		0.166	0.066	0.065	0.044	0.043	0.043	0.045
a (deg)	126	1.5	1.7	1.9	1.0	1.1	1.1	1.0
a_{max} (deg)		11.6	11.1	11.8	7.4	7.3	7.5	6.5
d (deg)	34	2.8	3.2	3.7	2.5	3.0	3.1	2.6
d_{max} (deg)		18.1	14.5	17.1	16.5	19.1	19.4	19.2
dipole moment (Debye)	47	0.23	0.25	0.23	0.39	0.34	0.35	0.37

[a]Heats of formation ΔH_f at 298.15 K, bond lengths r, bond angles a, and dihedral angles d; max stands for maximum absolute deviation. [b]Number of comparisons. [c]AM1, PM3, and PDDG/PM3 values from ref 39 unless noted otherwise. [d]Calculated using atomic electronic energies fitted to the 25 atomization energies of training set 1. [e]Calculated using Jorgensen's strategy and eisols fitted to all 622 experimental heats of formation. [f]Values are taken from a reparametrization of PDDG/PM3 to all 622 molecules as described in ref 53.

$$\Delta H_f^0 = E^{tot} - \sum_a E_a^{atom} + H^{contr} \quad (16)$$

where H^{contr} represents the molecular and atomic heat capacity corrections.

In Table 9, mean and maximal absolute deviations for 622 experimental heats of formation are shown. As expected from the results above, 3OB substantially improves over the MIO parametrization, although it does not reach the accuracy of PDDG/PM3.

Using the fitted atomic energies E_a^{atom} as described in section 4.2 in eq 16, the errors for the MIO parametrization are much lower. Also, 3OB benefits (see third and fourth lines of Table 9).

Jorgensen and co-workers have utilized eq 16 differently,[53] using only atomic heat capacity corrections for H^{contr}, while the E_a^{atom} are fitted, i.e. all molecular heat capacity corrections are included in the element-dependent E_a^{atom} (called eisol). Applying this correction and optimizing E_a^{atom} to yield the lowest MAD for all 622 heats of formation results in a further reduction of the error with the exception of DFTB2/MIO (see fifth and sixth lines of Table 9; the fitted atomic values are given in the Supporting Information) and reduces the MAD for 3OB from 6.1 kcal/mol for the standard approach to only 3.7 kcal/mol. Again, we find that PDDG/PM3 is superior with errors of only 3.2 kcal/mol,[39] after a refit of its parameters to all 622 molecules, even reaching 2.8 kcal/mol.[53]

The largest errors for 3OB are found again for molecules containing C≡N triple bonds, NO bonds, and also for peroxides, irrespective of the atomic energies applied. This is different for bicyclo-hydrocarbons, cyclopentanes, cyclohexanes, and highly branched alkanes that reveal relatively large errors when using the calculated atomic energies while they are significantly reduced by applying the fitted ones.

In summary, we recommend using the fitted atomic contributions including the molecular heat capacity corrections when the aim is to calculate heats of formation. Note, however, that for reactions or relative energies, atomic contributions cancel out, and no specific values for E_a^{atom} are needed.

4.6.2. Geometries and Dipole Moments. Geometrical data are compared to MP2/6-31G(d) results.[51] Similar to that for the G2/97 test set, DFTB performs very well for bond lengths, and no extraordinary outliers are detected, the maximal absolute deviation for 3OB being 0.045 Å. The largest errors are found for situations where the reference MP2/6-31G(d) and B3LYP/cc-pVTZ differ the most, that is, hydrazine and N_2, where the NN distances are 1.486 and 1.131 Å for MP2 and 1.436 and 1.091 Å for B3LYP. The same is found for the C≡N triple bonds, which are consistently calculated to be about 0.03 Å shorter with B3LYP than with MP2. The DFTB results are generally closer to the B3LYP results as DFTB is a DFT method and is fitted to B3LYP geometries.

DFTB bond and dihedral angles are in very good overall agreement with the MP2 results. For 3OB, there are only three out of 126 angles that deviate more than 4°, and three outliers are found for dihedral angles, namely, the CCCO dihedral angle of propanal skew that is overestimated by 19.2°, the CCOH angle in gauche ethanol that is underestimated by 12.9°, and the HNCO dihedral angle in urea that is overestimated by 11.7°. Further errors are within 10°. These errors are very similar for the MIO parametrization and were already reported earlier.[39]

Finally, 47 dipole moments have been compared to experimental values. Neither DFTB3 nor the new parametrization bring significant changes to the older parametrizations which have been examined in detail in refs 39 and 54. It should be noted that DFTB dipole moments are routinely computed from Mulliken charges, and refinement of charge calculation can improve related properties substantially .[54−56]

4.6.3. Isomerization Enthalpies and Conformational Energetics. Jorgensen et al. have also compiled tests for isomerization enthalpies and conformational energetics of small molecules. It was also shown that vibrational and thermal corrections for enthalpies are negligible for DFTB;[39] thus, only potential energy differences are compared to the experimental references. For 13 hydrocarbon and 11 CHO containing isomerizations, the results improve over the older parametrization with MADs of 3.8 and 3.3 kcal/mol for 3OB versus 4.8 and 5.7 kcal/mol for MIO/calc. Both parametrizations perform similar for CHN containing reactions, revealing a MAD of about 5 kcal/mol. PDDG/PM3 results from ref 39 show smaller average errors with 2.4, 3.1, and 2.8 kcal/mol for hydrocarbon, CHO, and CHN isomerizations but slightly larger errors for conformational energetics with a MAD of 1.8 kcal/mol versus 1.1 kcal/mol with both DFTB parametrizations. Details can be found in the Supporting Information.

dx.doi.org/10.1021/ct300849w | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 10. Mean Absolute Deviations (MADs) in kcal/mol for the GMTKN24-hcno Test Sets

set	N^a	OM3[b]	OM3-D[b]	DFTB2/MIO	MIO/calc	MIO/fit	3OB	3OB-D
ACONF	15	0.9	0.3	0.2	0.2	0.2	0.3	0.5
BH76	38	8.7	9.0	15.9	16.4	16.3	13.2	13.3
BH76RC	17	6.2	6.3	10.0	10.5	10.5	7.2	7.2
BHPERI	22	8.8	7.5	7.3	7.7	7.7	4.3	4.2
DARC	14	4.9	8.3	3.6	3.4	3.4	8.9	7.4
DC9	6	13.2	12.3	15.2	15.2	15.0	14.8	14.0
G21EA	11	9.9	9.9	10.3	18.3	31.0	18.6	18.6
G21IP	13	12.7	12.7	8.9	15.9	16.5	15.3	15.3
G2RC	12	4.5	4.0	10.3	9.6	9.5	9.3	9.2
IDISP	6	6.7	8.4	2.9	2.9	2.9	8.5	3.4
ISO34	34	4.4	4.4	4.7	4.9	4.9	3.6	3.5
MB08−165	21	21.3	21.9	13.2	15.7	15.6	14.5	14.5
O3ADD6	6	11.0	11.1	7.5	7.7	7.7	7.1	7.0
PA	8	11.9	11.7	14.6	7.0	6.8	6.7	6.7
PCONF	10	1.3	2.1	1.6	2.1	2.2	1.7	1.1
RSE43	28	5.2	5.0	7.0	5.9	6.0	5.0	5.0
S22	22	3.6	1.1	3.5	2.9	2.9	3.4	2.0
SCONF	17	1.3	1.4	2.1	1.6	1.6	2.0	2.1
SIE11	4	5.0	5.5	16.0	16.8	16.6	13.4	13.6
W4−08woMR	39	11.8	11.7	17.9	18.4	18.3	6.4	6.4
WATER27	27	9.2	7.8	22.9	5.7	5.6	7.9	6.1
OVMAD		7.9	7.7	10.1	9.4	9.7	7.6	7.2
WTMAD		6.4	6.2	7.9	7.3	7.5	6.4	5.7

[a]Number of comparisons. [b]Data taken from ref 59.

4.7. GMTKN24. GMTKN24 and GMTKN30 are comprehensive databases for general main group thermochemistry, kinetics, and noncovalent interactions that have recently been compiled to evaluate DFT methods.[57,58] They consist of 24/30 different, chemically relevant subsets based on theoretical or experimental reference values. A comparison between methods is carried out by calculating single-point energies at geometries from high-level ab initio methods. Another study used a subset of this database, called GMTKN24-hcno, that includes only elements C, H, N, and O to benchmark the most common and widely used semiempirical methods.[59] A careful inspection was carried out showing that this subset is well suited for evaluating semiempirical methods. It was found that among AM1,[60] PM3,[61] PDDG,[18] PM6,[62] DFTB2, and the OMx methods,[19,63−65] OM3 shows the best overall performance, being comparable to GGA-DFT calculations using the PBE functional with a triple-ζ basis, but is outperformed by B3LYP.

The GMTKN24-hcno test set includes 21 subsets, which include relative energies of alkanes (ACONF), tripeptides (PCONF), sugar conformers (SCONF), noncovalent interactions (IDISP, S22, WATER27), reaction energies of small molecules (G2RC, BH76RC, part of O3ADD6) and Diels−Alder reactions (DARC), isomerization (ISO34) and atomization energies (W4−08), electron affinities (G21EA), ionization potentials (G21IP), proton affinities (PA), and radical stabilization energies (RSE43). Important are also barrier heights as they reveal the performance for nonequilibrium structures and are represented in the subsets for substitution and association reactions (BH76), pericyclic reactions (BHPERI), and ozone reactions (part of O3ADD6). Furthermore, there are difficult cases for DFT (DC9, SIE11) and a set with randomly generated artificial molecules aiming for a benchmark of the overall transferability and robustness in a broad variety of chemical environments (MB08−165). For more details, see the original publication[57] and its Supporting Information.

To follow up on the work done by Korth and Thiel,[59] we evaluate our new 3OB parameters and compare the results to the MIO parametrization and OM3 as the best-performing semiempirical method. We use the same setup, i.e., from all data points of the full GMTKN24 set containing elements C, H, N, and/or O; three further points are excluded. These are O_3 and C_2 of atomization energy subset W4-08 because of their multireference character, the set then being called W4-08woMR (note, however, that O_3 is included within the ozone reaction energies and barriers subset O3ADD6), and the C_{20} cage/bowl isomerization. The latter is an exceptionally large outlier for semiempirical methods and also poses a problem for high-level ab initio methods.[59] For OM3, for example, the error is 206.9 kcal/mol, which would add about 0.56 kcal/mol to the overall mean absolute deviation (OVMAD). For DFTB3/3OB, the error is smaller but still large with −30.9 kcal/mol.

The standard DFTB3 formalism is applied including several extensions. Note that besides using DFTB3 instead of DFTB2, the following computational details differ from the protocol used in ref 59:

- Use of H−H-mod: As explained in detail in section 3.2, the 3OB H−H repulsive potential gives large errors for the energy of H_2, but the special parametrization H−H-mod resolves that problem. The H_2 molecule is present in 39 of 370 data points of the database (often with a high stoichiometric coefficient); therefore a large error for that molecule has a tremendous impact on the OVMAD. Applying this single modification, the DFTB3/3OB OVMAD drops by almost 8 kcal/mol, mostly affecting the performance for the MB08-165 and G2RC subset, but also IDISP, BH76, BH76RC and W4-08woMR. This strong dependence on one molecule shows a limitation of this database. We believe that the unfavorable performance for other semiempirical methods such as PDDG and PM6 stems, to a significant extent, from this problem as

dx.doi.org/10.1021/ct300849w | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

well, as the errors for the H_2 atomization energy are larger than 20 kcal/mol, while the overall performance on atomization energies/heats of formation is known to be one of the strengths of these methods.[39,41,62] Note that for DFTB2/MIO and DFTB3/MIO, there are also H−H-mod parameters available from www.dftb.org which were used in the present study.

- Use of spin-polarization corrections: For calculating atomic energies, eq 14 was applied, which includes the atomic spin-polarization energy E^{spin} as calculated from the PBE functional. For open-shell systems other than atoms, the spin-polarization formalism from Köhler and co-workers was invoked,[66,67] which is available within DFTB+[68] and our in-house software. With this modification, we are able to calculate triplet and quartet states and also to more adequately describe doublet states. Thus, all 370 data points of the GMTKN24-hcno set are included in our comparisons in contrast with the earlier study[59] that excluded triplets and quartets for DFTB2. Note that for MIO, spin-polarization constants and E^{spin} as calculated from PBE were used. Details are provided in the Supporting Information.
- Use of H−N-mod: For calculating proton affinities of sp^3 hybridized nitrogen species, we have developed the H−N-mod parameter set (see section 3.2). Within GMTKN24-hcno, this applies only to one data point, the proton affinity of ammonia where the error is reduced by about 10 kcal/mol, and thus is of minor importance for the OVMAD.

The results are shown in Table 10. DFTB3/MIO performs worse than DFTB2/MIO for ionization potentials and electron affinities (G21IP, G21EA) but improves for the hydrogen binding energies (WATER27) and proton affinities. The MAD for PA is still considerably high, which is due to the fact that mostly proton affinities of hydrocarbons are included (six out of eight data points). This may indicate a general problem with proton affinities in semiempirical methods, which may be due to the minimal basis set applied.

The new parametrization 3OB further improves over MIO in most cases; however, it is significantly worse for DARC and IDISP. While the large error for IDISP is reduced when applying dispersion (3OB-D) as this is of particular importance for this subset, the error for DARC is inherited from the DFT origin of DFTB. As has been pointed out earlier, Diels−Alder reactions (which DARC is composed of) pose a problem for most density functionals due to the tendency of overestimating nonbonded repulsion of bridgehead carbons of the Diels−Alder products.[69]

With MIO, this error is masked within the overbinding tendency; i.e., while the simplest Diels−Alder reaction of ethene and butadiene gives an error of almost −13 kcal/mol, it is much less for the larger bicyclic variants which represent the majority of data points. In contrast, 3OB gives a very good reaction energy for ethene and butadiene with an error of only −2 kcal/mol—an effect of the removed overbinding—but therefore yields errors of around +10 kcal/mol for the bicyclic Diels−Alder reactions leading to an overall higher MAD for that subset.

As already mentioned, we have also carried out an evaluation including London-type dispersion corrections as available in DFTB[34] (denoted as 3OB-D in Table 10). As one would expect, this significantly lowers the error for the dispersion related subsets IDISP and S22 but also is important for DARC and the larger water clusters in WATER27.

In comparison to the best-performing semiempirical method OM3, DFTB3/3OB gives a very similar OVMAD. The weighting of the different subsets summarized in the weighted mean absolute deviation (WTMAD) as proposed in the original GMTKN24 publication suggests the same conclusion. An interpretation of the MADs of each subset is not easy, as there might be a few outliers or over-represented molecules that dominate the MAD. Nevertheless, it seems to support earlier findings that OM3 is superior for electron affinities, ionization potentials, and cases where DFT fails (SIE, DARC), while strengths of DFTB3/3OB are hydrogen binding energies (WATER27) and proton affinities (PA). An encouraging fact pointed out by Thiel and Korth[59] is that OM3—and with the present study also DFTB3/3OB—seems to lose only little accuracy in comparison to the computationally much more demanding PBE density functional calculations. The OVMAD for PBE with the TZVP basis set[70] was reported to be 6.6 kcal/mol; for B3LYP/TZVP, the OVMAD is significantly lower with 4.8 kcal/mol.[59] Using the double-ζ basis set 6-31+G(d,p) gives similar results; the OVMADs are 6.4 and 4.3 kcal/mol for PBE and B3LYP. For details, see Supporting Information.

4.8. Crystalline Systems. The 3OB parameters were optimized for organic and biomolecules; however, DFTB is also frequently used for crystalline systems (for a review, see e.g. the recent issue of *Phys. Status Solidi B*, ref 71). While a thorough evaluation for these systems including a test for a broad range of electronic properties is beyond the scope of this work, we have performed a few calculations on the diamond and graphite modification of carbon. We have scanned the total energy for different lengths of the unit vectors of the experimentally found lattices to find the minimum energy structure. In Table 11, a comparison of the optimized lattice parameters and the cohesive energy to experimental values are presented. The cohesive energy per atom is defined in analogy to the atomization energy of eq 11 as

$$E^{coh} = -\frac{E^{tot}}{n} + E^{atom} \tag{17}$$

where n is the number of carbon atoms in the unit cell. For graphite, we have applied the dispersion correction[34] because otherwise the graphite sheets dissociate. Note that within diamond and graphite there is no charge transfer between the carbon atoms, and therefore the non-self-consistent version of DFTB, DFTB2, and DFTB3 give the same results.

Table 11 shows that geometric properties are in good agreement with experimental values,[72] with a slight disadvantage for the 3OB parameter set. Consistent with the increased Pauli repulsion, the lattice constants increase in comparison to MIO.

Table 11. Cohesive Energies in eV/atom, Nearest Neighbor Distances rCC, and Lattice Parameters (a, c) in Å for Diamond and Graphite in Comparison to Experimental Values

	diamond			graphite			
	E^{coh}	rCC	a	E^{coh}	rCC	a	c
exptl[a]	7.35	1.545	3.567	7.37	1.421	2.461	6.709
PBC	6.86	1.587	3.665	5.55	1.410	2.443	6.645
MIO	6.70	1.587	3.666	5.45	1.410	2.442	6.635
3OB	6.00	1.609	3.716	5.12	1.406	2.436	6.945

[a]Experimental values taken from refs 72 and 73.

dx.doi.org/10.1021/ct300849w | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 12. Atomization Energies in kcal/mol: Deviation to G3B3 for Small Species with Different Charge States

molecule	G3B3	MP2[a]	B3LYP[b]	BLYP[b]	BLYP[c]	BLYP[d]	DFTB2/MIO	MIO/calc	MIO/fit	3OB
H^+	−314.4	+0.8	−0.8	+2.1	3.5	3.5	+12.3	−7.1	−4.4	−12.1
H_2	109.8	−7.4	+0.3	−0.5	−0.4	−0.4	+3.4	+3.4	+3.4	+0.0
H^-	11.4	−5.3	+9.6	+8.1	−40.3	−40.3	−14.0	+5.4	+2.7	+0.4
CH_3^+	81.6	−7.2	−0.3	−0.6	3.1	−0.8	−3.6	−10.0	−10.6	−24.0
CH_4	419.7	−12.4	+0.9	−3.1	−2.2	−4.1	+17.6	+18.6	+18.6	−0.2
CH_3^-	305.5	−9.8	+6.1	+3.7	−33.1	−0.9	+16.1	+52.3	+282.2	+42.1
NH_4^+	194.7	−11.1	+2.0	+3.8	2.2	−0.1	+9.7	+4.5	+5.1	−22.0
NH_3	296.8	−10.7	+3.9	+4.7	−5.7	−1.7	+21.7	+26.0	+25.3	+4.4
NH_2^-	197.3	−9.1	+7.6	+10.5	−38.2	1.4	−1.1	+38.3	+29.9	+24.9
H_3O^+	88.7	−2.7	−2.4	+0.5	−3.7	−7.5	+27.6	+21.8	+22.4	−10.2
H_2O	231.9	−2.6	−1.3	+0.6	−11.8	−7.2	+15.6	+22.6	+22.5	+0.5
OH^-	147.9	−1.8	+2.6	+6.6	−49.4	−0.9	−13.1	+22.2	+28.7	+10.8

[a]Basis set G3large. [b]Basis set aug-cc-pVTZ. [c]Basis set 6-31G(d). [d]Basis set 6-31+G(d).

3OB underestimates experimental cohesive energies[73] significantly and also gives a wrong qualitative picture of the stability between diamond and graphite. MIO improves with respect to the absolute errors but still predicts a wrong relative stability. There is another parameter set called PBC that mainly focused on a proper description of SiC solids.[74] For the test cases given here, PBC reveals quite similar characteristics to those of MIO. In summary, while 3OB is expected to give reasonable results for geometric data in crystalline systems, a careful evaluation of energetic properties seems necessary when considering a specific application. Perhaps a more rigorous parametrization focusing on solids will be unavoidable.

4.9. Limitations of Methodology and Parametrization. As discussed in the previous sections and publications mentioned earlier, DFTB3 improves over DFTB2, and the new parametrization 3OB supersedes the MIO set. Nevertheless, severe shortcomings remain for ionic species. Table 12 shows atomization energies of small molecules in neutral and ionic form. For ions, the atomization energies are calculated similarly as for neutral systems, as additional electrons do not contribute energetically. This is a quite challenging test as is obvious from the large errors when comparing MP2, B3LYP, and BLYP to the high-level G3B3 results; however, it reveals some trends for DFTB3/3OB. For neutral systems, the error for DFTB3/3OB is small because of the reduced overbinding tendency in comparison to MIO; for cationic species, the atomization energy is underestimated by 10 kcal/mol and more, while it is overestimated for anionic species. For H^+, the error is about 12 kcal/mol; i.e., when calculating proton affinities this error must be compensated for by the ionic species. That is indeed true for hydroxide and hydronium, which deviate about 10 kcal/mol from G3B3. For the nitrogen analogs, the error is larger than 20 kcal/mol, characterizing the not understood problem of sp^3 nitrogen. When using the H−N-mod potential, the atomization energy errors for NH_4^+, NH_3, and NH_2^- are +16, +33, and +44 kcal/mol, thus being larger, but the potential is more balanced in the sense that errors are smaller for proton affinities; i.e., the deviation for proton affinities of NH_2^- and NH_3 are about −5 and +1 kcal/mol, respectively. For the methyl anion, the third-order formalism breaks down, which causes the Mulliken charge on carbon to be too large. Consequently, the charge-fluctuation contribution to the energy is overestimated. This effect is most pronounced for MIO/fit, because of the significantly smaller Hubbard derivative of carbon in comparison to the calculated one.

This analysis suggests methodological extensions for decreasing atomization energy errors of ionic species, for example increasing the flexibility of the model by introducing two different Hubbard parameters per element, one for positively charged and one for negatively charged atoms based on ionization potentials and electron affinities, or by including higher order terms. While the basis set effects from BLYP results in Table 12 indicate different trends, a further examination on the underlying physics concerning basis set and the monopole approximation within DFTB3 will be necessary.

5. CONCLUSION

Over the years, the DFTB model has been extended continuously, creating an increased flexibility without an excessive increase in the number of free parameters, which have to be optimized. For DFTB, there are the two parameters for wave function and density confinement, r^{wf} by r^{dens}, as well as the parameters entering the repulsive potentials V_{ab}^{rep}. r_a^{wf} defines the minimal basis set, i.e., it can be compared to the exponents in usual SCF basis sets, while r_a^{dens} is a real new parameter which has to be optimized. However, this parameter does not influence molecular properties dramatically; i.e., it cannot be used to tune the method's performance in a wide range. Nonetheless, slight adjustments of these electronic parameters can have some impact on the model performance, as shown in this work, which is concerned with the geometries for nonbonded systems [please note that bond distances changed only within 0.1 Å; this shows the impact that the choice of these parameters typically has] and overbinding of molecules. The repulsive energy contribution, V_{ab}^{rep}, on the other hand, is crucial for the performance of DFTB for total energies and properties dependent on their derivatives. However, this holds only for direct bonding interactions since these interactions have a range of 2−3 Å and do not influence angular properties. V_{ab}^{rep} is therefore the key to optimizing bond energies, bond distances, and vibrational stretch frequencies, and recent developments to improve the fitting tools have been shown to successfully increase the performance of DFTB in this respect.[10]

DFTB2 results from a second order expansion of DFT, however, contacting the perturbation to an interaction point charge. By doing this, the complex interaction of the electron densities has to be mapped onto a scaled Coulomb interaction between atomic point charges. This introduces a new parameter, the Hubbard U_a, which is calculated from DFT and therefore is not a free parameter to optimize in order to improve the model performance. On the other hand, the Coulomb scaling law

implicit in function γ_{ab} in eq 1 assumes an inverse relationship between the chemical hardness of an atom modeled by U_a and the size of that atom.[7,20] This relation holds very well within one row of the periodic table but completely fails for interactions of atoms from different rows, in particular for interactions with hydrogen. Therefore, a different scaling law has been proposed, which includes one additional parameter which has to be optimized.

DFTB3 extends the DFT expansion to third order and includes the modified scaling law for γ_{ab}. The monopole approximation leads to an additional parameter U_a^d, the derivative of U_a, which again is calculated from DFT and therefore is not subject to optimization.

Up to now, DFTB3 has been applied in combination with a parametrization of V_{ab}^{rep} called MIO, which originally was developed for DFTB2. DFTB3/MIO has been shown to extend the flexibility of the method; i.e., it greatly improves the description of hydrogen bonded and charged molecules, which does not go back to solely expanding the free parameter space but to physically well motivated extensions of the formalism.

We have presented a new parameter set called 3OB for carbon, hydrogen, nitrogen, and oxygen based on DFTB3 that is available at www.dftb.org. In comparison to the older MIO parametrization, 3OB improves the description of the geometry for nonbonded interactions as, e.g., the hydrogen bond length in the water dimer while introducing only a minimal error for the respective binding energy. Furthermore, the new parametrization removes the overbinding for almost all bond types, leading to an adequate description of atomization energies. Exceptions are found only for the C≡N triple bond and specific cases such as carbon dioxide and carbon monoxide that overbind by about 10 kcal/mol.

Extensive benchmark tests have been carried out for a variety of energetic and geometric properties showing an overall improvement of the new parametrization. We have shown that DFTB3 can reach the accuracy of or even outperform DFT methods with small basis sets (double-ζ + polarization: DZP). But the brute force optimization also shows the limitations, which are inherent to the simplicity of the DFTB3 model:

- While the overall performance for atomization energies is quite good, a closer look at reaction energies indicates the limits of the method: Several reactions, which have errors of more than 10 kcal/mol using DZP-DFT as well, can be well described when going to DFT with larger basis sets. This may indicate one of the limitations of DFTB3 using a minimal basis.
- DFTB3 is derived from DFT; therefore it clearly suffers from the well-known DFT deficits. The DFT self-interaction error shows up in the DC9, SIE11, and DARC subsets of the GMTKN24 test suite. Related problems show up for excited states when using a TD-DFTB implementation.[75,76] Current work is therefore concerned with the implementations of self-interaction corrected functionals or long-range extended ones.[77,78]
- DFTB3 clearly reaches the accuracy of SCF-DFT with DZP basis sets, however, not for all properties simultaneously. In particular energies and vibrational frequencies seem to be in conflict. Vibrational stretch frequencies of C=C, C=N, and C=O double bonds deviate by 100 or more wavenumbers. Therefore, a compatible special parametrization for these repulsive potentials called 3OB-f has been developed which improves these stretch frequencies while maintaining an excellent description of geometries though suffering a loss of accuracy for energetics.
- To compute the binding energy of the H−H covalent bond, a special (but compatible) parametrization H−H-mod becomes necessary. For other uses, the standard parameters are recommended because they prevent artificial H−H bond formation as could possibly be happening during a molecular dynamics simulation using H−H-mod.
- Proton affinities of oxygen turned out to be excellently reproduced; however, those of carbon (PA set in GMTKN24 test suite) and nitrogen pose a problem. For proton affinities of sp^3 hybridized nitrogen species, consistent errors of about 14 kcal/mol are detected. This problem again seems to be related to the minimal basis description, since the electronic structure seems to require the use of d orbitals. As a temporary workaround, we have shifted the H−N repulsive potential in a special parametrization H−N-mod, reducing this error significantly. As a result we recommend the use of H−N-mod only for this specific property.
- Atomization energies of ionic species are poorly described. A consistent trend has been detected. Atomization energies for cations are underestimated, while those for anions are overestimated. This could be related to the monopole description in the higher order terms, or the simple use of one U for left and right derivatives in the DFTB3 formalism.

Despite these deficiencies, we have briefly discussed that DFTB3/3OB can compete with and complements classical semiempirical methods. As expected, we have found that PBE and B3LYP generally perform better than DFTB3/3OB. However, when using small basis sets for DFT calculations, this trend is often reversed. Reaction and atomization energies of small molecules are more accurately described by DFTB3/3OB. As we have shown in a previous study, this holds also true for cases where a large basis set superposition error is expected because DFTB3 does not exhibit this systematic error.[3] For practical applications, the use of popular density functionals with an "affordable" basis set might thus not be the best choice. This is especially noteworthy, as even with small basis sets, PBE and B3LYP require at least 250 times more computation time.[79]

As a result of this study, we present a new parameter set for CHNO containing molecules called 3OB. The expertise gained during the parametrization process of DFTB3 in combination with the semiautomatized optimization protocol and the good performance of the results motivates the parametrization of further elements. Additional methodological developments that broaden the applicability of DFTB3 have been suggested[22,78,80] and carried out[55,56] recently. In case a future reparametrization becomes necessary, the presented strategies and setups will probably be helpful.

■ ASSOCIATED CONTENT

Supporting Information

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

dx.doi.org/10.1021/ct300849w | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

AUTHOR INFORMATION

Corresponding Author

*E-mail: marcus.elstner@kit.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Prof. W. L. Jorgensen for providing us with compilations and geometries of earlier studies. We appreciate fruitful discussions with Dr. Bálint Aradi and Prof. Qiang Cui. This work was supported by Deutsche Forschungsgemeinschaft (Project DFG EL 206/11-1).

REFERENCES

(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) Seifert, G. *J. Phys. Chem. A* **2007**, *111*, 5609−5613.

(3) Gaus, M.; Cui, Q.; Elstner, M. *J. Chem. Theory Comput.* **2011**, *7*, 931−948.

(4) Seifert, G.; Eschrig, H. *Phys. Status Solidi B* **1985**, *127*, 573−585.

(5) Seifert, G.; Eschrig, H.; Bieger, W. *Z. Phys. Chem. (Leipzig)* **1986**, *267*, 529−539.

(6) Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. *Phys. Rev. B* **1995**, *51*, 12947−12957.

(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) Elstner, M. Ph.D. thesis, Universität-Gesamthochschule Paderborn, Paderborn, Germany, 1998.

(10) Gaus, M.; Chou, C.-P.; Witek, H.; Elstner, M. *J. Phys. Chem. A* **2009**, *113*, 11866−11881.

(11) Gaus, M. Ph.D. thesis, Karlsruhe Institute of Technology, Karlsruhe, Germany, 2011. Publicly available at www.bibliothek.kit.edu (accessed May 31, 2012)

(12) Bodrog, Z.; Aradi, B.; Frauenheim, T. *J. Chem. Theory Comput.* **2011**, *7*, 2654−2664.

(13) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.

(14) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.

(15) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(16) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. *J. Phys. Chem.* **1994**, *98*, 11623−11627.

(17) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100.

(18) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601−1622.

(19) Scholten, M. Ph.D. thesis, University of Düsseldorf, Düsseldorf, Germany, 2003.

(20) Elstner, M. *Theor. Chem. Acc.* **2006**, *116*, 316−325.

(21) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063−1079.

(22) Bodrog, Z.; Aradi, B. *Phys. Status Solidi B* **2012**, *249*, 259−269.

(23) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007−1023.

(24) Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **1999**, *110*, 7650−7657.

(25) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618−622.

(26) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764−7776.

(27) Winget, P.; Selçuki, C.; Horn, A. H. C.; Martin, B.; Clark, T. *Theor. Chem. Acc.* **2003**, *110*, 254−266.

(28) Krüger, T.; Elstner, M.; Schiffels, P.; Frauenheim, T. *J. Chem. Phys.* **2005**, *122*, 114110.

(29) Małolepsza, E.; Witek, H. A.; Morokuma, K. *Chem. Phys. Lett.* **2005**, *412*, 237−243.

(30) Witek, H. A.; Morokuma, K.; Stradomska, A. *J. Theor. Comput. Chem.* **2005**, *4*, 639−655.

(31) Witek, H. A.; Morokuma, K. *J. Comput. Chem.* **2004**, *25*, 1858−1864.

(32) Witek, H. A.; Morokuma, K.; Stradomska, A. *J. Chem. Phys.* **2004**, *121*, 5171−5178.

(33) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; 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.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision A.1; Gaussian Inc.: Wallingford, CT, 2009.

(34) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149−5155.

(35) Klopper, W.; Van Duijneveldt-van De Rijdt, J. G. C. M.; Van Duijneveldt, F. B. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2227−2234.

(36) Boese, A. D.; Jansen, G.; Torheyden, M.; Höfener, S.; Klopper, W. *Phys. Chem. Chem. Phys.* **2011**, *13*, 1230−1238.

(37) Pitoňák, M.; Neogrády, P.; Řezáč, J.; Jurečka, P.; Urban, M.; Hobza, P. *J. Chem. Theory Comput.* **2008**, *4*, 1829−1834.

(38) Niehaus, T. A.; Elstner, M.; Frauenheim, T.; Suhai, S. *J. Mol. Struct.: THEOCHEM* **2001**, *541*, 185−194.

(39) Sattelmeyer, K. W.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. A* **2006**, *110*, 13551−13559.

(40) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 42−55.

(41) Otte, N.; Scholten, M.; Thiel, W. *J. Phys. Chem. A* **2007**, *111*, 5751−5755.

(42) Hehre, W. J.; Ditchfield, K.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257−2261.

(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*, 506−511.

(45) Brothers, E. N.; Scuseria, G. E. *J. Chem. Theory Comput.* **2006**, *2*, 1045−1049.

(46) Riley, K. E.; Op't Holt, B. T.; Merz, K. M., Jr. *J. Chem. Theory Comput.* **2007**, *3*, 407−433.

(47) Wang, N. X.; Wilson, A. K. *J. Chem. Phys.* **2004**, *121*, 7632−7646.

(48) Raymond, K. S.; Wheeler, R. A. *J. Comput. Chem.* **1999**, *20*, 207−216.

(49) Lee, J. S. *Chem. Phys. Lett.* **2002**, *359*, 440−445.

(50) Russell D. Johnson III. *NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101*, Release 15b, August 2011. http://cccbdb.nist.gov/ (accessed May 31, 2012)

(51) Sattelmeyer, K. W.; Tubert-Brohman, I.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2006**, *2*, 413−419.

(52) Elke Goos, A. B.; Ruscic, B. *Extended Third Millennium Ideal Gas and Condensed Phase Thermochemical Database for Combustion with Updates from Active Thermochemical Tables Update of Third Millennium Ideal Gas and Condensed Phase Thermochemical Database for Combustion with Updates from Active Thermochemical Tables Alexander Burcat and Branko Ruscic*, Report ANL 05/20 and TAE 960 September 2005; Technion-IIT Aerospace Engineering, and Argonne National Laboratory Chemistry Division. ftp://ftp.technion.ac.il/pub/supported/aetdd/thermodynamics (accessed May 31st 2012).

(53) Tirado-Rives, J.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2008**, *4*, 297−306.

(54) Kalinowski, J. A.; Lesyng, B.; Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 2545−2549.

(55) Kaminski, S.; Giese, T. J.; Gaus, M.; York, D. M.; Elstner, M. *J. Phys. Chem. A* **2012**, *116*, 9131−9141.

(56) Kaminski, S.; Gaus, M.; Elstner, M. *J. Phys. Chem.* In press

(57) Goerigk, L.; Grimme, S. *J. Chem. Theory Comput.* **2010**, *6*, 107−126.
(58) Goerigk, L.; Grimme, S. *J. Chem. Theory Comput.* **2011**, *7*, 291−309.
(59) Korth, M.; Thiel, W. *J. Chem. Theory Comput.* **2011**, *7*, 2929−2936.
(60) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(61) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209−220.
(62) Stewart, J. J. P. *J. Mol. Model.* **2007**, *13*, 1173−1213.
(63) Kolb, M.; Thiel, W. *J. Comput. Chem.* **1993**, *14*, 775−789.
(64) Weber, W. Ph.D. thesis, University of Zürich, Zürich, Switzerland, 1996.
(65) Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495−506.
(66) Köhler, C.; Seifert, G.; Gerstmann, U.; Elstner, M.; Overhof, H.; Frauenheim, T. *Phys. Chem. Chem. Phys.* **2001**, *3*, 5109−5114.
(67) Köhler, C.; Frauenheim, T.; Hourahine, B.; Seifert, G.; Sternberg, M. *J. Phys. Chem. A* **2007**, *111*, 5622−5629.
(68) Aradi, B.; Hourahine, B.; Frauenheim, T. *J. Phys. Chem. A* **2007**, *111*, 5678−5684.
(69) Johnson, E. R.; Mori-Sánchez, P.; Cohen, A. J.; Yang, W. *J. Chem. Phys.* **2008**, *129*, 204112.
(70) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829−5835.
(71) Large scale atomistic simulations of materials: from bio-nano to solids: Seifert, G.; Elstner, M.; Deák, P. *Phys. Status Solidi B* **2012**, *249*, 225−415.
(72) King, H. W. In *CRC Handbook of Chemistry and Physics*, 93rd ed.; Haynes, W. M., Bruno, T. J., Lide, D. R., Eds.; CRC Press/Taylor and Francis: Boca Raton, FL, Internet Version 2013; p 12.15.
(73) Brewer, L. *The Cohesive Energies of Elements*; Technical Report, Lawrence Berkeley National Laboratory Reports LBL-3720, Berkeley, CA, 1977.
(74) Rauls, E.; Elsner, J.; Gutierrez, R.; Frauenheim, T. *Solid State Commun.* **1999**, *111*, 459−464.
(75) Niehaus, T. A.; Suhai, S.; Della Sala, F.; Lugli, P.; Elstner, M.; Seifert, G.; Frauenheim, T. *Phys. Rev. B* **2001**, *63*, 085108.
(76) Wanko, M.; Garavelli, M.; Bernardi, F.; Niehaus, T. A.; Frauenheim, T.; Elstner, M. *J. Chem. Phys.* **2004**, *120*, 1674−1692.
(77) Hourahine, B.; Sanna, S.; Aradi, B.; Köhler, C.; Niehaus, T.; Frauenheim, T. *J. Phys. Chem. A* **2007**, *111*, 5671−5677.
(78) Niehaus, T. A.; Della Sala, F. *Phys. Status Solidi B* **2012**, *249*, 237−244.
(79) Elstner, M.; Gaus, M. In *Computational Methods for Large systems: Electronic Structure Approaches for Biotechnology and Nanotechnology*; Reimers, J. R., Ed.; John Wiley and Sons: Hoboken, NJ, 2011; pp 287−308.
(80) Giese, T. J.; York, D. M. *Theor. Chem. Acc.* **2012**, *131*, 1145.