JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Parameterization of the DFTB3 Method for Br, Ca, Cl, F, I, K, and Na in Organic and Biological Systems

Maximilian Kubillus,[†] Tomáš Kubař,[†] Michael Gaus,[‡] Jan Řezáč,[§] and Marcus Elstner*[,†]

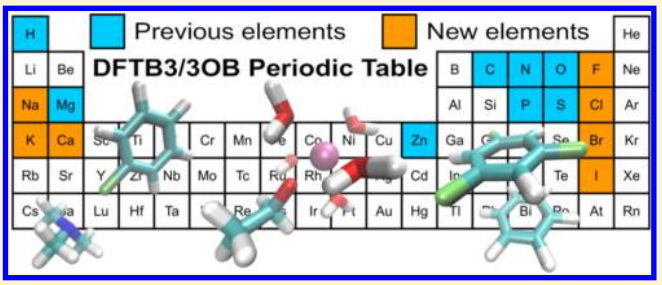
[†]Institute of Physical Chemistry, Karlsruhe Institute of Technology, Kaiserstr. 12, 76021 Karlsruhe, Germany

[‡]Institute of Theoretical Chemistry, University of Wisconsin—Madison, 702 West Johnson Street, Madison, Wisconsin 53715, United States

[§]Department of Computational Chemistry, Institute of Organic Chemistry and Biochemistry AS CR, v.v.i., Flemingovo nám. 2, CZ-166 10 Prague 6, Czech Republic

Supporting Information

ABSTRACT: We present an extension to the recent 3OB parametrization of the Density Functional Tight Binding Model DFTB3[1,2] for biological and organic systems. Parameters for the halogens F, Cl, Br, and I have been developed for use in covalently bound systems and benchmarked on a test set of 106 molecules (the 'OrgX' set), using bonding distances, bonding angles, atomization energies, and vibrational frequencies to assess the performance of the parameters. Additional testing has been done with the X40 set of 40 supramolecular systems containing halogens,[3] adding a simple correction for the halogen bonds that are strongly overbound in DFTB3. Furthermore, parameters for Ca, K, and Na as counterions in biological systems have been created. To benchmark geometries as well as ligand binding energies a test set 'BioMe' of 210 molecules has been created that cover coordination to various functional groups frequently occurring in biological systems. The new DFTB3/3OB parameter set outperforms DFT calculations with a double-ζ basis set in terms of energies and can reproduce DFT geometries, with some minor deviations in bond distances and angles due to the use of a minimal basis set.



1. INTRODUCTION

The Density Functional Tight Binding (DFTB) models are derived from density functional theory (DFT) using an approximation of the DFT total energy functional and were additionally extended by a Taylor expansion of the latter in later models. Denoting the DFT ground-state electron density as ρ, the energy is expanded around a reference density ρ_0 in terms of density fluctuations $\delta\rho = \rho - \rho_0$ in the respective orders as

$$E[\rho] = E^0[\rho_0] + E^2[\rho_0, (\delta\rho)^2] + E^3[\rho_0, (\delta\rho)^3] \quad (1)$$

Here, ρ_0 is chosen to be the superposition of neutral atomic densities. The methods constructed in this way may include terms up to the first, second, or third order and are designated DFTB1,[4,5] DFTB2 (formerly SCC-DFTB),[6] DFTB3,[1,7] respectively. To derive a computationally efficient scheme, the energy contributions in eq 1 are subject to further approximations. The resulting theory allows for electronic structure calculations that are 2 to 3 orders of magnitude faster compared to commonly used DFT methods with medium-sized basis sets (e.g. double-ζ valence quality complemented with polarization functions). Recently, the advantages and limitations of the different models have been reviewed,[8,9] and the application areas for such efficient methods have been discussed in detail.[10]

The individual energy contributions in eq 1—after the introduction of several approximations—involve various parameters. The second-order term E^2 contains the Hubbard parameter U (a measure for the chemical hardness) of the atoms, while the derivatives U are needed at the third order (E^3). These parameters may be obtained from DFT calculations and need not be optimized empirically. For the calculation of parameters appearing in the zeroth-order term E^0, a minimal atomic basis set has to be specified, and this involves two parameters per atom type. The most significant effort is required to parametrize the contributions constituting the so-called repulsive energy E^{rep}, which consists of a sum of repulsive potentials V_{ab}^{rep} that are parametrized as splines for each atom pair.

In principle, E^0 does not depend on the order of the method; that is, it should be the same for DFTB1, DFTB2, and DFTB3. In practice, however, the accuracy of a DFTB model may be improved by optimizing the confinement radii, thus changing overlaps and Hamiltonian matrix elements and therefore the DFTB total energy. So, the "MIO" parameter set[6] was developed with optimized confinement radii for the DFTB2 model. Following the implementation of DFTB3, the basis set was reoptimized, yielding the "3OB" set.[2] Specifically, the

Received: October 14, 2014

DOI: 10.1021/ct5009137
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

confinement radii were modified slightly in order to reduce overbinding and improve intermolecular distances dominated by van der Waals interactions.[2] Owing to the different underlying basis sets, the parameters contained in "MIO" and "3OB" sets themselves cannot be combined in a single DFTB calculation due to the different Hamiltonians and overlaps when using different confinement radii in each DFTB model. Representing the state of the art of DFTB calculations of organic and biological molecular systems, the "3OB" parameter set for the DFTB3 model has been created involving the following elements so far: O, N, C, and H,[2] extended for P and S,[11] and further for Zn and Mg.[12] In this work, the "3OB" parametrization for the halogens F, Cl, Br, I, and the metals Na, Ca, and K is presented.

These elements are of great importance in several areas of biological research. From enzyme catalysis involving alkaline and alkaline earth metals[13] to the relevance of competing binding strengths of metals in ion channels,[14] there is a broad range of applications for DFTB3 parameters. Also, another notable example of the effects of metals on proteins is lymphotactin where lower concentration of ions can induce dimerization of the protein.[15] Halogens are commonly used in drug design studies.[16] Aside from the electrostatic effects of halogens on substrates, the utilization of halogen bonds has been of great interest in research lately.[17] With this extension of the 3OB parametrization of DFTB3 the user is enabled to conduct extensive studies on the mentioned topics, including simulations on biological systems where standard force-field based methods might not be sufficient.

2. THEORY

The DFTB3 methodology was detailed recently;[1,8,9] therefore, only a short summary is given here to elucidate the meaning of the various parameters in DFTB3 as well as the parametrization process.

The DFTB models use a minimal atomic orbital basis set and the basis functions are obtained explicitly as one-electron wave functions from DFT calculations on a free atom, applying an additional confining potential:

$$\left[\hat{T} + V^{\text{eff}}[\rho_a] + \left(\frac{r}{r_0}\right)^{\sigma}\right]\phi_\mu = \epsilon_\mu \phi_\mu \tag{2}$$

Here, $V^{\text{eff}}[\rho_a]$ is the effective Kohn−Sham potential for atom a using the Perdew, Burke, and Ernzerhof (PBE)[18] functional for the exchange-correlation part. The DFTB basis set is composed of orbitals yielded by this calculation, and as such, it is determined by the parameters r^0 and σ, which define the confinement potential. Two different sets of parameters r_0 and σ are usually considered, one each (i) for the calculation of electron density ρ_a^0 of atom a, and (ii) to obtain the atomic orbitals ϕ_μ. It should be noted that the on-site energies (atomic orbital energies) ϵ_μ are obtained from yet another DFT calculation without any confinement in eq 2 ($r^0 = \infty$). In this work, the zeroth-order regular approximation (ZORA)[19] was used to treat the relativistic effects in Br and I, while nonrelativistic calculations were employed for all of the other elements. The off-diagonal Hamiltonian matrix elements are computed using a two-center approximation, assuming superposition of atomic densities, as

$$H^0_{\mu\nu} = \langle \phi_\mu | \hat{T} + V[\rho_a^0 + \rho_b^0] | \phi_\nu \rangle \tag{3}$$

and the overlap matrix is obtained as $S_{\mu\nu} = \langle \phi_\mu | \phi_\nu \rangle$.

In the DFTB3 framework the expansion of Kohn−Sham molecular orbitals in a linear combination of atomic orbitals $\psi_i = \Sigma_\mu c_{i,\mu}\phi_\mu$ leads to an expression for the total energy:

$$E^{\text{DFTB3}} = E^0 + E^1 + E^2 + E^3 = \frac{1}{2}\sum_{ab} V_{ab}^{\text{rep}} + \sum_i \sum_{\mu\nu} c_\mu^i c_\nu^i H^0_{\mu\nu} + \frac{1}{2}\sum_{ab} \Delta q_a \Delta q_b \gamma_{ab} + \frac{1}{3}\sum_{ab} \Delta q_a^2 \Delta q_b \Gamma_{ab} \tag{4}$$

Here, Δq_a is the negative of the Mulliken net charge of atom a, and the functions γ_{ab} and $\Gamma_{ab} = ((\partial\gamma_{ab})/(\partial q_a))|_{q_a^0}$ describe the interaction of density fluctuations on the individual atoms, that is, of the electron density that has been transferred between atoms. The γ function and its derivative are evaluated from their analytical forms at run-time, as described in ref 16. The zeroth-order term E^0 is approximated by a sum of pairwise functions of interatomic distances, so-called repulsive potentials V_{ab}^{rep}, which are represented by fourth-order splines. The spline coefficients, cutoff and division points are determined with the parametrization protocol by Gaus et al. implemented in the *erepfit* program to parametrize splines.[1,20] Each of these splines approaches zero at a certain interatomic distance (cutoff point), and it is divided into several intervals in the parametrization. Continuity conditions are imposed to V_{ab}^{rep} as well as its first three derivatives. From input geometries, atomization and reaction energies as well as additional equations (assigning specific values at certain points of the spline or its derivatives), a linear equation system is created, which is then solved via singular value decomposition to obtain the spline coefficients.

Finally, we complemented the DFTB3 energy with a correction for halogen bonds,[21] which are overbound in DFTB strongly due to the missing residual Pauli repulsion between the confined (tightly bound) DFTB basis set orbitals. The overall root-mean-square deviation (RMSD) of interaction energies in the X40 set of supramolecular interactions involving halogens is 2.2 kcal/mol and the major sources of error are halogen bonds (RMSD 4.3 kcal/mol) and hydrogen bonds (RMSD 2.6 kcal/mol). While the errors in hydrogen bonds emerge mainly in hydrogen halogenides coordinating to organic ammonia derivatives, halogen bonds are overestimated systematically. DFTB inherits those deviations mainly from the PBE functional where halogen bond energies are overestimated markedly even with a double-ζ basis set (see also the results section 4.2). The error is of the same nature as found with the semiempirical method PM6[22] and can be rationalized as an insufficient residual Pauli repulsion (both electrostatic and dispersion contributions are attractive). A simple empirical correction—the "X correction"—was developed for PM6, adding a repulsive interatomic potential between halogens and each possible acceptor (see eq 5). This potential is parametrized so that the corrected method reproduces dissociation curves of halogen bonds. This simple and efficient solution is applied to DFTB3 as well. We introduce the atomic pair-potential in a slightly modified form and find a DFTB3-X energy term

$$E_{ab}^X = \frac{1}{2}\sum_{A,B} c_1 \cdot \exp[-c_2(R - d_{ab})^{c_3}] \tag{5}$$

DOI: 10.1021/ct5009137
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 1. Electronic Parameters for All Atoms Parameterized in This Work (In Atomic Units Where Not Dimensionless)[a]

param.	Ca	K	Na	F	Cl	Br	I
			definition of the basis set				
l_{max}	1	1	1	1	2	2	2
n_{max}	2	2	1	2	2	2	2
a_0	0.50	0.50	0.50	0.50	0.50	0.50	0.50
a_1	1.26	7.65	1.08	1.03	1.21	1.45	1.60
a_2	3.16	3.08	2.35	2.12	2.92	4.18	5.15
a_3	7.95	1.24	5.08	4.37	7.04	12.10	16.52
a_4	20.00	19.00	11.00	9.00	17.00	35.00	53.00
r^s	6.4	7.0	4.4	2.8	4.3	6.0	6.3
r^p	9.9	7.9	4.4	2.8	4.3	6.0	6.3
r^d					5.7	6.0	7.8
r^{dens}	12.0	12.0	8.6	5.0	13.0	10.0	11.0
σ	13.6	15.8	2.0	2.0	9.0	6.4	2.0
			calculated parameters				
ϵ_s	−0.1379	−0.0852	−0.1007	−1.0958	−0.7544	−0.7394	−0.6382
ϵ_p	−0.0521	+0.0148	−0.0274	−0.4087	−0.3164	−0.2888	−0.2606
ϵ_d					+0.0222	+0.0243	+0.0223
E^{spin}	−0.0236	−0.0056	−0.0081	−0.0146	−0.0080	−0.0069	−0.0057
U	+0.1764	+0.1356	+0.1651	+0.5584	+0.3668	+0.3277	+0.2842
U^{der}	−0.0340	−0.0339	−0.0454	−0.1623	−0.0697	−0.0573	−0.0433

[a]Parameters a_i, E^{spin}, n_{max}, U, U^{der}, and ϵ_i are not subject to optimizations. All of the values r_i and σ are fitted to reproduce the reference data. The calculated parameters were obtained with the procedures described in our previous work.[2,11]

with c_1 to c_3 being universal parameters for the 3OB parametrization and d_{ab} being a parameter for the element pair a and b. To apply the correction only for nonbonding atom pair interactions, we add a switching function to the X correction. Also, analytical gradients can be obtained from the energy term to calculate forces, as described in detail in the Appendix. Finally, it was confirmed that the correction does not introduce any artifacts in the bond dissociation energy profiles, for all pairs of atoms including those covalently bound. A modified version of the DFTB+[23] package with an implementation of the X correction was also created, including gradients for geometry optimizations and molecular dynamics simulations.

3. PARAMETRIZATION OF DFTB3

The DFTB3 energy consists of the electronic contributions E^0 to E^3, which determine the electronic structure of the system, and the repulsive energy part E^{rep}, which takes the form of a sum of pair potentials.

The electronic contributions contain only monatomic parameters, which are needed to compute the diagonal as well as the off-diagonal elements of the Hamiltonian, while the repulsive potentials of E^{rep} are of two-body nature. Therefore, determination of the latter is a problem quadratic in the number of elements in principle. In practice, however, not all of the combinations of elements are needed, which reduces the parametrization efforts.

All electronic parameters introduced in this work are calculated with the *slateratom* and *sktwocnt* programs developed at the University of Bremen, and their values are summarized in Table 1. Listed first are parameters defining the basis set: All of the a_i values, l_{max}, and n_{max} define the Slater type orbitals, while the confinement radii r and the exponent σ define the confinement potential. With these parameters a Slater type basis is created and then used to compute the off-diagonal overlaps $S_{\mu\nu}$ and PBE Hamilton matrix elements $H^0_{\mu\nu}$. Note that the diagonal elements $\epsilon = H^0_{\mu\mu}$ are computed without any confinement potential. The remaining one-center parameters are calculated from DFT, also using the PBE functional: orbital energy eigenvalues ϵ, which enter the Hamiltonian diagonal; the spin-polarization energy E^{spin}, which is only needed to compute the atomization energy;[2] and the Hubbard parameters U and its derivative U^{der}, entering the second and third order terms.

The exponents a_i, which correspond to the effective nuclear charges of the Slater-like atomic orbitals, were not optimized. Much the same as in the previous 3OB parametrizations,[2,11] each chemical element features several generally different confinement radii; these are r^s, r^p, and r^d for the s, p and d orbitals, respectively, and there is another radius r^{dens} applied in the calculation of electron density of the neutral atom, as described above. All of the confinement radii were subject to optimization. As a starting value in the optimization, the value of twice the covalent radius r_{cov} was used, which had been shown to be a sensible choice in previous parametrizations.[24] Since alkaline earth metals in biomolecular systems occur as solvated ions in the most cases, the ionic radii r_{ion} proved themselves as better starting values for the parametrization.

The confinement potential had been usually chosen to be quadratic ($\sigma = 2$). The use of a higher exponent would bring on a more distinct decline of the wave function and/or density with increasing distance from the nuclei. Indeed, it was shown recently that such a modified confinement may result in a better performance of the parametrization.[25] In this work, most of the exponents have been taken from Table 1 in ref 25, including the value for I, which is $\sigma = 2$ by coincidence. The exceptions to this are Na and F, for which the previously conventional choice of $\sigma = 2$ was retained.

While the application of a minimal basis set is always preferable within DFTB due to the favorable computational efficiency, the description of some bonding situations may be greatly improved by the addition of polarization functions. The performance of the current parametrization for systems containing Na and Ca has improved consistently upon addition

DOI: 10.1021/ct5009137
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 2. Training Sets for the Parameterization of Repulsive Potentials Using Molecular Geometries and Reactions[a]

atom	molecules	reactions
Ca	Ca_2H_2, CaH_2, $HCaCH_3$, $Ca(NH_2)_2$, $Ca(OH)_2$, $HCa(SH)$, $HCa(PH_2)$, $Ca(HPO_4)$, $[Ca(NH_3)]^{2+}$, $[Ca(OH_2)]^{2+}$, $[Ca(OH_2)_4]^{2+}$, $[Ca(SH_2)]^{2+}$, $[Ca(SH_2)_4]^{2+}$, $[Ca(SH_2)_6]^{2+}$, CaMg, CaZn, HCaK, HCaNa	$Ca^{2+} + H_2O \rightarrow [Ca(OH_2)]^{2+}$ $Ca^{2+} + NH_3 \rightarrow [Ca(NH3)]^{2+}$ $Ca^{2+} + H_2S \rightarrow [Ca(SH_2)]^{2+}$
K	K_2, KH, $K(OH)_{lin}$, $[K(NH_3)]^+$, $K(CH_3)$, $K(PH_2)$, K(SH), HMgK, HZnK, KNa	$K^+ + H_2O \rightarrow [K(OH_2)]^+$ $K^+ + NH_3 \rightarrow [K(NH_3)]^+$ $K^+ + H_2S \rightarrow [K(SH_2)]^+$
Na	Na_2, NaH, $Na(OH)_{lin}$, $Na(NH_2)$, $Na(CH_3)$, $Na(PH_2)$, $[Na(SH_2)]$+, NaMgH, NaZnH	$Na^+ + H_2O \rightarrow [Na(OH_2)]^+$ $Na^+ + NH_3 \rightarrow [Na(NH_3)]^+$ $Na^+ + H_2S \rightarrow [Na(SH_2)]^+$
F	F_2, HF, CH_3F, NH_2F, OHF, PH_2F, SHF, NaF, MgF, KF, CaHF, ZnHF	
Cl	Cl_2, HCl, CH_3Cl, C_2HCl, CCl_4, NH_2Cl, ClOH, $HClO_4$, ClO_4^-, PH_2Cl, ClSH, NaCl, ClMgH, ClF, KCl, ZnHCl, CaHCl	$HClO_4 + H_2 \rightarrow HClO_3 + H_2O$
Br	Br_2, HBr, CH_3Br, NH_2Br, BrOH, BrO^-, BrO_4^-, PH_2Br, BrSH, NaBr, MgHBr, BrF, BrCl, KBr, ZnHBr, CaHBr	
I	I_2, HI, CH_3I, C_2HI NH_2I, IOH, IO^-, PH_2I, SHI, NaI, MgHI, IF, ICl, IBr, KI, ZnHI, CaHI	

[a]All of the remaining data needed to reproduce the repulsive potentials are given in the Supporting Information.

of p-type functions to the basis set on these elements; the geometries of studied molecular systems were reproduced in accordance with the DFT reference data. On the other hand, the addition of d-orbitals seemed to lead to an unbalanced basis set, resulting in large deviations of the molecular structures.

The pairwise repulsive potentials reach from a distance of zero up to a cutoff point where the potential reaches zero. This range of distances is divided by knot points into several intervals, and the repulsive potential takes the form of a fourth-order polynomial on each of the intervals. Continuity conditions are imposed on the polynomials as well as its first three derivatives. The most of the repulsive potentials are defined on a sequence of four intervals. Some potentials, like the halogen-carbon interactions, are parametrized with a larger number of intervals due to the different binding situations at different bond distances, that is, the halogen binding to sp, sp^2, and sp^3 carbon. In special cases, the number of intervals was reduced to two, and the potential was fitted for a single geometry to an atomization energy that was made more repulsive artificially (see below). The cutoff point was, in most cases, chosen as 75% of the difference between a typical bonding and second neighbor distance, added to the shortest bonding distance for the respective pair of elements that was present in the training set. This ensures that the spline mostly affects only chemical bonds and no long-range interactions. In cases with no typical second neighbor distance in a biological environment, the bonding distance was multiplied with a factor of 1.4 to obtain a suitable cutoff for that potential. The detailed knot point definition for each repulsive potential can be found in the Supporting Information. A set of molecular systems to train the repulsive potentials was set up for the metal−oxygen pair potential, considering geometries obtained with DFT optimizations on the B3LYP level of theory with triple-ζ basis sets. Reference atomization and reaction energies were computed using G3B3, G4, CCSD(T)/CBS, and B3LYP/def2-TZVPP levels of theory. Details on which references were used to train each potential can be found in the Supporting Information. The training sets for all atoms parametrized in this work are shown in Table 2. Simple uncharged molecules were chosen where the accuracy of the resulting repulsive potential was sufficient in benchmarks. For most interactions involving coordination to N, O, or S, geometries of charged complexes were used (e.g., water or ammonia complexes), together with their respective ligand binding energy. In some cases, the splines were divided into multiple intervals and the parametrization involved geometries of multiple complexes with an increasing solvation shell (binding additional water or ammonia molecules) to improve upon bond distance accuracy. In case of the metal−metal, halogen−halogen, and metal−halogen potentials, the binding energies were artificially modified to be more repulsive, in order to prevent the formation of spurious bonds between ions, observed in pilot simulations, bonds that were not dissociating again. For halogens, some adjustments in the splines between halogens and carbon were made mainly to improve bond distances and energies for molecules containing halogens bound to sp^2- or sp-hybridized carbon. Further information needed to reproduce the repulsive potentials exactly, such as the atomization energies used to parametrize the splines, geometries, as well as force and energy weights, can be found in the Supporting Information.

In case of metal ions coordinating to O, N, or S, the quantity being benchmarked to the reference was the incremental binding energy, which corresponds to the energy of binding an additional water molecule to a cluster formed by the metal ion and several water molecules that are already bound. A range of wave function confinement radii between r_{ion} and 2.5 r_{cov} was scanned, testing several values for the density compress radius r^{dens} together with each value of the wave function radius.

Regarding the halogens, a minimal basis set configuration $2s^2$ $2p^5$ was considered for F, while a set of valence d-functions was applied for Cl, Br, and I as polarization functions, the 3OB basis for those halogens having the configuration Ns^2 Np^5 Nd^0, with N being 3, 4, and 5, respectively. This was shown to be necessary for the description of hypervalent systems and intermolecular interactions. Further, as discussed in ref 26, the rotational barrier of hydrogen in the halogen oxo acids is not correctly reproduced using a minimal basis for halogens. Adding d-type polarization functions to halogens improved upon this barrier, and a qualitatively correct representation for perchloric acid (ClO_4H) is found, in contrast to the wrong minimum obtained with a minimal basis set; see Figure 1. Still, the rotational barrier is overestimated by 0.8 kcal/mol in DFTB3/3OB in comparison with the PBE/cc-pVTZ reference. This seems to be an effect emerging from the use of a minimal basis, as illustrated by the calculations performed with DFTB2/MIO and B3LYP/STO-3G, which show a minimum at a dihedral angle of 0°.

The confinement radii for the halogens were determined by scanning the wave function radii from r_{cov} and 2.5 r_{cov}. The performance was assessed on a set of relevant molecules and

DOI: 10.1021/ct5009137
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

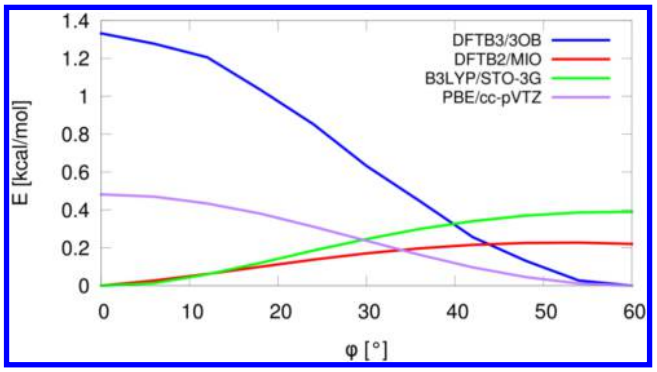


Figure 1. Energy of a ClO_4H molecule as a function of the dihedral angle H−O−Cl−O (φ) obtained on different levels of theory. Upon the addition of d-type basis functions, the rotational barrier is reproduced qualitatively (albeit overestimated largely). A wrong conformation is predicted with a minimal basis set (DFTB2/MIO, B3LYP/STO-3G).

complexes. Once a suitable training set for all repulsive potentials of an atom was found, all confinement radii for the density and for the d-type polarization functions were reoptimized to tune the atomization energies and the vibrational frequencies. This was done by scanning confinement radii around the previously chosen values in an interval of ±1 au. New repulsive potentials were created to complement the modified electronic parameters for each new combination of confinement radii, using the same protocol as described above.

A collective parametrization of the repulsive potentials was chosen in most of the cases. This means that all of the repulsive parameters were determined for a small training set simultaneously, aiming at the reproduction of reference binding energies and geometries. In this way, the reference energies are reproduced more consistently in general, compared to an alternative setup with a separate fitting of each potential.[2,20]

The reference atomization energies were obtained on the G3B3, G4, or CCSD(T) level of theory. Several repulsive potentials involving halogens were fitted to artificially increased binding energies, to compensate for the overbinding tendency of DFTB, which was observed for some of these bonding situations. In the case of Ca, K, and Na, the energies were fitted to minimize the deviations of incremental ligand binding energies from the reference, see further the following section.

An exception is the Ca−P potential, which was fitted to the geometry of the $Ca^{2+}\cdots PO_4H_2^-$ ion pair in order to avoid interference with the Ca−O repulsive interaction. All of the other repulsive potentials involving Ca were included in the collective fitting described above. Another noteworthy detail concerns the design of metal−metal and metal−ion interactions: stronger repulsives were created here so that no bonds between two counterions may form in molecular dynamics simulations. In consequence, the energetic description of these interactions within the range of repulsive potentials is insufficient. This will not influence most applications in biological systems but will make the application of the parameters for solids unfeasible.

The halogen bond correction (eq 5) was parametrized for the noncovalent interactions in the X40 × 10 set of dissociation curves,[3] taken from the Benchmark Energy and Geometry Database (BEGDB).[27] The parameters for interactions between halogen acceptors Cl, Br, and I and the donors O and N were found by fitting to molecular complexes 19 to 24 and are summarized in Table 3. Halogen-sulfur bonds are described

Table 3. Parameters Used for the D3(BJ) Correction (For Details See Ref 30) and Parameters Defining the Halogen Correction As Described in Equation 5 (Given in the Å and kcal/mol Unit System)

DFTB-D3(BJ) param.		
a1	a2	s8
0.746	4.191	3.209
halogen bond correction		
c1	c2	c3
7.761	0.050	4.518
element pair		d_{ab}
O−Cl		1.237
O−Br		1.099
O−I		1.313
N−Cl		1.526
N−Br		1.349
N−I		1.521

sufficiently without any corrections. The CCSD(T)/CBS interaction energies from the original work were used as a reference. In the DFTB3 calculation, D3 dispersion correction[28] with Becke-Johnson damping[29] and parameters for the DFTB3/3OB parametrization with ζ = 4.0[30,31] was applied together with the X correction. The DFTB3/3OB performance was compared to PBE-D3/def2-SVP calculations. All energies were obtained from geometries optimized using gradients of the respective method and compared to the CCSD(T)/CBS reference energies from ref 3.

4. RESULTS

4.1. Calcium, Potassium, and Sodium. The first test consisted in a study of small gas-phase complexes of the ions coordinated with one to six molecules of water, ammonia or hydrogen sulfide. All of the geometries were obtained with DFT optimization, using the B3LYP functional and the def2-TZVPP basis set for Ca and K,[32] and the complexes of Na were optimized with the Gaussian-B3LYP functional and the cc-pVTZ basis set. The incremental binding energies of the complexes of Ca were obtained on the CCSD(T) level of theory,[33] using a two-point extrapolation method[34,35] and the basis sets cc-pVTZ and cc-pVQZ.[36−38] All of these calculations were carried out with the Turbomole 6.4 package, using the resolution of identity implementation wherever possible.[39−44] Energies of the complexes of Na and K were obtained from G3B3[45] and G4[46,47] calculations (because of the high computational cost, all of the energies of clusters $[Ca(H_2S)_n]^{2+}$ were obtained with G3B3 instead of CCSD(T)/CBS), respectively, using the *Gaussian 09* package.[48] All DFTB3 calculations were carried out using the DFTB+ 1.2.2 package.[23]

The incremental ligand binding energy benchmark results are presented in Table 4. Significant deviations are generally seen for the incremental energies binding to hydrogen sulfide molecules. The largest RMSD is found for the $[K(H_2S)_n]^{2+}$ complexes with 10.4 kcal/mol, with an error as large as 53% of the binding energy for the incremental energy from a coordination number of three to four. This is mainly due to the failure of DFTB3 to reproduce the geometries of those clusters as well as the geometries of ammonia or water binding complexes. DFTB3 exhibits notably small errors for the incremental binding energy of the higher-coordinated species for ammonia and water complexes, and larger errors tend to

DOI: 10.1021/ct5009137
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 4. Incremental Ligand Binding Energies for Metal Complex Clusters[a]

calcium clusters			sodium clusters			potassium clusters		
H_2O	CCSD(T)	Δ3OB	H_2O	G3B3	Δ3OB	H_2O	G4	Δ3OB
1 → 2	−49.0	−5.9	1 → 2	−22.6	−0.6	1 → 2	−15.8	−0.6
2 → 3	−44.9	−2.6	2 → 3	−19.4	−1.4	2 → 3	−14.0	−1.5
3 → 4	−37.7	−1.9	3 → 4	−16.9	−0.8	3 → 4	−12.4	−0.9
4 → 5	−31.1	1.0	4 → 6	−24.8	−0.8			
5 → 6	−28.2	1.0						
MSD	MAD	RMSD	MSD	MAD	RMSD	MSD	MAD	RMSD
−0.6	1.2	1.8	−0.7	0.7	0.9	−1.0	1.0	1.1
NH_3	CCSD(T)	Δ3OB	NH_3	G3B3	Δ3OB	NH_3	G4	Δ3OB
1 → 2	−53.9	−3.9	1 → 2	−27.5	1.8	1 → 2	−17.1	−3.1
2 → 3	−48.8	−1.1	2 → 3	−22.6	−1.7	2 → 3	−14.8	−4.8
3 → 4	−39.5	−4.5	3 → 4	−18.7	−3.2	3 → 4	−12.6	−5.5
4 → 5	−30.3	0.6	4 → 5	−12.2	−1.1			
5 → 6	−25.9	1.0	5 → 6	−10.6	−1.5			
MSD	MAD	RMSD	MSD	MAD	RMSD	MSD	MAD	RMSD
−2.0	2.2	2.8	−0.8	1.5	1.9	−4.5	4.5	4.6
H_2S	G3	Δ3OB	H_2S	G3B3	Δ3OB	H_2S	G4	Δ3OB
1 → 2	−37.7	−7.6	1 → 2	−15.1	1.8	1 → 2	−21.6	10.1
2 → 3	−32.4	−5.2	2 → 3	−2.9	−6.1	2 → 3	−15.3	6.1
3 → 4	−28.4	0.6	3 → 4	−21.1	0.7	3 → 4	−25.5	13.6
4 → 5	−23.0	3.8	4 → 5	−9.1	−1.5			
5 → 6	−21.8	4.3	5 → 6	−9.3	0.7			
MSD	MAD	RMSD	MSD	MAD	RMSD	MSD	MAD	RMSD
−0.8	4.3	4.8	−1.9	2.2	3.0	9.9	9.9	10.4

[a]Presented are reference values together with the level of theory on which they were obtained, and the deviations (in kcal/mol) from reference values of the results obtained with the current 3OB parameterization of DFTB3.

Table 5. Reaction Energies of Ligand Exchange Reactions of the Fully Coordinated Complex $[Ca(H_2O)_6]^{2+}$ (Ca_{aq}^{2+})[a]

	B3LYP[b]	ΔDFTB3	ΔPBE[c]
$Ca_{aq}^{2+} + NH_3 \rightarrow [Ca_{aq}(NH_3)]^{2+} + H_2O$	1.5	−1.4	−1.0
$Ca_{aq}^{2+} + H_2S \rightarrow [Ca_{aq}(SH_2)]^{2+} + H_2O$	12.1	−1.3	+3.5
$Ca_{aq}^{2+} + C_2H_4O \rightarrow [Ca_{aq}(C_2H_4O)]^{2+} + H_2O$	8.3	+2.8	−5.7
$Ca_{aq}^{2+} + C_2H_5O^- \rightarrow [Ca_{aq}(C_2H_5O)]^{+} + H_2O$	209.3	−4.3	+9.7

[a]Deviation of reaction energies obtained with the new DFTB3/3OB parameterization as well as with PBE/def2-SVP from reference calculations (hybrid DFT); values in kcal/mol. All energies were obtained from geometry optimizations with the respective method. [b]Using basis set aug-cc-pVTZ on a structure optimized with B3LYP/cc-pVTZ. [c]Using basis set def2-SVP.

occur for the lower coordination numbers. It has to be noted here that this is a desired trade-off because the former situation is highly relevant in biomolecular systems while the latter is not. Mostly, the metal ion possesses a full or nearly full coordination sphere, and the reactions taking place consist in either removal of a ligand or substitution of a ligand by another, while all of the other ligands remain coordinated. This point is detailed in Table 5, where several ligand exchange reactions of solvated Ca^{2+} are studied. The reactions exchanges one ligand from a solvation shell of six waters and replaces it with a different binding motif: ammonia, hydrogen sulfide, ethanolate, or acetaldehyde. In most cases, DFTB3 performs slightly better than PBE/def2-SVP, the exception being the ammonia ligand exchange. A notable example also is the ethanolate exchange reaction where the coordination angles in the gas-phase alcoholate−calcium complexes have large systematic deviations (see below).

Further, DFTB3 was tested for 210 complexes, which feature the interactions of the ions with O, N, S, and P atoms of small molecules. Also included were several complexes featuring the interaction of a metal ion with the π-system of a small aromatic molecule. The detailed analysis of bond distances, bond angles, atomization energies, and vibrational frequencies is presented in the Supporting Information, and Table 6 summarizes the results. The geometries obtained with DFTB3 are slightly worse than those yielded by PBE/def2-SVP calculations, while the incremental binding energies of complexes that were sufficiently reproduced by the 3OB parameters are better than PBE/def2-SVP. Overall, the deviations from reference data are small, the bond distances deviating by up to 12% in extreme cases such as a secondary nitrone coordinating to potassium while PBE/def2-SVP deviates only by 6%. As mentioned above the ligand binding energies of hydrogen sulfide have larger deviations, similar to PBE/def2-SVP, but the binding energies of water and ammonia are notably better reproduced than with PBE. Therefore, DFTB3 seems to be well suited to describe the interactions of small organic molecules with Na, K, and Ca ions.

The performance, however, differs for different functional groups. Therefore, the performance of DFTB3 and PBE/DZ was analyzed in more detail, and the results are shown by the example of Ca in Figures 2 and 3. Perhaps caused by the small basis sets employed in DFTB3, there are issues with the

DOI: 10.1021/ct5009137
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 6. Performance of the New DFTB3/3OB Parameters for Na, K, and Ca[a]

	DFTB3/3OB			PBE/def2-SVP		
property	Ca	K	Na	Ca	K	Na
bond distances	0.049	0.074	0.048	0.018	0.029	0.016
bond angles	9.5	5.3	11.6	3.8	3.9	2.5
LBE[b] H_2O	2.5	1.0	0.7	9.2	6.2	5.0
LBE[b] NH_3	2.2	4.5	1.5	6.8	5.3	1.8
LBE[b] H_2S	4.3	9.9	2.2	6.6	9.4	1.4

[a]Errors are presented as mean absolute deviations (MAD) from the results of reference calculations in the AA and kcal/mol unit system. Equivalent data for the PBE functional with a double-zeta basis set are presented for comparison. For the particular absolute values, types of reference data and the numbers of test systems, see the Supporting Information. All benchmark properties were obtained from geometry optimizations with the respective method. [b]Incremental ligand binding energies.

coordination of lone pairs, especially of oxygen atoms. Owing to the monopole approximation, the electron charge transferred between atoms is forced to a spherical shape. Consequently, the potential energy as a function of the coordination angle of the oxygen anion features wrong minima. Especially high deviations from the reference geometries are observed in the alcoholate and nitrone complexes studied. An interesting observation is that DFT calculations performed with a small basis set show similar effects, see Figure 4. Double- and triple-ζ optimized geometries (molecules b and c in Figure 4) for gas-phase sodium ethanolate show a sodium coordination angle around 170° and a binding distance between sodium and the alcohol oxygen of 1.96 Å. With a minimal basis set (molecule a in Figure 4), the bond distance and angle show large deviations of 66.4° and 0.056 Å compared to the triple-ζ geometry. As mentioned above, this error emerges from an insufficient description of oxygen lone-pairs with a minimal basis set. Another point is that not all bonding angles are reproduced with sufficient accuracy compared to the reference geometries. Importantly, most of these problems vanish as soon as a water shell is added to the metal. In these cases, even the binding energies are in agreement with hybrid DFT calculation, and DFTB3/3OB performs better than double-ζ basis set PBE calculations in most cases.

The application of DFTB3/3OB seems to be problematic for the alkyne−calcium π interaction. In the example of 2-butyne in Figure 5, the $C_3-C_2-C_1$ angle predicted by DFTB3/3OB is only 141° as compared to 174° of the reference. The distance between the calcium ion and carbon decreases from 2.439 to 2.383 Å, suggesting the formation of a weak bond and breaking of the carbon−carbon triple bond. However, this is an exceptional occurrence, which has been observed neither with sodium nor for any double bond tested. Concerning the Ca−S interaction, DFTB3 seems to be problematic with the description of H_2S containing systems. $[Ca(H_2S)_6]^{2+}$, optimized with DFTB3, shows a Ca−S−H bond angle of about 128°. This is very similar to that in water clusters (127°). Hybrid DFT calculations (B3LYP/def2-TZVPP), on the other hand, predict an angle of 105.7°. This suggests that DFTB3/3OB fails to predict geometries in systems weakly coordinating to H_2S since the structures obtained from optimizations are very similar to water clusters.

4.2. Halogens. The parameters for halogens were benchmarked using a test set of 88 organic molecules and 18 halogen oxides (oxoacids containing 2 to 4 oxygen atoms as well as their deprotonated anions), where each of the elements F, Cl, Br, and I is contained in at least 22 molecules. The summary of performance of current parametrization for bond distances and angles, atomization energies, and vibrational frequencies for the organic molecules is given in Table 7. Halogen oxides were only parametrized to reproduce B3LYP geometries and are not included in Table 7. The new parametrization performs well in terms of bonding distances but it suffers from larger errors with bonding angles. It is to note that those errors mainly result from increasing the wave function confinement radii as compared to the DFTB2/MIO[26] parametrization, which was done to improve upon supramolecular interactions. All in all, a similar picture as for the metal ions is found: While DFTB3 provides slightly better energies than PBE with a double-ζ basis set does, it shows a slightly worse performance on geometries. Contrary to the C, H, N, and O parameters, there is no separate parameter set optimized for vibrational frequencies provided. The deviations of vibrational frequencies observed with the "all-purpose" parametrization are slightly larger than in the PBE comparison but well within the usual accuracy of the 3OB parametrization of DFTB3.

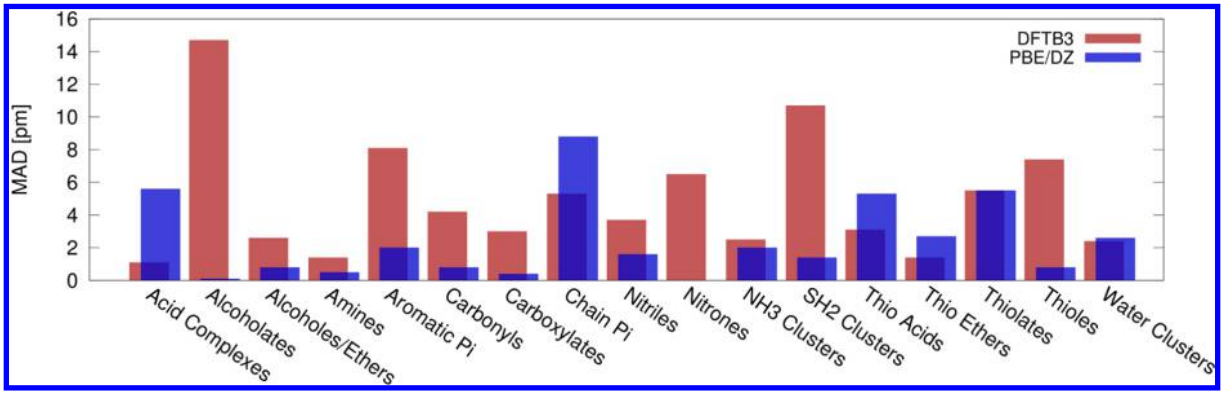


Figure 2. Accuracy of bonding distances in complexes of calcium with various organic functional groups reproduced by the current DFTB3 parametrization The deviations observed with PBE/def2-SVP calculations are shown for comparison. The reference data were obtained with DFT optimizations on the level B3LYP/def2-TZVPP using Turbomole 6.4. Errors expressed as mean absolute deviations (MAD) in pm. The test systems contained in each category mentioned here are detailed further in the Supporting Information.

DOI: 10.1021/ct5009137
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

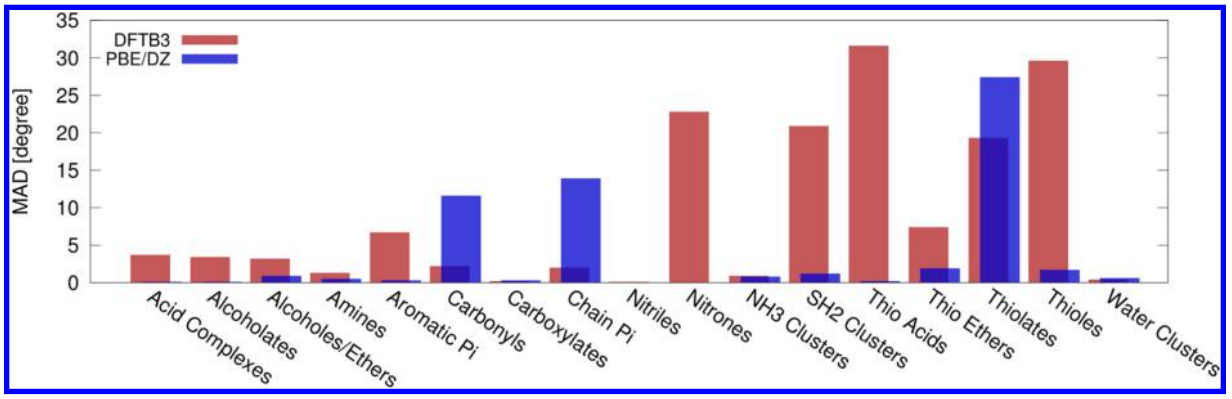


Figure 3. Accuracy of bonding angle in complexes of calcium with various organic functional groups reproduced by the current DFTB3 parametrization The deviations observed with PBE/def2-SVP calculations are shown for comparison. The reference data were obtained with DFT optimizations on the level B3LYP/def2-TZVPP using Turbomole 6.4. Errors expressed as mean absolute deviations (MAD) in degrees. The test systems contained in each category mentioned here are detailed further in the Supporting Information.

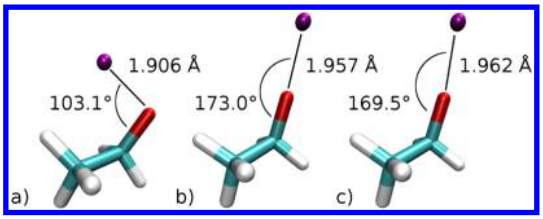


Figure 4. Bond distances and angles in the sodium ethanolate complex obtained from calculations involving different basis sets. (a) STO-3G,[50,51] (b) def2-SVP, and (c) def2-TZVPP. All of the optimizations were done using the PBE functional in Turbomole 6.4.

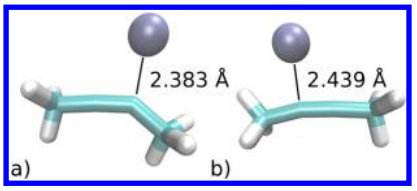


Figure 5. π-Interaction of 2-butyne with calcium. (a) DFTB3/3OB suggests the double bond and strong coordination to calcium instead of the carbon−carbon triple bond. (b) Gaussian B3LYP/cc-pVTZ as a reference.

Multiple tests have been performed to check the accuracy and robustness of the halogen bonding correction, using geometries from the X40 set.[3] The results, analyzed by different types of interactions are summarized in Table 8. The RMSD of interaction energies in the X40 set upon optimizing the geometries decreases from 3.1 to 1.9 kcal/mol upon adding the correction. This error includes the problematic hydrogen bonds that are underbound in some cases. With the X correction the RMSD of hydrogen bonds increases from 2.5 to 3.5 kcal/mol. For halogen bonds only, the RMSD decreases from 4.6 to 1.0 kcal/mol. The RMSD of single-point calculations over halogen bonds not included in the parametrization is 2.4 and 0.9 kcal/mol, including the interactions with sulfur for which we do not correct. This demonstrates that the correction is robust for halogen bonds and yields good results even for chemically different systems such as very strong halogen bonds of fluorinated compounds (complexes 16, 17, and 18). The description of some hydrogen bonds with hydrogen halogenides is worsened slightly upon the application of the correction. Although the hydrogen bonds affected by this are less likely to appear in relevant biological environments, the suitability of this correction should be considered carefully before applying it to larger and more complex systems.

All geometries from the X40 data set (no. 13 to 24) were optimized with the corrected DFTB3-D3X method. The RMSD is reduced from 0.312 to 0.132 Å. The geometry of most complexes is conserved, only the intermolecular distance and angles change slightly (see below). The only exception is the weakest complex (no. 13) of chloromethane and formaldehyde where the geometry changes to a more stable one defined by the dipole−dipole interaction. As before, the description of hydrogen bonds is slightly worsened with the halogen bond correction (RMSD increases from 0.119 to 0.149 Å). The hydrogen bond of system no. 40 (methanole and chloromethane) is unstable with DFTB3 (with as well as without the halogen bond correction) and rearranges to a more stable complex dominated by dispersion interactions. Excluding the two problematic systems (nos. 13 and 40) from the analysis we find a slight improvement in bonding distances with the halogen bond correction (mean absolute deviation decreases from 0.169 to 0.118 Å but a slightly increased RMSD of 0.148 Å from 0.129 Å).

Because the directional character of a halogen bond is determined by the presence of the σ hole,[21] it is important to know the intramolecular angular dependence is described in the DFTB framework. We compared the interaction energy profile along an angular scan in the methyl bromide···formaldehyde complex (no. 14 in X40 set). While the reference calculations yield an energy minimum at angle of 168.1°, DFTB3 predicts the complex to be almost linear at 178.6°. This deviation in the bonding angle can be explained by the simplified description of the oxygen atom lone pairs inherent to the monopole approximation in DFTB. Besides this small difference, DFTB3 describes the angular dependence correctly. Neither the DFTB3 electrostatics nor the dispersion can be the cause of the angular dependency of the halogen bond since both are only dependend on interatomic distances; the angle-dependent potential energy surface emerges from the energy contribution of the core DFTB Hamiltonian.

Finally, we tested how the correction affects intramolecular angles, namely, the repulsion of atoms in 1−4 position. We compared the geometries of 1,1-diodo-2-methoxyethene, optimized by corrected and uncorrected DFTB-D3 and a B3LYP/def2-QZVPP reference. The sum of the O−C−C and C−C−I angles is 242.8° in DFT-D3, uncorrected DFTB3

DOI: 10.1021/ct5009137
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 7. Performance of the Current Parameterization of Halogens Expressed As Mean Absolute Deviations (MAD) from Reference Data[a]

	DFTB3/3OB				PBE/def2-SVP			
property	F	Cl	Br	I	F	Cl	Br	I
bond distances [Å]	0.011	0.007	0.013	0.034	0.005	0.006	0.007	0.019
bond angles [deg]	1.1	1.8	2.7	1.8	0.5	0.4	0.4	0.5
at. energies [kcal/mol]	5.2	11.2	17.0	14.6	31.2	24.7	17.0	8.5
vib. freq. [cm^{-1}]	45.9	46.6	108.4	25.3	43.3	15.6	11.3	3.1

[a]The particular absolute values, types of references and numbers of test systems can be found in the Supporting Information. All benchmark properties were obtained from geometry optimizations with the respective method.

yields 237.3° and after the correction, the value is 240.7°. It is clear that, although hydrogen bonds are affected negatively by the repulsion, the X correction improves not only the halogen bonds themselves but the intramolecular interactions, too.

5. CONCLUSIONS

A new test set—the 'OrgX' set—has been created to benchmark performance of quantum-chemical methods in terms of geometries, atomization energies, and vibrational frequencies of 106 molecules containing halogens. The geometries were obtained from B3LYP calculations with triple-ζ basis sets, energies were obtained either from G3B3 (for the fluorine containing molecules) or CCSD(T) calculations (for chlorine up to iodine), and all vibrational frequencies were calculated with geometries optimized using BP86 and triple-ζ basis sets.

The 'BioMe' test set deals with the coordination of alkaline and alkaline earth metal ions to important functional groups in 210 bio-organic model complexes. Geometries were obtained in the same way as mentioned above, whereas the incremental ligand binding energies of small molecules were used instead of atomization energies.

Given these test sets, an extension of the 3OB parametrization for DFTB3 was carried out and the performance was assessed. Also, a comparison to PBE/def2-SVP calculations was made; this method was chosen because it can be applied to fairly large systems at reasonable computational cost, especially in molecular dynamics. The performance of the new DFTB3 parametrization is close to that of PBE concerning the geometries, but it is even outperforming it in terms of energies.

Table 8. Binding Energy Deviations (in kcal/mol) Compared to CCSD(T)/CBS Values of All Systems of the X40 Data Set[3] and Its Subgroups[a]

	DFTB3-D3X/3OB			PBE-D3/def2-SVP		
property	MSD	MAD	RMSD	MSD	MAD	RMSD
dipole−dipole	−0.1	0.1	0.2	−3.8	3.8	4.1
dispersion systems	0.1	0.5	0.9	−1.6	1.6	1.8
halogen bonds	0.0	0.8	1.0	−5.2	5.2	6.1
hydrogen bonds	2.3	2.3	3.5	−4.5	4.5	4.6
π systems	−1.1	1.1	1.1	−3.8	3.8	4.1
overall	0.3	1.1	1.9	−4.0	4.0	4.7

[a]All energies were calculated from geometries that were optimized with gradients obtained by the respective method. For both benchmarks the D3 dispersion correction was used, with DFTB3 the D3(BJ) parameters for 3OB with ζ = 4.0 was combined with the newly parameterized halogen bond correction (X correction). All energies were obtained from geometry optimizations with the respective method.

However, there are also a couple of limitations, which mostly are related to the use of a minimal basis. These include the bond distances in complexes of alkali ions with anionic oxygen or sulfur-containing compounds. The deviations in these worst cases are about 10% of the bond distances; it has to be noted that most applications of DFTB3 in organic and biological systems will be unaffected since the energetic description of the systems is fairly accurate. The deviations in bond angles may be less severe in actual applications because in condensed phase several ligands are attached to the alkali ion, restoring the correct coordination for many basis set issues seen in the small gas-phase test systems.

We also attempted to benchmark rock salts with the current 3OB parametrization for alkaline, alkaline earth metals, and halogens, but as mentioned before, metal−metal and metal−halogen binding energies were modified to be more repulsive. This avoids the formation of undesired complexes of ions in solution. Hence the performance of the new parameters is insufficient for solids, and at this point, no studies on these type of matrials should be conducted with 3OB alkaline metals, alkaline earth metals, and halogens.

Overall, the performance is sufficient and the parameters are applicable to biomolecular systems where both halogens and alkaline and alkaline earth ions play a vital role.

APPENDIX

To avoid breaking of any covalent bonds, the X correction interaction between sigma-holes and electron donors needs to be switched to a constant value so that the gradients vanish. We use a sum of van der Waals radii by Bondi[49] of the halogen bond atom pair a and b and define a switching interval from R_0 to R_1:

$$R_0 = 0.7R^{vdw} \quad \text{and} \quad R_1 = 0.8R^{vdw} \tag{6}$$

$$\text{with} \quad R^{vdw} = R_a^{vdw} + R_b^{vdw} \tag{7}$$

In order to achieve the desired behavior, we chose a polynomial that smoothly varies from 0 to 1 within the interval between R_0 and R_1:.

$$\sigma(x) = \begin{cases} 0 & \text{if} \quad R < R_0 \\ 1 & \text{if} \quad R > R_1 \\ -20x^7 + 70x^6 - 84x^5 + 35x^4 & \text{otherwise} \end{cases} \tag{8}$$

using a reduced bond distance $x = (R - R_0)/(R_1 - R_0)$. Renaming the energy term in eq 5 to $f_{ab}^{X}(R)$, the halogen bond correction energy reads.

DOI: 10.1021/ct5009137
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

$$E_{ab}^{X} = \frac{1}{2}\sum_{A,B}\sigma(x) f_{ab}^{X}(R) + \frac{1}{2}\sum_{A,B}[1-\sigma(x)] f_{ab}^{X}(R_0) \quad (9)$$

In order to obtain forces, gradients of the energy correction are needed. Introducing the notation.

$$\Delta s_i := \begin{cases} +(s_a - s_b) & \text{if} \quad i = a \\ -(s_a - s_b) & \text{if} \quad i = b \end{cases} \quad (10)$$

for the spatial coordinate s and with the derivatives of the repulsive energy term.

$$\frac{\partial f_{ab}^{X}}{\partial s_i}(R) = c_2 c_3 (R - d_{ab})^{c_3 - 1} f_{ab}^{X}(R) \frac{\Delta s_i}{R} \quad (11)$$

and the derivative of the switching function

$$\frac{\partial \sigma}{\partial s_i}(x) = \begin{cases} 0 & \text{if} \quad R < R_0 \text{or} R > R_1 \\ -140x^6 + 420x^5 - 420x^4 + 140x^3 & \text{otherwise} \end{cases} \quad (12)$$

we find the gradient of the halogen bond correction as

$$\frac{\partial E_{ab}^{X}}{\partial s_i} = \frac{1}{2}\sum_{A,B}\sigma(x)\frac{\partial f_{ab}^{X}}{\partial s_i}(R) + \frac{\partial \sigma}{\partial s_i}(x) f_{ab}^{X}(R) + \frac{1}{2}\sum_{A,B}\frac{\partial \sigma}{\partial s_i}(x) f_{ab}^{X}(R_0) \quad (13)$$

■ ASSOCIATED CONTENT

Supporting Information

Spreadsheet including all benchmark data mentioned in the text, including bond distances and angles, atomization energies, incremental ligand binding energies, vibrational frequencies, and X40 data set benchmarks. Different levels of theory used as references. All data needed to reproduce each repulsive potentials. Zip compressed file of geometries, featuring all optimized reference and DFTB3/3OB geometries in xyz file format. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

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

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank the Bremen Center for Computational Materials Science (BCCMS) of the university of Bremen for supplying the *slateratom* and *sktwocnt* programs to create electronic parameters for DFTB. J. Řezáč acknowledges the support of the Institute of Organic Chemistry and Biochemistry AS CR, Research Project RVO:61388963, and of the Czech Science Foundation (P208/12/G016).

■ REFERENCES

(1) Gaus, M.; Cui, Q.; Elstner, M. *J. Chem. Theory Comput.* **2011**, *7*, 931−948.
(2) Gaus, M.; Goez, A.; Elstner, M. *J. Chem. Theory Comput.* **2013**, *9*, 338−354.
(3) Řezáč, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2012**, *8*, 4285−4292.
(4) Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. *Phys. Rev. B* **1995**, *51*, 12947−12957.
(5) Seifert, G.; Porezag, D.; Frauenheim, T. *Int. J. Quantum Chem.* **1996**, *58*, 185−192.
(6) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260−7268.
(7) Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 5614−5621.
(8) Elstner, M.; Seifert, G. *Phil. Trans. R. Soc. A* 2014, *372*.
(9) Gaus, M.; Cui, Q.; Elstner, M. *WIREs Comput. Mol. Sci.* **2014**, *4*, 49−61.
(10) Cui, Q.; Elstner, M. *Phys. Chem. Chem. Phys.* **2014**, *16*, 14368−14377.
(11) Gaus, M.; Lu, X.; Elstner, M.; Cui, Q. *J. Chem. Theory Comput.* **2014**, *10*, 1518−1537.
(12) Lu, X.; Gaus, M.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2014**, DOI: 10.1021/jp506557r.
(13) Andreini, C.; Bertini, I.; Cavallaro, G.; Holliday, G.; Thornton, J. *JBIC, J. Biol. Inorg. Chem.* **2008**, *13*, 1205−1218.
(14) Lockless, S. W.; Zhou, M.; MacKinnon, R. *PLoS Biol.* **2007**, *5*, 1079−1088.
(15) Formaneck, M. S.; Ma, L.; Cui, Q. *J. Am. Chem. Soc.* **2006**, *128*, 9506−9517.
(16) Sirimulla, S.; Bailey, J. B.; Vegesna, R.; Narayan, M. *J. Chem. Inf. Model.* **2013**, *53*, 2781−2791.
(17) Lu, Y.; Liu, Y.; Xu, Z.; Li, H.; Liu, H.; Zhu, W. *Expert Opin. Drug Discovery* **2012**, *7*, 375−383.
(18) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(19) Kellö, V.; Sadlej, A. J.; Hess, B. A. *J. Chem. Phys.* **1996**, *105*, 1995−2003.
(20) Gaus, M.; Chou, C.-P.; Witek, H.; Elstner, M. *J. Phys. Chem. A* **2009**, *113*, 11866−11881.
(21) Clark, T.; Hennemann, M.; Murray, J.; Politzer, P. *J. Mol. Model.* **2007**, *13*, 291−296.
(22) Řezáč, J.; Hobza, P. *Chem. Phys. Lett.* **2011**, *506*, 286−289.
(23) Aradi, B.; Hourahine, B.; Frauenheim, T. *J. Phys. Chem. A* **2007**, *111*, 5678−5684.
(24) Elstner, M. *Theor. Chem. Acc.* **2006**, *116*, 316−325.
(25) Wahiduzzaman, M.; Oliveira, A. F.; Philipsen, P.; Zhechkov, L.; van Lenthe, E.; Witek, H. A.; Heine, T. *J. Chem. Theory Comput.* **2013**, *9*, 4006−4017.
(26) Kubař, T.; Bodrog, Z.; Gaus, M.; Köhler, C.; Aradi, B.; Frauenheim, T.; Elstner, M. *J. Chem. Theory Comput.* **2013**, *9*, 2939−2949.
(27) Řezáč, J.; Jurečka, P.; Riley, K. E.; Černý, J.; Valdes, H.; Pluháčková, K.; Berka, K.; Řezáč, T.; Pitoňák, M.; Vondrášek, J.; Hobza, P. *Collect. Czech. Chem. Commun.* **2008**, *73*, 1261−1270.
(28) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
(29) Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(30) Grimme, S. *Angew. Chem., Int. Ed.* **2013**, *52*, 6306−6312.
(31) Risthaus, T.; Grimme, S. *J. Chem. Theory Comput.* **2013**, *9*, 1580−1591.
(32) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.
(33) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479−483.
(34) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J. Chem. Phys.* **1997**, *106*, 9639−9646.
(35) Fabiano, E.; Della Sala, F. *Theor. Chem. Acc.* **2012**, *131*, 1278.
(36) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(37) Prascher, B. P.; Woon, D. E.; Peterson, K. A.; Dunning, J.; Thom, H.; Wilson, A. K. *Theor. Chem. Acc.* **2011**, *128*, 69−82.
(38) Koput, J.; Peterson, K. A. *J. Phys. Chem. A* **2002**, *106*, 9595−9599.
(39) TURBOMOLE V6.4 2012, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989−2007,

DOI: 10.1021/ct5009137
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.
(40) Treutler, O.; Ahlrichs, R. *J. Chem. Phys.* **1995**, *102*, 346−354.
(41) Von Arnim, M.; Ahlrichs, R. *J. Comput. Chem.* **1998**, *19*, 1746−1757.
(42) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283−290.
(43) Hättig, C.; Weigend, F. *J. Chem. Phys.* **2000**, *113*, 5154−5161.
(44) Hättig, C.; Hellweg, A.; Köhn, A. *J. Am. Chem. Soc.* **2006**, *128*, 15672−15682.
(45) Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **1999**, *110*, 7650−7657.
(46) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *126*, 084108.
(47) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *127*, 124105.
(48) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; J. R. Cheese-man, Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; M. Cari-cato, 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, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, 2009.
(49) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441−451.
(50) Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657−2664.
(51) Hehre, W. J.; Ditchfield, R.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1970**, *52*, 2769−2773.

DOI: 10.1021/ct5009137
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX