This is an open access article published under an ACS AuthorChoice License, which permits copying and redistribution of the article or any adaptations for non-commercial purposes.

JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

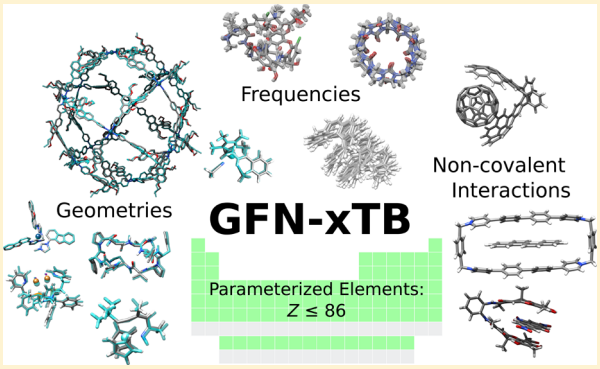
A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1−86)

Stefan Grimme,* Christoph Bannwarth, and Philip Shushkov

Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Universität Bonn, Beringstr. 4, D-53115 Bonn, Germany

Supporting Information

ABSTRACT: We propose a novel, special purpose semiempirical tight binding (TB) method for the calculation of structures, vibrational frequencies, and noncovalent interactions of large molecular systems with 1000 or more atoms. The functional form of the method is related to the self-consistent density functional TB scheme and mostly avoids element-pair-specific parameters. The parametrization covers all spd-block elements and the lanthanides up to Z = 86 using reference data at the hybrid density functional theory level. Key features of the Hamiltonian are the use of partially polarized Gaussian-type orbitals, a double-ζ orbital basis for hydrogen, atomic-shell charges, diagonal third-order charge fluctuations, coordination number-dependent energy levels, a noncovalent halogen-bond potential, and the well-established D3 dispersion correction. The accuracy of the method, called Geometry, Frequency, Noncovalent, eXtended TB (GFN-xTB), is extensively benchmarked for various systems in comparison with existing semiempirical approaches, and the method is applied to a few representative structural problems in chemistry.



1. INTRODUCTION

In quantum chemistry, geometry optimizations, vibrational frequency calculations, and molecular dynamics simulations are currently dominated by Kohn−Sham density functional theory (DFT) for system sizes up to a few hundred atoms. These computations serve as a starting point for investigations of various spectroscopic and/or thermochemical properties, possibly with higher-level wave function theory (WFT) methods. For fairly large systems with 1000 or more atoms, however, even simplified schemes such as the PBEh-3c hybrid DFT[1,2] and Hartree−Fock/minimal basis set (HF-3c)[3] composite methods, which we have proposed in the past few years, become computationally unfeasible. For instance, many interesting proteins with 2000−5000 atoms are out of reach for routine DFT optimizations, and such investigations are conducted at the force-field (FF) level.[4−6] The disadvantages of the FF treatment are numerous, including difficulties with metals, proton transfer and protonation states, polarization and chemical reactions, and can only be circumvented by a quantum mechanical (QM) treatment of the electrons.

Semiempirical quantum mechanical (SQM) methods[7−9] provide a well-known alternative route because they are at least 2 orders of magnitude faster than conventional DFT treatments. SQM methods approximate single-reference Hartree−Fock (HF) or first-principles DFT theory and have been investigated extensively in the 1980s and 1990s. The development of SQM methods has recently seen a renewed interest triggered primarily by the advent of the density functional tight binding (DFTB) method pioneered by Seifert, Elstner, and Frauenheim.[10−15] In the most recent self-consistent charge (SCC) form with third-order charge fluctuation terms,[12] the method is relatively accurate for a wide range of chemical systems[16−22] and has been implemented in several quantum chemical codes.[23−27] SQM methods derived from HF theory neglect to a varying degree the differential overlap between atomic basis functions, which results in vanishing classes of one- and two-electron integrals and thus substantial computational savings. Depending on the extent of approximations, these SQM methods fall into various categories such as complete or intermediate neglect of differential overlap (CNDO or INDO). We refer the reader to ref 9 for an excellent review of SQM methods, including noncovalent interactions (NCI). Other notable recent developments in the field of SQM methods include modifications of PM6[28−33] and OM2[34−38] and the development of PM7,[39] INDO/X,[40] and

Received: February 4, 2017
Published: April 18, 2017

Downloaded via 62.153.224.77 on November 11, 2024 at 13:25:45 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

© 2017 American Chemical Society

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009

MSINDO-sCIS.[41] For a recently proposed semiempirical minimal-basis scheme in between DFT and DFTB employed in embedding approaches, see ref 42. Recently we have also proposed an extended special purpose TB variant, which uses an augmented atomic orbital (AO) basis set, for the fast computation of electronic spectra.[43] The encouraging performance of this method, called sTDA-xTB, has stimulated the current work.

The objective of the paper is to propose a new semiempirical TB method, called GFN-xTB, where GFN indicates the design of the approach to yield reasonable Geometries, vibrational Frequencies, and Noncovalent interactions, and 'x' stands for extensions in the AO basis set and the form of the Hamiltonian. In general, GFN-xTB provides for molecules from the whole periodic table higher accuracy for the target properties than existing 'general-purpose' semiempirical approaches. It is applicable to a wider range of systems and is computationally and numerically more robust than other schemes with comparable accuracy. The premise is that SQM methods cannot provide the same level of accuracy for a broad range of chemical properties, such as molecular structures and chemical reaction energies. Thus, the compromise established in conventional SQM methods is often inaccurate and likely limits their use in many applications. This has led to the poor reputation of SQM methods in the 1990s, which were eventually surpassed by the successful performance of DFT. Special purpose SQM methods that aim at a specific chemical or physical property from the outset have been pioneered by Truhlar and co-workers for computing reaction dynamics,[44] atomic charges[45] (see also ref 46), and solvation free energies.[47] Along these lines, we found recently that certain SQM methods can be useful in gas-phase reaction dynamics to simulate electron-impact mass spectra (QCEIMS),[38,48] and the applications to this problem have fueled the current development.

The method described herein has the following characteristics:

- Structural properties around equilibrium, such as vibrational frequencies and noncovalent interactions, are the main target quantities. These are difficult to obtain with ab initio procedures for systems with more than 1000 atoms and are important in many chemical problems. Chemical reaction energies are not part of the training set and are thus less accurate than with general purpose SQM methods. However, some relevant thermochemical properties outside of the training set (dissociation energies, relative proton affinities (PA), and relative ionization potentials (IP)) are obtained reasonably accurately such that potential energy surfaces (PES) are useful for high-temperature molecular dynamics calculations in modeling mass spectra. Furthermore, reasonable enthalpies of formation can be obtained by correcting the computed atomization energies with a sum over atom-specific terms. The accuracy for conformational energies is comparable or better than existing SQM methods.
- Most semiempirical methods employ minimal AO basis sets, which limit the accuracy and can lead sometimes to unphysical results, especially for heavier elements. The GFN-xTB basis set consists of a minimal basis set of atom centered, approximate Slater functions augmented with a second s-function for hydrogen to improve the description of hydrogen bonding and d-polarization functions for higher row elements to facilitate the description of hypervalent bonding arrangements. The Slater functions are approximated as contractions of standard primitive Gaussian functions (STO-*m*G).[49]
- The number of parameters is kept to a minimum. Thus, pair-specific potentials, which are characteristic of DFTB, are avoided, and mainly global and element-specific parameters are employed. For example, for each of the second row elements B−F, only 11 parameters need to be determined, the two atomic energy levels (2s/2p) and corresponding Slater AO exponents, two repulsion parameters, two off-diagonal Hamiltonian polynomial factors, two chemical hardness parameters, and one shell-charge scaling parameter. The method is therefore easy to parametrize using reference data generated by hybrid DFT calculations. This parametrization scheme is flexible and can be adjusted to specific systems or chemical properties.
- The method is applicable to diverse chemical systems composed of elements from different corners of the periodic table and ranging from organic molecules to inorganic main group compounds and (transition-)metal complexes. Biochemical systems like (metallo-)proteins, RNA and DNA, and their inter- and intramolecular interactions are of primary interest. Electronically degenerate (multiconfigurational) situations are handled by a finite electronic temperature approach,[50] which allows the qualitatively correct description of biradicals, orbital-symmetry forbidden reactions, and dissociation processes.[38,51]

The paper is organized as follows: A detailed description of the GFN-xTB Hamiltonian is provided in Section 2.1, and the parametrization procedure is outlined in Section 2.2. The GFN-xTB performance for structures and NCIs is compared to contemporary semiempirical methods in Section 3.

2. THEORY

2.1. The GFN-xTB Hamiltonian. In the following, atomic units are used, if not stated otherwise. The total GFN-xTB energy expression consists of electronic (el), atom pairwise repulsion (rep), dispersion (disp), and halogen-bonding (XB) terms:

$$E = E_{\mathrm{el}} + E_{\mathrm{rep}} + E_{\mathrm{disp}} + E_{\mathrm{XB}} \tag{1}$$

The electronic energy E_{el} is given by

$$E_{\mathrm{el}} = \sum_{i}^{\mathrm{occ.}} n_i \langle \psi_i | H_0 | \psi_i \rangle + \frac{1}{2} \sum_{\mathrm{A,B}} \sum_{l(\mathrm{A})} \sum_{l'(\mathrm{B})} p_l^{\mathrm{A}} p_{l'}^{\mathrm{B}} \gamma_{\mathrm{AB},ll'} + \frac{1}{3} \sum_{\mathrm{A}} \Gamma_{\mathrm{A}} q_{\mathrm{A}}^3 - T_{\mathrm{el}} S_{\mathrm{el}} \tag{2}$$

where ψ_i are the valence molecular orbitals (MOs) with occupation numbers n_i (which may be fractional, see eq 6), and H_0 is the zero-order Hamiltonian. The electronic free energy $T_{\mathrm{el}}S_{\mathrm{el}}$ term is discussed below. The SCC contributions are represented by a second-order term and a third-order diagonal contribution (the second and third term in eq 2), where q_{A} is the Mulliken charge of atom A and Γ_{A} is the charge derivative of the atomic Hubbard parameter η_{A}. The summation in the second-order term is over all shells l and l' located on atoms A

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009

and B with p_l^A being the charge distributed over the orbital shells with angular momenta l on atom A:

$$p_l^A = p_l^{A_0} - \sum_{\nu}^{N_{AO}} \sum_{\mu \in A, \mu \in l} S_{\mu\nu} P_{\mu\nu} \tag{3}$$

where N_{AO} is the total number of AOs and $p_l^{A_0}$ is the reference shell occupation of the free atom (e.g., for carbon $p_{2s}^{A_0} = 2$, $p_{2p}^{A_0} = 2$). Instead of the ground-state shell occupations of the free atoms, the canonical ones are chosen that result from the 'Aufbau' principle together with ignoring well-known exceptions (e.g., for Cr we take a $3d^4 4s^2$ instead of the ground-state $3d^5 4s^1$ occupation). The shell atomic charges sum to the partial atomic charges, $\sum_{l \in A} p_l^A = q_A$. Please refer to refs 52 and 53 for more details of the DFTB treatment of d-block elements. The third-order density fluctuations terms are described in refs 11 and 12. Our approach is restricted to a diagonal third-order term, which reduces the numerical complexity but keeps the most essential part. The combination of l-dependent second-order and diagonal third-order density fluctuation terms is used for the first time in a DFTB context.

The distance dependence of the Coulomb interaction is described by the generalized Mataga–Nishimoto–Ohno–Klopman[54−57] formula, which is given by

$$\gamma_{AB,ll'} = \left(\frac{1}{R_{AB}^{k_g} + \eta^{-k_g}} \right)^{1/k_g} \tag{4}$$

where R_{AB} is the interatomic distance, k_g is a global parameter, and the average chemical hardness η of the two atoms A and B is defined as

$$\eta = 2 \left(\frac{1}{(1 + \kappa_A^l)\eta_A} + \frac{1}{(1 + \kappa_B^{l'})\eta_B} \right)^{-1} \tag{5}$$

where η_A and η_B are taken as element-specific fit parameters, and the parameters κ_A^l and $\kappa_B^{l'}$ are element-specific scaling factors for spd levels.

In case of fractional orbital occupations, an electronic temperature times entropy term, $-T_{el}S_{el}$ enters the electronic energy (see eq 2). The electronic temperature serves as an adjustable parameter (default: T_{el} = 300 K), and the so-called Fermi smearing is employed to achieve fractional occupations for systems with near-degenerate orbital levels. The orbital occupations for the spin orbital ψ_i are given by[50]

$$n_i(T_{el}) = \frac{1}{\exp[(\epsilon_i - \epsilon_F)/(k_B T_{el})] + 1} \tag{6}$$

where ϵ_i is the orbital energy of the spin orbital ψ_i and ϵ_F is the Fermi level. Higher temperatures may be used in certain cases, e.g., to investigate thermally forbidden reaction pathways (see Section 3.4). In this way, an electronic free energy, including an entropy contribution, is variationally minimized in the SCC procedure. This should not be confused with a true thermostatistical treatment because the Fermi smearing technique is used only as a way to take static correlation into account.

Expressing the molecular orbitals ψ_i in eq 2 as a linear combinations of atom-centered orbitals (LCAO):[58,59]

$$\psi_i = \sum_{\mu}^{N_{AO}} c_{\mu i} \phi_\mu(\zeta, \text{STO} - m\text{G}) \tag{7}$$

and variational minimization of the energy expression in eq 2 with respect to the linear coefficients $c_{\mu i}$ leads to the general eigenvalue problem

$$\mathbf{FC} = \mathbf{SC}\epsilon \tag{8}$$

where $\mathbf{F}$ denotes the approximate TB Fock/Kohn–Sham matrix in the AO basis, $\mathbf{S}$ is the AO overlap matrix, and $\mathbf{C}$ are the MO coefficients with eigenvalues ϵ. For reasons of computational efficiency and robustness, open-shell systems are treated in a restricted manner by a fractional occupation of a single set of spatial orbitals (see below). The spherical Slater-type AOs ϕ are approximated by Stewart's Gaussian expansions[49] with m = 4–6 primitives per AO. They are available up to principal quantum number n = 6 (for l = 2 up to n = 5). The exponents ζ are element-specific, fitted parameters. A description of the AO basis is given in Table 1. For most main group elements of third and higher rows, which can form hypervalent electronic structures, an additional single d-polarization function is provided. The same holds for earth alkaline metals, which require d-functions in certain bonding situations (e.g., bent MX_2 molecules where X is a halogen). Polarization functions on elements H to F do not substantially improve the results and increase the computational cost considerably (by a factor of 5–10). For the group IIb elements, Zn to Hg, the d-electrons are not treated as valence (sp-shell only). As we do not focus on magnetic and electronic spectroscopic properties, the lanthanides are treated as 4d transition metals without explicit consideration of the f-electrons (i.e., all lanthanides have three valence electrons). This 'f-in-core' approximation is rather accurate according to ab initio pseudopotential calculations.[60] The actinides, for which this approximation is not possible, are presently not covered.

Table 1. Description of the Slater-Type AO Basis Sets[a]

element	basis functions
H	ns, $(n+1)$s
He	ns
Be–F, group I, Zn, Cd, Hg, Tl–Bi	$n$sp
groups II–VIII	$n$sp, $(n+1)$d
transition metals and lanthanides	$(n-1)$d, $n$sp

[a]The n denotes the principal quantum number of the valence shell of the element.

The matrix elements of the GFN-xTB Hamiltonian are computed similarly to the second- and third-order SCC-DFTB3 methods:[10,11,52]

$$\langle \phi_\mu | F | \phi_\nu \rangle = \langle \phi_\mu | H_0 | \phi_\nu \rangle + \frac{1}{2} S_{\mu\nu} \sum_{C} \sum_{l''} (\gamma_{AC,ll''} + \gamma_{BC,l'l''}) p_{l''}^C + \frac{1}{2} S_{\mu\nu} (q_A^2 \Gamma_A + q_B^2 \Gamma_B) \quad (\mu \in l(A), \nu \in l'(B)) \tag{9}$$

where indices μ and ν label the AOs with corresponding angular momenta l and l' and the second sum runs over all atoms C and their shells l''.

The H_0 elements are given by

$$\langle \phi_\mu | H_0 | \phi_\nu \rangle = K_{AB} \frac{1}{2}(k_l + k_{l'}) \frac{1}{2}(h_A^l + h_B^{l'}) S_{\mu\nu} (1 + k_{EN} \Delta EN_{AB}^2) \Pi(R_{AB,ll'}) \quad (\mu \in l(A), \nu \in l'(B)) \tag{10}$$

where k_l and $k_{l'}$ are the well-known 'Hückel' constants taken as free parameters for each angular momentum (l for A, l' for B)

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989–2009

and h_A^l and $h_B^{l'}$ are effective atomic energy levels. For s−p interactions and those with the hydrogen 2s function, two parameters k_{sp} and $k_{2s,H}$ replace the average $\frac{1}{2}(k_l + k_{l'})$ in eq 10. The electronegativity difference of the two atoms, $\Delta EN_{AB} = EN_A - EN_B$, computed with standard Pauling values modifies the matrix elements with the global parameter k_{EN} as a coefficient of proportionality. K_{AB} is a scaling constant, which differs from unity only for a limited number of element pairs. We observed that after the optimization of element-specific and global parameters, the bond strength between certain elements was slightly unbalanced, and we introduced eight element-pair-specific and intermetallic K_{AB} parameters (see Table 2). We would like to emphasize that we avoided element-pair-specific parameters in general and introduced the ones listed in Table 2 to fine-tune the accuracy for properties such as conformational energies, which were significantly improved by reducing the value of K_{HH}. In our opinion, the element-specific parameters form a reasonably accurate parametrization suitable for general chemistry, and users can customize the Hamiltonian by varying certain K_{AB} parameters.

Table 2. Global Empirical Parameters Defining the GFN-xMethod[a]

parameter	value
k_s	1.85
k_p	2.25
k_d	2.00
k_{sp}	2.08
$k_{2s,H}$	2.85
k_g	2.0
k_f	1.5
$k_{CN,s}$	0.006
$k_{CN,p}$	−0.003
$k_{CN,d}$	−0.005
k_{EN}	−0.007
a_1(D3)	0.63
a_2(D3)	5.0
s_8(D3)	2.4
k_{XR}	1.3
k_{X2}	0.44
K_{HH}[b]	0.96
K_{BH}[b]	0.95
K_{NH}[b]	1.04
K_{PB}[b]	0.97
K_{SiB}[b]	1.01
K_{NiH}[b]	0.90
K_{ReH}[b]	0.80
K_{PtH}[b]	0.80
K_{3d-3d}[b,c]	1.10
$K_{4d/5d/4f-4d/5d/4f}$[b,c]	1.20

[a]The parameters are either dimensionless or in atomic units. [b]Special H_0 scaling paramaters K_{AB} (see eq 10) for certain element pairs and inter transtion-metal intactions (for 3d and 4*d*/5d/4f block elements). [c]$K_{3d-4d/5d/4f} = \frac{1}{2}(K_{4d/5d/4f-4d/5d/4f} + K_{3d-3d})$.

$\Pi(R_{AB,ll'})$ is a distance and *l*-dependent function given by

$$\Pi(R_{AB,ll'}) = \left(1 + k_{A,l}^{poly}\left(\frac{R_{AB}}{R_{cov,AB}}\right)^{1/2}\right)\left(1 + k_{B,l'}^{poly}\left(\frac{R_{AB}}{R_{cov,AB}}\right)^{1/2}\right) \quad (11)$$

where $R_{cov,AB} = r_{cov,A} + r_{cov,B}$ is the covalent distance, r_{cov} is the covalent atomic radius,[61] and $k_{A,l}^{poly}$ and $k_{B,l}^{poly}$ are element-specific parameters. Finally, the effective atomic energy levels of nonmetals, h_A^l, depend linearly on the geometric D3 coordination number, *CN*,[62] of atom A:

$$h_A^l = H_A^l(1 + k_{CN,l}CN_A) \quad (l \in A) \quad (12)$$

where $k_{CN,l}$ are three global scaling parameters for spd shells. The dependence of the zero-order Hamiltonian on $\Pi(R_{AB,ll'})$, ΔEN, and CN_A/CN_B, with the latter two contributions specific to our TB scheme and at variance with the widely used DFTB3, introduces more flexibility in the covalent binding part without the need for element pair-specific parametrization. A larger ΔEN will reduce the extent of 'covalent' binding and effectively increase the ionic character for element pairs with a nonzero ΔEN. The dependence of h_A^l on the coordination number CN_A changes the energetic gap between the different *l* shells of atom A. This is of particular importance for the s and p levels as this affects the hybridization in different bonding situations (thus the dependency on the CN_A).

For the repulsion energy in eq 1, we employ an atom pairwise potential similar to the one proposed in ref 63:

$$E_{rep} = \sum_{AB} \frac{Z_A^{eff} Z_B^{eff}}{R_{AB}} e^{-(\alpha_A\alpha_B)^{1/2}(R_{AB})^{k_f}} \quad (13)$$

where Z^{eff} are effective nuclear charges, k_f is a global parameter, and α are element-specific parameters. Z^{eff} are fitted parameters, and the optimized values deviate by <20−30% from the nuclear charge, *Z*, for most elements. An exception from this general trend is the lanthanides, for which the difference is about 50%.

The dispersion energy in eq 1 is computed by the well-established D3 method[62] in the BJ-damping scheme[64] without three-body terms (see ref 65 for a recent review). The two short-range damping parameters a_1 and a_2, and also s_8 are adjusted to yield good noncovalent interaction energies and geometries.

Halogen-bonding (XB) interactions are described poorly[9] by the point charge approximation of electrostatic interactions in all standard DFTB schemes. This holds also, e.g., for PM6, for which corrections have been developed to account for the deficiencies.[32,66] We use a pairwise repulsive correction of a modified Lennard-Jones form:

$$E_{XB} = \sum_{XB} f_{dmp}^{AXB} k_X \left(1 + \left(\frac{R_{cov,AX}}{R_{AX}}\right)^{12} - k_{X2}\left(\frac{R_{cov,AX}}{R_{AX}}\right)^6\right) \Big/ \left(\frac{R_{cov,AX}}{R_{AX}}\right)^{12} \quad (14)$$

where $R_{cov,AX} = k_{XR}(r_{cov,A} + r_{cov,X})$ is an effective covalent distance and k_{XR} and k_{X2} are global parameters. The correction takes into account the halogens Br, I, and At, and N or O atoms are the donors B. The atom A closest to the halogen defines the AXB angle θ. The damping function f_{dmp}^{AXB} is given by

$$f_{dmp}^{AXB} = \left(\frac{1}{2} - \frac{1}{4}\cos\theta_{AXB}\right)^6 \quad (15)$$

such that the correction vanishes for nonlinear arrangements, and the magnitude of the interaction is determined by the parameter k_X.

The values of 16 global parameters for GFN-xTB, as obtained from a nonlinear fit to reference data (see Section 2.2), are gathered in Table 2. The roughly 1000 element-specific parameters for the periodic table up to $Z = 86$ are listed

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009

in the Supporting Information. This intermediate degree of parametrization is typical for many SQM methods, such as PM6[28] or MSINDO,[67] which rely on element-specific parameters. The K_{AB} values deviating from one are given in Table 2.

The new method is implemented in a standalone general quantum chemical program, called `xtb`. It features analytical nuclear gradients, shared memory parallelization, an efficient geometry optimizer[68] based on approximate normal coordinates (ANC) from Lindh's model Hessian,[69] molecular dynamics, conformational searches, and reaction path optimization modules. It has been tested in full protein structure optimizations with up to 3000 atoms, a few hundred ps dynamics for about 500 atoms, and in single-point computations of the fractional orbital density (FOD)[51] for about 10,000 atoms. In addition, an efficient and computationally inexpensive solvation model supplemented with analytical nuclear gradients is developed. It is based on the generalized Born model for the dielectric response of the solvent,[70] a term proportional to the solvent accessible surface area (SA),[71] and a short-range hydrogen-bonding contribution. Details of the new solvation model together with applications to (bio)chemical systems will be provided in a separate publication.[72] The complete program can be obtained upon request.[73]

2.2. Parametrization and Technical Details. The parametrization involved the determination of the global parameters in Table 2 and the element-specific parameters. To keep the total number of parameters to a minimum, the values for the lanthanides were obtained by linear interpolation with the nuclear charge Z, where only the start (Z = 58, Ce), and end points (Z = 71, Lu) were fitted.

The parameters were determined by a minimization of the root-mean-square deviation (RMSD) between calculated and reference data using the Levenberg−Marquardt algorithm.[74,75] Most molecular structures were optimized at the PBEh-3c hybrid DFT level,[1] which yields very accurate equilibrium (R_e) structures for a wide range of systems. PBE0-D3(BJ)/def2-TZVP[76,77] was the reference structure level for the rare gas systems. Global and atomic parameters were simultaneously optimized for the elements H, C, N, O. The rest of the atomic parameters were optimized element-wise while keeping all existing parameters fixed. The fitting for the periodic table was thus continued with the halogens, after which elements were treated in more or less canonical order while trying to minimize interdependencies. About 50−100 reference data points (see below) were used on average per fitted parameter. The reference molecules were predominantly of closed-shell character and covered common bonding situations. The human work time required to parametrize an element, including the generation of reference molecular structures, was about 1−2 days.

Five types of reference data were employed in the parametrization: (a) equilibrium structures, (b) distorted geometries with energies with respect to the equilibrium of a few kcal mol^{-1}, (c) harmonic vibrational frequencies (mainly for small molecules with <10 atoms), (d) CM5[78] atomic charges, and (e) noncovalent interaction (NCI) energies and structures. For (a) and (b) a force matching procedure was applied where all atomic force components of the reference molecule were fitted. This procedure was not sufficient to determine the NCI related parameters. Therefore, full structure optimizations were carried out for a few small complexes from the S22 benchmark set,[79,80] and the all-atom-structure RMSD from the reference was included in the fit. Spline-interpolated CCSD(T) data (minima and interaction energies) for the S66 × 8 NCI benchmark set[81] and interaction energies for the L7 set[82] were also used. The Mulliken charge-derived CM5 atomic charges were obtained by GFN-xTB as described previously.[43] The charges were used to stabilize the fit and to detect physically inconsistent parameter sets. Their relative contribution to the training set RMSD was within about 10% or less, which was smaller than the contribution of parts (a−c). It was observed during optimization that reasonably accurate charges (mean relative deviation from absolute PBE0/def-TZVP based reference values <10−20%) were obtained when reference structures and frequencies were described well. All global parameters in Table 2 except those related to halogen bonding were obtained by fitting to the HCNO training set. It comprised about 260 molecules ranging from diatomics to systems with about 100 atoms. The k_{XR} and k_{X2} halogen-bond parameters were determined by fitting to energies and forces of the XB18 benchmark set.[83] The GFN-xTB harmonic vibrational frequencies were not scaled and were fitted to the corresponding PBEh-3c values scaled by 0.95.[1]

We used the TURBOMOLE suite of programs[84−86] (version 7.0) to conduct most of the ground-state DFT calculations and geometry optimizations. For the calculation of the Hirshfeld[87] based CM5 charges at the PBE0/def-TZVP level the ORCA code[88] was used. We employed standard exchange−correlation functional integration grids (*grid*2 and *finalgrid*4 in ORCA and *m*4 in TURBOMOLE if not noted otherwise), typical self-consistent field (SCF) convergence criteria (10^{-7} E_h), and the resolution of the identity (RI) integral approximation[89−91] in the DFT calculations.

Calculations for comparisons were conducted with the `DFTB+`[23] (DFTB3[12] with the 3OB parametrization[92−94]), `MOPAC16`[95] (PM6-D3H4X),[28,31,32] `MSINDO`[67] (version 3.6), and `MNDO99`[96] (OM2)[34] codes. The DFTB3 method was used in its self-consistent version, which we refer to as DFTB3. Other DFTB parametrizations[97] comprising some 3d transition-metal elements[98] are available; however, since we are focusing on organic systems, we employed the 3OB parametrization[92−94] throughout. PM6-D3H4X is applied as implemented (with the identical keyword) in the `MOPAC16` code with the D3 correction in the zero-damping scheme.[31,62] The electronic energies obtained from DFTB3, MSINDO, and OM2 were corrected by the D3[62] approach using the rational/Becke−Johnson damping function.[99] For clarity, we use the suffix "-D3(BJ)" for these rationally damped schemes. The same correction is used in GFN-xTB, which is an integral part of the new method, and therefore no suffix is used here. For MSINDO, we used the neglect of diatomic differential overlap (NDDO) Hamiltonian and augmented it with the H+ hydrogen-bond correction[100] (denoted by the "H+" suffix) as implemented in the `MSINDO` code. The D3(BJ) and H+ parameters for DFTB3-D3(BJ), OM2-D3(BJ), and MSINDO-D3(BJ)H+ were taken from refs 38, 101, and 16 and are listed in the Supporting Information. Our standalone `dftd3` code[102] was used for the calculations of the D3(BJ) corrections. As in GFN-xTB, the Fermi smearing technique (T_{el} = 300 K) was employed in the DFTB3 and MSINDO procedures, which did not affect energies or structures of most of the systems considered here. Structures of systems that do not belong to established benchmark sets were compared to PBEh-3c,[1] if not stated otherwise. In these cases, the grid in TURBOMOLE was

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009

reduced to *m*3, and *g*-functions were removed from the auxiliary basis sets for the PBEh-3c calculations.

3. RESULTS AND DISCUSSION

In this section, the performance of the GFN-xTB approach is benchmarked for structures, noncovalent interaction energies, and conformational energies. Existing SQM methods are assessed on the same sets and compared to the performance and computational cost of GFN-xTB. The discussion continues with an exploration of characteristic features of a number of PESs and concludes with a demonstration of the efficiency of GFN-xTB in geometry optimizations of large chemical systems.

3.1. Structures. The geometry optimization of structures is one of the most important tasks for SQM methods, in particular for large systems, where it becomes prohibitively expensive at a first-principles level such as DFT. We assess the performance of GFN-xTB for the computation of structures of systems consisting of main group elements and some organometallic complexes.

3.1.1. Organic and Main Group Molecules. We choose ROT34, LB12, and HMGB11 test sets to assess the performance for molecular structures of systems containing only main group elements. The ROT34[103−105] set consists of the experimentally derived rotational constants, B_e, of 12 small organic molecules.

The ROT34 set is a sensitive measure for the accuracy of chemical structures because small changes in the bond lengths and angles can result in significant deviations from accurately measured rotational constants. Overestimated values indicate underestimated bond lengths, owing to the reciprocal dependence of B_e on the moment of inertia. The performance of GFN-xTB is compared to other dispersion-corrected semiempirical methods, including DFTB3-D3(BJ),[12,92,101] HF-3c,[3] MSINDO-D3(BJ)H+,[16,67] PM6-D3H4X,[28,31,32] and OM2-D3(BJ),[34,38] and the results are listed in Table 3. Well-performing density functional methods, such as PBE0-D3(BJ)/def2-TZVP,[62,76,77,99,106,107] yield a mean unsigned relative deviation (MURD) and standard relative deviation (SRD) of less than 0.5%.[1,103,104] The error obtained for semiempirical methods is naturally larger. The performance of GFN-xTB is very good, owing partly to the fact that a few systems of this set were part of the training set. The MURD of GFN-xTB is the lowest one (1.1%), even outperforming HF-3c. There is no large systematic shift of the bond lengths with a mean relative deviation (MRD) of only 0.4%. DFTB3-D3(BJ) gives a slightly lower SRD than GFN-xTB at the expense of systematically elongated bonds. This behavior has been discussed before[1,103] and is shared by DFTB, (semi)local density functionals, and HF-3c. The NDDO-based methods PM6-D3H4X, MSINDO-D3(BJ)H+, and OM2-D3(BJ) show MURDs and SRDs ≥ 2.5% and give less reliable geometries compared to DFTB3-D3(BJ), HF-3c, and GFN-xTB.

The LB12[1] (a set of 12 molecules each containing a single long bond between two atoms) and HMGB11[1] (11 heavy main group bond) sets are sensitive to the performance for heavier elements and bonding features that were not included in the training set. The molecules in each set can be found in ref 1 (see Supporting Information for detailed results). We restrict the comparison to GFN-xTB, PM6-D3H4X, MSINDO-D3(BJ)H+, and HF-3c, which have available parameters for all elements. The semiempirical methods underestimate the bond lengths in both LB12 and HMGB11 (see Table 3) sets. GFN-xTB produces the smallest mean absolute deviation (MAD) in both sets (MAD = 13.2 and 3.0 pm) surpassing PM6-D3H4X, MSINDO-D3(BJ)H+, and even HF-3c. The GFN-xTB mean deviation (MD) is close to zero in the HMGB11 set, and the bond lengths in the LB12 set are on average underestimated by 10.9 pm. The errors are systematic because the standard deviation (SD) is the lowest of all methods (SD = 18.8 and 3.9 pm). HF-3c performs similarly and only slightly worse than GFN-xTB (MADs and SDs are larger by 1−3 pm and 1−5 pm). PM6-D3H4X gives two outliers in the LB12 set, where the metal−metal bond dissociates during optimization, and these values are excluded from the statistical analysis.

Table 3. Comparison to Experimental Data of Optimized Geometries with Different Semi-Empirical Methods[a]

	GFN-xTB	PM6-D3H4X	MSINDO-D3(BJ)H+	OM2-D3(BJ)	DFTB3-D3(BJ)	HF-3c
ROT34 (Deviations in %)[b]						
MRD:	0.4	−1.6[c]	1.8	1.6[c]	−1.3	−1.4
MURD:	1.1	2.5[c]	3.1	2.3[c]	1.5	1.4
SRD:	1.5	2.5[c]	3.7	2.8[c]	1.3	1.0
MAXR:	5.8	6.1[c]	11.7	8.5[c]	4.6	4.5
LB12 (Deviations in pm)[d,h]						
MD:	−10.9	−5.5[e]	−17.6[f]	–	–	−5.2
MAD:	13.2	20.5[e]	27.7[f]	–	–	14.1
SD:	18.8	30.0[e]	31.9[f]	–	–	23.6
MAX:	55.7	64.2[e]	74.6[f]	–	–	61.1
HMGB11 (Deviations in pm)[g,h]						
MD:	−0.5	−3.8	−12.0	–	–	5.5
MAD:	3.0	10.1	18.3	–	–	5.8
SD:	3.9	17.3	25.7	–	–	5.3
MAX:	8.5	42.6	71.7	–	–	15.7

[a]Relative deviations in the rotational constants of medium-sized molecules (ROT34) as well as deviations in the bond lengths of very long intramolecular bonds (LB12) and covalent bonds of heavy main group elements (HMGB11) are considered. MRD = mean relative deviation, MURD = mean unsigned relative deviation, SRD = standard relative deviation, MAXR = maximum unsigned relative deviation, MD = mean deviation, MAD = mean absolute deviation, SD = standard deviation, and MAX = maximum absolute deviation. [b]Rotational constants B_e (excluding vibrational effects) from ref 104 with an estimated reference error of 0.2%. [c]Statistical data discarding the isoamyl-acetate for which a wrong conformer is obtained. [d]Bond lengths of long bonds as used in ref 1. [e]Statistical data discarding the transition-metal-containing systems HAPPOD[1] and KAMDOR. In both cases, the optimization resulted in cleavage of the metal−metal bond. [f]Statistical data discarding the transition-metal-containing system HAPPOD[1] where the optimization yielded a structure with dissociated (and reordered) ligands. [g]Bond lengths of covalent bonds comprising heavy main group elements ($n_{val} \geq 3$) as used in ref 1. [h]The reference bond lengths have an estimated uncertainty of 2 pm possibly resulting from crystal packing and/or vibrational effects.

Overall, GFN-xTB performs similarly or even better in all three sets compared to the significantly more expensive HF-3c approach. GFN-xTB furthermore has the advantage of being applicable to transition metals, which we demonstrate in the following subsection.

3.1.2. Transition-Metal Complexes. The computation of transition-metal and organometallic complex geometries is a challenge for both single reference QM and SQM methods. As a benchmark for 3d transition-metal systems, we use a set of 32 complexes with 50 bond distances compiled by Bühl and Kabrede, dubbed TMC32.[108] The complexes are $Sc(acac)_3$, $TiCl_4$, $TiMeCl_3$, $TiMe_2Cl_2$, $Ti(BH_4)_3$, VOF_3, VF_5, $VOCl_3$,

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009

$V(NMe_2)_4$, $V(Cp)(CO)_4$, CrO_2F_2, CrO_2Cl_2, $CrO_2(NO_3)_2$, $Cr(C_6H_6)_2$, $Cr(C_6H_6)(CO)_3$, $Cr(O)_4$, MnO_3F, $MnCp(CO)_3$, $Fe(CO)_5$, $Fe(CO)_3$, $Fe(CO)_2(NO)_2$, $FeCp_2$, $Fe(C_2H_4)(CO)_4$, $Fe(C_5Me_5)$, $CoH(CO)_4$, $Co(CO)_3(NO)$, $Ni(CO)_4$, $Ni(acac)_2$, $Ni(PF_3)_4$, $CuCH_3$, CuCN, and $Cu(acac)_2$. OM2-D3(BJ) and DFTB3-D3(BJ) are excluded in the comparison because of missing parameters, and HF-3c is not considered either due to SCF convergence problems.

Table 4 contains the statistical data on the TMC32 set obtained with GFN-xTB, PM6-D3H4X, and MSINDO-D3(BJ)H+. All methods have a negative MD indicating underestimated bond lengths. The MDs and MADs of PM6-D3H4X and GFN-xTB are much smaller in magnitude compared to MSINDO-D3(BJ)H+. GFN-xTB gives on average slightly shorter bonds with more systematic deviations than PM6-D3H4X. This good performance of GFN-xTB is not due to the finite temperature Fermi smearing because orbital occupations remain integral.

Table 4. Comparison of Calculated and Experimental Ground-State Equilibrium Bond Distances R_e (in pm) for 3d Transition-Metal Complexes (TMC32)[a]

	GFN-xTB	PM6-D3H4X	MSINDO-D3(BJ)H+
MD:	−2.7	−1.0	−12.7
MAD:	5.1	6.6	12.7
SD:	5.9	10.7	7.4
MAX:	16.5	44.8	28.3

[a]The set list taken from ref 108. MD = mean deviation, MAD = mean absolute deviation. SD = standard deviation, MAX = maximum absolute deviation.

The structures of 10 transition-metal complexes[63] computed with GFN-xTB and PBEh-3c[1] as reasonably accurate DFT reference are overlaid in Figure 1. The agreement between both methods is very good with RMSDs ranging between 0.05 and 0.41 Å. Two structures (TM4 and TM6) show larger RMSDs of about 1 Å. In TM4, this is due to rotated phenyl groups. TM6 is a flexible binuclear complex with separate, non-covalently interacting ligands. The hydrogen bonds between the oxalate and diaminocyclohexane make this a challenging case. PM6-D3H4X shows smaller RMSDs for TM4 and TM6 (0.51 and 0.54 Å), but yields larger RMSDs than GFN-XTB for all the other complexes (see Supporting Information). Notable cases are TM5 and TM10. Here the RMSDs with PM6-D3H4X are larger by a factor of 4 and 5, respectively, compared to GFN-xTB. In all complexes, the overall structure of the transition-metal complexes is reproduced by GFN-xTB without significant reorganization or dissociation.

Together with the TMC32 set (Table 4), this shows that GFN-xTB is able to capture the bonding situations in organometallic systems rather reasonably and significantly better than NDDO-type methods without being specifically modified for such systems. The Fermi smearing technique (see eq 6) also allows the treatment of systems with small HOMO−LUMO gaps, which are problematic for standard single reference treatments. More detailed tests on transition-metal complexes will follow in a separate paper. We also discuss a few cases where static electron correlation is important in Section 3.4.

3.2. Noncovalent Interactions. *3.2.1. Noncovalent Interaction Energies.* A number of test sets to assess the performance of quantum chemistry methods for London dispersion interactions,[65,109] which are missing in mean-field methods like HF and semilocal DFT, have been developed in the past decade. Here, we apply uniformly all SQM methods in combination with the atom-pairwise D3(BJ)[62,99] correction (D3 with zero damping[62] for PM6-D3H4X) for London dispersion interactions. The SQM methods, however, differ in the way they account for the electrostatic, Pauli exchange repulsion, and induction interactions, which contribute together to the total noncovalent interaction. Different approaches to improve the treatment of hydrogen bonds are also applied. PM6-D3H4X and MSINDO-D3(BJ)H+ use a force field-type hydrogen-bonding potential, and DFTB3 and GFN-xTB incorporate modifications in the electronic Hamiltonian, such as modified Coulomb interaction for hydrogen atoms in DFTB3 and additional 2s function for hydrogen in GFN-xTB.

The S22[79] set consists of 22 noncovalently bound, small size complexes and serves as a measure for the accuracy of quantum chemistry methods to describe noncovalent interactions that are dominated by London dispersion and hydrogen bonding. We use the most recent reference values for association energies.[80]

Well-performing, dispersion-corrected DFT methods typically give MADs of ≤0.5 kcal mol^{-1}.[65] HF-3c and PM6-D3H4X

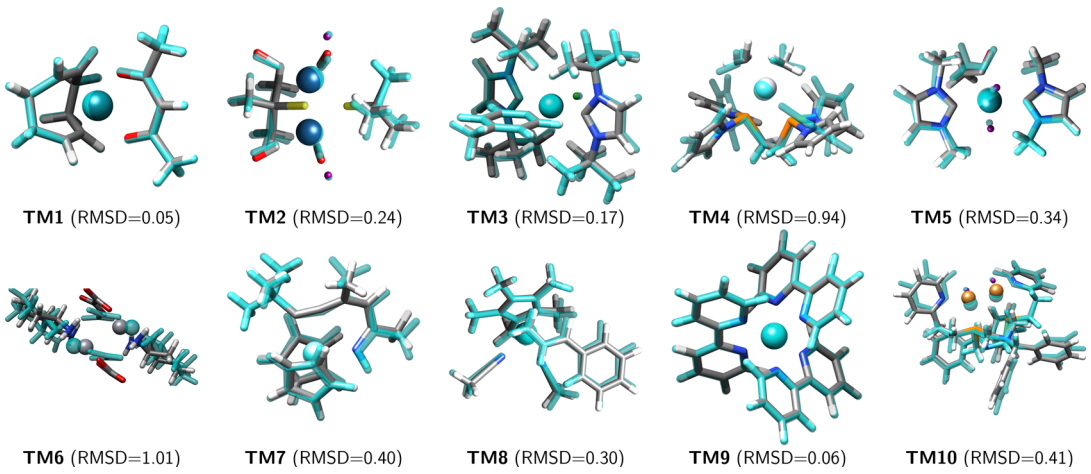


Figure 1. Overlay of 10 larger transition-metal complex[63] structures optimized with PBEh-3c (transparent cyan) and GFN-xTB (color coded atoms). The RMSD for an all-atom best fit is given in Å.

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009

provide similar accuracies with no significant systematic deviations (Table 5). Most of the other studied SQM methods

Table 5. Association Energies of Noncovalently Bound Complexes[a]

	GFN-xTB	PM6-D3H4X	MSINDO-D3(BJ)H+	OM2-D3(BJ)	DFTB3-D3(BJ)	HF-3c
			S22[b]			
MD:	1.3	0.3	0.4	0.8	0.7	−0.1
MAD:	1.3	0.6	1.6	0.9	1.0	0.5
SD:	1.1	0.7	2.0	1.2	1.5	0.7
MAX:	3.6	2.1	3.6	3.9	5.0	1.4
			S12L[c]			
MD:	−2.5	−6.0	−7.9	−4.4[d]	−3.6[e]	−3.8
MAD:	7.5	6.8	8.6	4.5[d]	4.7[e]	5.3
SD:	9.4	8.5	7.5	4.3[d]	5.4[e]	5.5
MAX:	23.2	21.8	19.1	12.8[d]	13.0[e]	10.2
			WATER27[f]			
MD:	−0.4	0.1	−4.0	0.2	0.1	−4.4
MAD:	1.4	0.9	4.0	0.9	0.8	4.4
SD:	1.9	1.3	3.3	1.2	1.0	4.3
MAX:	6.1	4.2	12.1	4.3	1.9	14.8
			C15[g]			
MD:	2.1	0.7	0.2	1.8	3.9	−3.7
MAD:	2.1	1.3	1.8	1.8	4.1	3.7
SD:	1.5	1.5	2.3	1.2	3.5	2.5
MAX:	5.2	3.6	4.0	4.6	12.5	8.3
			PL24[h]			
MD:	−1.7	−7.9	−4.6	–	−1.0	1.2
MAD:	3.5	8.2	5.5	–	3.0	1.7
SD:	4.5	7.8	6.1	–	3.8	1.9
MAX:	12.5	23.2	16.7	–	8.8	5.3

[a]The sets are S22 and S12L as well as sets with emphasis on hydrogen bonding, such as the WATER27, C15, and PL24. Deviations of association energies are given in kcal mol^{-1}. [b]Reference data taken from ref 80. [c]Reference structures are taken from ref 110, while the reference energies are taken from refs 65 and 111. [d]Excluding systems **3b** and **7a** due to missing parameters. [e]Excluding system **7a** due to missing parameters. [f]Reference energies taken from refs 112 and 113. Energies are given per water (hydronium/hydroxide) molecule. [g]Charged hydrogen-bonded dimers. Reference data taken from ref 31. [h]24 protein−ligand interactions with reference data taken from ref 114.

yield MADs close to 1.0 kcal mol^{-1}. This holds also for GFN-xTB where the deviation indicates a systematic underbinding tendency. GFN-xTB shows the largest relative deviations for π-stacked structures (see Supporting Information), such as benzene−cyanide and ethene dimer, which results from the monopole approximation and the poor description of benzene and ethene quadrupole moments. The NDDO-based methods PM6-D3H4X and OM2-D3(BJ), and also HF-3c, perform much better for these systems because of the more accurate electrostatic interactions. GFN-xTB describes rather well the hydrogen-bonded systems with largest deviation observed for the stacked uracil dimer (3.5 kcal mol^{-1}, 16%) and the best overall performer of GFN-xTB—the formic acid dimer (>0.2 kcal mol^{-1}, > 1%). Therefore, the description of hydrogen bonds facilitated by an additional s-basis function on hydrogen performs reasonably well. The S22 (or the related S66)[81] set was used as a fitting set for most of the considered methods, and their performance has to be assessed for larger system.[16]

To this end, we use the S12L[110] set, which consists of large, experimentally accessible supramolecular complexes. The magnitudes of the association energies, derived from experimental free association energies, are in the range of 20−35 kcal mol^{-1} for the neutral complexes, and the values for the charged complexes are significantly larger—about 80 kcal mol^{-1} for **6a** and **6b**, and 133 kcal mol^{-1} for **7a**. Complex **7a** (see Supporting Information) is excluded from the statistical analysis for DFTB3-D3(BJ) and OM2-D3(BJ) due to missing parameters for iron. The chlorine containing complex **3b** is also neglected for OM2-D3(BJ) for the same reason. Given that the S12L energies are close to 5 times larger than the S22 ones, the performance of DFTB3-D3(BJ) and OM2-D3(BJ) with MADs of 4.7 and 4.5 kcal mol^{-1} is consistent throughout the two sets. When the **7a** complex is included in the statistics, the average S12L energy becomes 6−7 times larger than the S22 energies. Therefore, the performance of the remaining SQM methods is also uniform throughout, except PM6-D3H4X and HF-3c, which show somewhat larger deviations for S12L. GFN-xTB demonstrates the smallest MD of all methods, which approaches zero if system **7a** is excluded from the analysis. The MAD values indicate consistent deviations for the S22 and S12L sets. The relatively larger standard deviation of GFN-xTB is due to complex **7a**, which is overstabilized by ≈23 kcal mol^{-1} (14%). If we exclude this system the SD goes down to 7.2 kcal mol^{-1}. The effect of the monopole approximation is less pronounced for the more charge delocalized systems of S12L, and neither HF-3c nor the NDDO methods perform systematically better than GFN-xTB or DFTB3-D3(BJ).

We consider next the association energies of neutral and also charged water clusters. The well-established WATER27 set[112] is used for this purpose, including the revised reference energies for the $(H_2O)_{20}$ subset.[113] Due to the varying cluster sizes, we normalize the association energies to the number of water (and hydronium/hydroxide) molecules (see Supporting Information). The electrostatic interactions and hydrogen bonding are very important for these highly polar complexes. MSINDO-D3(BJ)H+ and the computationally more demanding HF-3c approach show relatively large systematic overestimation of the binding energy, as the magnitudes of the MDs are equal to the MADs throughout. The largest part of these errors results from the subset of charged clusters incorporating either hydronium or hydroxide. In agreement with ref 16, DFTB3-D3(BJ) performs best with an MAD and SD of ≤1. PM6-D3H4X and OM2-D3(BJ) yield negligibly larger MAD and SD values. The respective values are larger by about 0.5 kcal mol^{-1} for GFN-xTB, which performs worse for the anionic clusters. Excluding these systems from the statistical analysis, GFN-xTB shows similar deviations as PM6-D3H4X and OM2-D3(BJ) on this subset (MAD and SD of 0.6−0.7 kcal mol^{-1}).

The C15 set contains charged hydrogen-bonded systems. PM6-D3H4X shows the best overall performance with an MAD of 1.3 kcal mol^{-1} and a relatively small MD and SD. OM2-D3(BJ) gives the smallest SD but with MD and MAD equal to 1.8 kcal mol^{-1} is only slightly better than GFN-xTB. GFN-xTB demonstrates a bit larger systematic deviation, MD and MAD of 2.1 kcal mol^{-1}, and an error spread that is the same as PM6-D3H4X, and thus the method describes the hydrogen bonds of charged systems quite well. Overall, NDDO-type methods perform relatively better than TB methods, especially if hydrogen-bonding corrections are used as in PM6-D3H4X and MSINDO-D3(BJ)H+.

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009

The last organic benchmark set listed in Table 5 is the protein−ligand binding motif set PL24.[114] HF-3c gives the smallest MAD of 1.7 kcal mol^{-1} and the smallest SD. DFTB3-D3(BJ) ranks second followed by GFN-xTB with an MAD of 3.5 kcal mol^{-1}. The TB methods demonstrate better accuracy than the NDDO methods PM6-D3H4X and MSINDO-D3(BJ). PM6-D3H4X, which performs very well on the hydrogen bonded sets, is worse than MSINDO-D3(BJ)H+ here. The good performance of GFN-xTB on PL24 is very encouraging and justifies further applications of GFN-xTB to biological systems.

We continue the analysis of the SQM methods accuracy for NCI energies with a few more challenging and unconventional sets listed in Table 6. OM2-D3(BJ) is not included in the discussion due to missing parameters. The first set is a collection of three subsets compiled by Truhlar and co-workers,[115] which contain systems with predominantly dipolar (DI6/04), charge-transfer interactions (CT7/04), and hydrogen-bonded complexes (HB6/04).

Table 6. Association Energies[a] of Noncovalently Bound Complexes, Including the HB6/04, CT7/04, and DI6/04 Sets of Truhlar and Co-workers[b], the X40[c] Set, a Set of Metal−Organic Framework Fragment Interactions with CO_2 (MOF-CO_2)[e], and a Set of Rare Gas−Cucurbituril Complexes (RG-CB_x)[f]

	GFN-xTB	PM6-D3H4X	MSINDO-D3(BJ)H+	DFTB3-D3(BJ)	HF-3c
			Nonbonded Interactions[b]		
MD:	0.3	6.7	2.5	0.0	2.4
MAD:	0.9	7.1	2.7	1.8	2.8
SD:	1.1	11.2	2.1	2.4	3.1
MAX:	2.1	31.4	6.0	5.9	9.1
			X40[c]		
MD:	0.7	0.7	−1.6[d]	−1.1	−1.2
MAD:	0.9	1.1	2.5[d]	1.8	1.9
SD:	0.9	2.2	5.0[d]	2.4	2.3
MAX:	3.7	11.4	21.5[d]	6.9	6.9
			MOF-CO_2[e]		
MD:	0.5	−0.5	−2.8	–	−1.4
MAD:	0.7	0.8	2.8	–	1.4
SD:	0.7	1.0	2.2	–	1.4
MAX:	1.7	2.7	11.8	–	6.3
			RG-CB_x[f]		
MD:	−0.1	13.5 (4.1)[g]	–	–	−8.1 (−0.1)[h]
MAD:	0.3	14.3 (5.0)[g]	–	–	9.8 (2.2)[h]
SD:	0.3	21.2 (6.0)[g]	–	–	18.0 (2.8)[h]
MAX:	0.6	61.8 (15.2)[g]	–	–	52.7 (4.5)[h]

[a]Deviations of association energies are given in kcal mol^{-1}. [b]This is a collection of subsets: hydrogen-bonded (HB6/04), charge-transfer (CT7/04), and dipole interacting (DI6/04) systems. Reference data taken from ref 115. [c]Halogen-bonded systems with reference data taken from ref 116. [d]Excluding iodine containing systems due to missing parameters. [e]Interactions between CO_2 and metal−organic framework (MOF) fragments. Reference data taken from ref 117. [f]Association energies of rare gas atoms with cucurbituril hosts (CB5 and CB6). Reference structures at the PBEh-3c level and reference energies from counterpoise-corrected DLPNO−CCSD(T)/def2-QZVPP[77,118,119] calculations (data will be published elsewhere). [g]Data without krypton containing systems. [h]Data without xenon containing systems.

This benchmark set contains the elements F, Cl, and S, apart from the standard elements H, C, N, and O. GFN-xTB is the front runner here with the smallest MAD and SD of 0.9 and 1.1 kcal mol^{-1}, very close to "chemical accuracy". This is remarkable because the magnitudes of the binding energies range from 1 to 16 kcal mol^{-1}, and no other SQM method tested yields comparable accuracy. For instance, DFTB3-D3(BJ), which ranks second, shows deviations that are larger by 1 kcal mol^{-1}. Both TB methods have no systematic deviations, unlike the NDDO methods and HF-3c.

The X40[116] set contains 40 halogen-bonded systems and presents a challenge to the approximate electrostatics and element pair-specific parametrization in SQM methods. GFN-xTB and PM6-D3H4X employ specific correction terms for halogen bonds (see eq 14), which are parametrized for related systems. This is reflected by the smallest MD magnitudes of 0.7 kcal mol^{-1} of both methods in Table 6. MSINDO-D3(BJ)H+, DFTB3-D3(BJ), and HF-3c have MD magnitudes > 1 kcal mol^{-1}. While the MADs of GFN-xTB and PM6-D3H4X are rather similar, the latter gives outliers and a relatively large SD of 2.2 kcal mol^{-1}), which is close to the one of DFTB3-D3(BJ) and HF-3c. GFN-xTB performs best with a MD, MAD, and SD below 1 kcal mol^{-1}. Thus, shortcomings of the monopole approximation are compensated by the halogen-bond correction in GFN-xTB.

MOF-CO_2 is a test set for noncovalent interactions of CO_2 with functional groups of metal−organic frameworks.[117] DFTB3 is not tested due to missing parameters for Li and Zn in the 3OB parametrization.[92−94] GFN-xTB performs the best here showing the smallest magnitudes of MD (together with PM6-D3H4X), MAD, and SD. GFN-xTB shows a slight underbinding tendency supported by a similar MAD and a small spread of errors. PM6-D3H4X ranks second with a larger SD, whereas HF-3c and especially MSINDO-D3(BJ)H+ perform worse.

The last set in Table 6 contains unpublished results for the noncovalently bound complexes of rare gas atoms and cucurbituril[x] hosts (x = 5,6). The reference energies are obtained by domain-based local pair natural orbital approximated coupled cluster calculations (DLPNO-CCSD(T)/def2-QZVPP),[77,118,119] including counterpoise correction (more details will be published elsewhere). MSINDO and DFTB3 calculations are not discussed due to missing parameters. The performance of GFN-xTB is exceptionally good with a MD close to zero and MAD and SD of only 0.3 kcal mol^{-1}. HF-3c and PM6-D3H4X provide large MADs around 10 kcal mol^{-1}. The largest outliers result from problems for specific elements, krypton for PM6-D3H4X and xenon for HF-3c. The performance of both methods is improved significantly if we exclude the corresponding complexes from the statistical analysis. The densely packed cucurbituril[5] (CB5) complexes show large errors that indicate systematic errors at short distances for these element-method combinations. While the reasons for the drastic destabilization of the krypton complexes with PM6-D3H4X (see Supporting Information) are unclear, the problems of HF-3c with xenon are attributed to the geometrical counterpoise correction (gCP). Both methods are outperformed by GFN-xTB, even when the problematic cases are excluded.

Finally, we summarize the performance on all test sets with emphasis on the range of applicability and consistent performance of the methods. In Figure 2, we plot the mean unsigned relative deviation (MURD) of all methods, which

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009

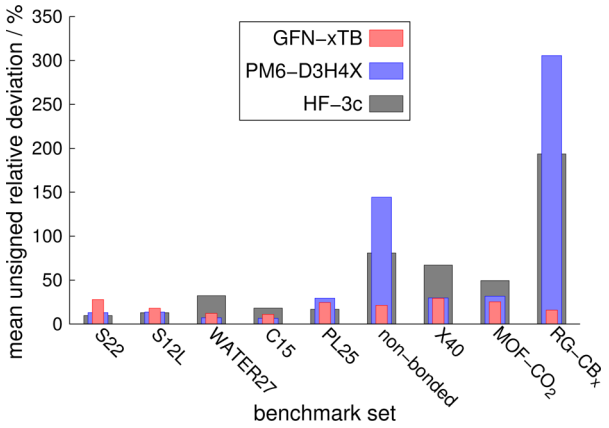


Figure 2. Mean unsigned relative deviations of the semiempirical methods GFN-xTB (red), PM6-D3H4X (blue), and HF-3c (gray) on the noncovalent interaction benchmark sets considered in this work. The deviations are in percent (%).

parametrization covers all test sets. Consequently, OM2-D3(BJ) and DFTB3-D3(BJ) are not included because even though they perform well for some sets, they do not have parameters for all elements in the sets. Figure 2 indicates that the 3 remaining methods give similar MURDs, <30%, for the first 5 sets, these are the sets in Table 5. GFN-xTB provides MURDs smaller than 30% for the last four sets as well, which correspond to the sets in Table 6. Unlike the consistent accuracy of GFN-xTB for all test sets, PM6-D3H4X and HF-3c demonstrate larger errors for some of the sets and therefore are not consistently accurate for NCI energies.

3.2.2. Structures of Noncovalently Bound Systems. This section is devoted to the performance of SQM methods for geometries of noncovalently bound systems. This is of practical significance because SQM methods are likely to be used for geometry optimizations followed by single-point calculations at a higher level of theory in a multilevel scheme. Because even HF-3c can become too computationally expensive for such applications, we focus on the performance of the cheaper SQM methods.

We optimize the dimers of S22[79] and give the deviations of the center-of-mass (CMA) distances in Table 7. HF-3c gives the smallest deviations, followed by DFTB3-D3(BJ), which underestimates the CMA distances on average, and GFN-xTB and PM6-D3H4X show comparable MADs and SDs. In general, methods that perform well for S22 energies provide reasonable structures, except the outlier OM2-D3(BJ). The vanishing MD and the small SD of the CMA distances observed for GFN-xTB suggest that this method produces PESs resembling the high-level reference curves.

Hobza and co-workers have presented a benchmark set, which contains structures of small isolated peptides with aromatic side chains (P26).[120] The relative stabilities of these structures depend significantly on the intramolecular noncovalent interactions. The average heavy atom RMSD is the best for DFTB3-D3(BJ) and GFN-xTB, followed by HF-3c. PM6-D3H4X and OM2-D3(BJ) show larger $\overline{\text{RMSD}}$s and perform better than MSINDO-D3(BJ)H+. We plot the heavy atom RMSDs of the individual systems for GFN-xTB, PM6-D3H4X, and DFTB3-D3(BJ) in Figure 3a (see Supporting Information for system abbreviations and results with the other SQM methods). GFN-xTB produces a few larger RMSDs (>0.2 Å), which arise from deviations in the dihedral angle of the

Table 7. Comparison of Structures of Noncovalently Bound Systems[a]

	GFN-xTB	PM6-D3H4X	MSINDO-D3(BJ)H+	OM2-D3(BJ)	DFTB3-D3(BJ)	HF-3c
S22[b] (CMA distance in pm)						
MD:	−5	3	−25	−17	−11	−2
MAD:	15	14	35	26	14	7
SD:	18	21	47	29	16	11
MAX:	50	58	186	66	45	34
P26[c] (Heavy Atom Position RMSD of Small Peptides in pm)						
$\overline{\text{RMSD}}$:	12	30	48	28	11	20
S66 × 8[d] (CMA Distance in %)						
MRD:	−0.6	0.6	−2.5	−5.0	−2.9	−1.2
MURD:	2.0	2.0	5.3	5.0	3.1	1.2
SRD:	2.7	2.6	6.0	2.7	2.8	1.1
MAXR:	7.0	6.7	14.5	12.4	10.9	4.2

[a]The CMA distances of the S22 complexes and the RMSD of the heavy atom positions of the small peptides from P26 are compared for fully optimized geometries. The CMA distances of the S66 complexes are estimated from cubic spline interpolations of energies computed on the S66 × 8 set. [b]Reference structures taken from ref 79. [c]Reference structures are taken from ref 120. [d]CMA distances determined from cubic spline interpolation (see ref 1) based on reference structures and energies from ref 81.

aromatic side chains, the case for system 5 (FGG_470). For most systems, GFN-xTB shows RMSDs that are smaller than the ones of DFTB3-D3(BJ). PM6-D3H4X gives larger RMSD values, including larger outliers compared to the TB models.

The S66 × 8[81] by Hobza and co-workers has been used to compare noncovalently bound complex equilibrium geometries at high theoretical level. The reference minimum CMA distances are determined by interpolation from the CCSD(T)/CBS energies. The same procedure is applied to the computations with the SQM methods, and the statistical results for the comparison of the equilibrium CMA distances are compiled in Table 5 (see Supporting Information for details). HF-3c performs the best with the smallest MURD and SRD and a slight tendency to underestimate the equilibrium distances by 1.2%. GFN-xTB and PM6-D3H4X rank second, both providing distances with MURDs of 2.0%; PM6-D3H4X slightly overestimating and GFN-xTB slightly underestimating the distances. OM2-D3(BJ), which performs quite well for energies (see Table 5), underestimates significantly the distances (−5%) similarly to the performance on S22. Together with the results on ROT34[104] (see Table 3), OM2-D3(BJ) does not confer enough accuracy for geometry optimizations in multilevel approaches.

Provided the encouraging performance of GFN-xTB for fast geometry optimizations, we test the method in a particular multilevel scheme. To this end, we reoptimize the PBEh-3c structures of the S30L[111] supramolecular test set with PM6-D3H4X, DFTB3-D3(BJ), and GFN-xTB, and after that recompute single-point energies with PBEh-3c. Because the geometries of the molecular fragments often do not differ much between the free and the associated state, deviations of the association energies, $\Delta\Delta E$, reflect changes of the intermolecular PES of SQM methods with respect to PBEh-3c. These deviations are then positive, indicating underbinding and effective cancellation of intramolecular contributions. The results are plotted in Figure 3b and absolute ΔE values can be found in the SI. The $\Delta\Delta E$ values are mostly positive, except two negative $\Delta\Delta E$ (systems 11 and 12) for GFN-xTB due to

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009

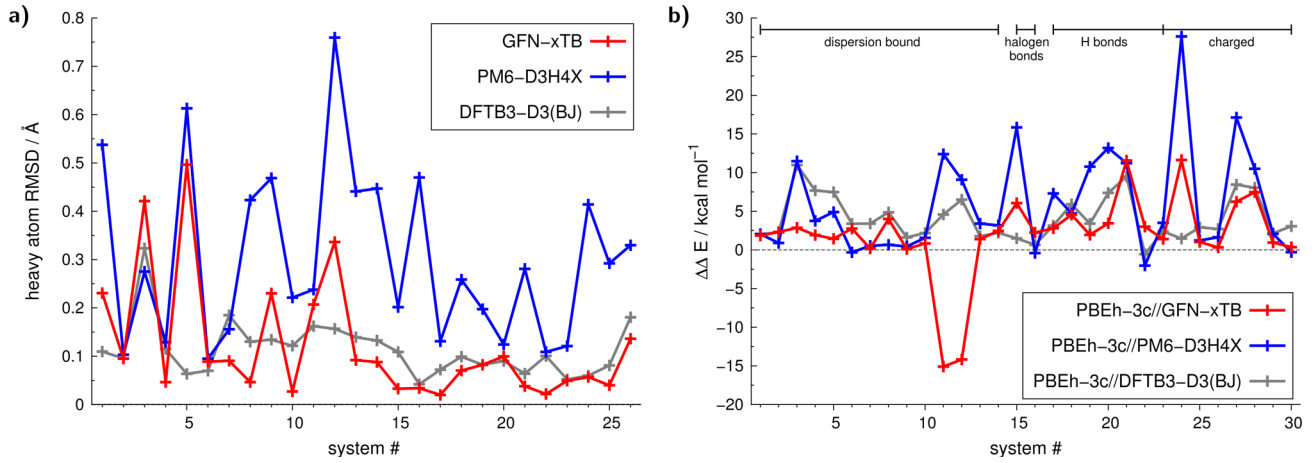


Figure 3. (a) RMSD of heavy atom positions relative to the MP2/aug-cc-pVTZ[121−123] reference structures for the P26[120] peptide benchmark set. The RMSDs obtained for geometries optimized with GFN-xTB, PM6-D3H4X, and DFTB3-D3(BJ) are given in red, blue, and gray, respectively. (b) Deviations in the association energies for 30 large noncovalent complexes from S30L[111] computed with PBEh-3c on geometries obtained with the semiempirical methods GFN-xTB, PM6-D3H4X, and DFTB3-D3(BJ). The deviations are given in kcal mol^{-1} as the difference $\Delta\Delta E = \Delta E_{\text{PBEh-3c//SQM}} - \Delta E_{\text{PBEh-3c//PBEh-3c}}$.

large relaxation of the pentakis(1,4-benzodithiino)-corannulene host upon binding. In fact, the free host with GFN-xTB deviates more from the PBEh-3c structure than the bound state leading to the erroneous stabilization. Apart from this outlier, GFN-xTB gives intermolecular PES that resemble quite well the PBEh-3c surface (excluding systems 11 and 12 results in MRD = MURD = SRD = 9.4%). The deviations with PM6-D3H4X are roughly twice as large, and the performance of DFTB3-D3(BJ) is quite good with MRD = 13.2%, MURD = 13.3%, and SRD = 11.1%. We note that the results may be different for other higher level methods, and we choose PBEh-3c for the comparison because similarly to GFN-xTB, the approach is designed to provide geometries and NCIs that are comparable to the results of dispersion-corrected hybrid-functional calculations in large basis sets. It is also a sound choice because of the computational cost of the method and the large system sizes considered. Possible problematic systems can be straightforwardly diagnosed by comparison of the association energies computed at the SQM geometries with the higher level method (PBEh-3c) and the SQM method (GFN-xTB). The typical GFN-xTB error is <30% for NCI energies (Figure 2), and the multilevel PBEh-3c//GFN-xTB approach gives errors below 10% for association energies with respect to a full PBEh-3c//PBEh-3c computation. For systems 11 and 12, the ΔE from GFN-xTB//GFN-xTB are −38.8 and −40.5 kcal mol^{-1} but are −54.1 and −53.9 kcal mol^{-1}, respectively, with PBEh-3c//GFN-xTB and can be identified as problematic for the multilevel treatment.

3.3. Conformational Energies. SQM methods find applications in the sampling and energetic ranking of conformers, which are relevant to biophysical chemistry, organic supramolecular chemistry, and organic crystal polymorph prediction.[124] We use a number of well-established benchmark sets for conformational energetics to gauge the performance of GFN-xTB. ACONF[125] is a set of conformers of butane, propane, and hexane, CYCONF[126] consists of gas-phase conformers of cysteine, PCONF[127] contains conformers of a small Phe-Gly-Gly tripeptide, and MCONF[128] has 51 conformers of melatonine. We consider, in addition, three hairpin conformers of long alkane chains, called the Hairpin set,[129] and 46 diuracilphosphate conformers that comprise the UpU46 set[130] as an example of a charged system. The statistical data for all methods are given in Table 8, and details are deferred to the Supporting Information.

The ACONF set tests mainly the balance between repulsive Pauli exchange and attractive London dispersion interactions and is characterized by relatively small conformational energies ranging between 0.5 and 5.0 kcal mol^{-1}. The NDDO methods OM2-D3(BJ) and PM6-D3H4X perform very well for this set showing small MADs and nearly no systematic deviation. The worst performer is the MSINDO-D3(BJ)H+ method, which strongly overestimates the conformational energies. The accuracy of GFN-xTB and DFTB3-D3(BJ) is much better than of MSINDO-D3(BJ)H+, somewhat worse than that of the other NDDO approaches, and on par with the much more computationally expensive HF-3c. The TB models underestimate the conformational energies on average with mean deviations of −0.7 and −0.8 kcal mol^{-1}, which points to softer short-range repulsion forces. GFN-xTB and DFTB3-D3(BJ), however, capture correctly the energetic ordering of the conformers as evidenced by Figure 4a, unlike HF-3c, which exhibits changes in the conformer energetic ordering (see Supporting Information).

CYCONF presents a more difficult test set than ACONF for SQM methods because cysteine conformational energies are sensitive to the description of electrostatic interactions and hydrogen bonds. OM2 lacks parameters for sulfur and is not included in the discussion. The best performing method is HF-3c showing hardly any systematic deviation and a spread <1.0 kcal mol^{-1}. GFN-xTB and DFTB3-D3(BJ) perform the best from the group of the more empirical models. Both methods underestimate the conformational energies with GFN-xTB, demonstrating slightly better accuracy than DFTB3-D3(BJ). The NDDO methods PM6-D3H4X and MSINDO-D3(BJ)H+ cannot compete with the accuracy of the TB models on this set.

MCONF tests similar interactions as CYCONF that have a stronger dispersion component due to the indole residue of melatonine and lacks third row elements. HF-3c retains the very good accuracy shown with CYCONF. GFN-xTB, PM6-D3H4X, and DFTB3-D3(BJ) compete closely for the second best position, whereas OM2-D3(BJ) ranks this time with the worst performer, which is MSINDO-D3(BJ)H+. The three

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009

Table 8. Comparison of Conformational Energies Computed with Different Semi-Empirical Methods Relative to High-Level Reference Calculations[a]

	GFN-xTB	PM6-D3H4X	MSINDO-D3(BJ)H+	OM2-D3(BJ)	DFTB3-D3(BJ)	HF-3c
			ACONF[b]			
MD:	−0.7	0.0	1.2	0.1	−0.8	−0.9
MAD:	0.7	0.5	1.2	0.2	0.8	0.9
SD:	0.4	0.6	0.6	0.3	0.5	0.4
MAX:	1.4	1.3	2.5	0.8	1.8	1.8
			CYCONF[c]			
MD:	−1.2	−2.8	−4.4	–	−1.7	−0.3
MAD:	1.4	3.3	4.7	–	1.7	0.7
SD:	1.3	2.7	3.5	–	1.1	0.7
MAX:	2.7	5.5	8.0	–	2.9	1.1
			MCONF[d]			
MD:	−1.6	−1.2	−2.6	−2.5	−1.1	−0.3
MAD:	1.6	1.5	2.8	2.6	1.8	0.9
SD:	0.9	1.2	2.3	1.8	1.9	1.1
MAX:	2.7	3.2	7.0	5.4	4.1	2.2
			PCONF[e]			
MD:	2.5	2.5	5.6	0.9	0.6	2.6
MAD:	2.5	2.6	5.7	1.1	1.2	2.6
SD:	1.9	1.8	3.1	0.9	1.4	1.1
MAX:	5.9	6.0	10.0	2.0	2.8	3.6
			UpU46[f]			
MD:	−0.1	2.2	1.9	–	−0.8	2.7
MAD:	1.4	2.6	2.3	–	1.4	3.0
SD:	1.8	2.3	2.0	–	1.6	2.1
MAX:	5.6	7.0	6.0	–	6.4	6.1
			Hairpin[g]			
MD:	−0.2	0.6	7.4	0.0	−2.4	−2.9
MAD:	0.2	0.6	7.4	0.2	2.4	2.9
SD:	0.2	0.2	0.6	0.2	0.3	0.3
MAX:	0.4	0.7	8.1	0.2	2.6	3.1

[a]The sets are ACONF, CYCONF, MCONF, and PCONF from the GMTKN30 benchmark database, and the UpU46 set and a recently introduced set of three large, linear alkanes, called hairpin. Deviations in the conformational energies are relative to the lowest-energy conformer calculated at the reference level and are given in kcal mol^{-1}. MD = mean deviation, MAD = mean absolute deviation, SD = standard deviation, and MAX = maximum absolute deviation. [b]Reference data taken from ref 125. [c]Reference data taken from ref 126. [d]Reference data taken from ref 128. [e]Reference data taken from ref 127. [f]Reference data taken from ref 130. [g]Energetic difference of the folded and linear, all-*anti* conformer.[129]

second best performers underestimate the melatonine conformational energies on average, and GFN-xTB shows somewhat larger (MAD = −1.6 kcal mol^{-1}) but more systematic deviations (SD = 0.9 kcal mol^{-1}) in the group.

The best performing methods on the tripeptide PCONF[127] set are OM2-D3(BJ) and DFTB3-D3(BJ) with MADs close to 1 kcal mol^{-1}. HF-3c, PM6-D3H4X, and GFN-xTB rank in second place, each of them giving an MAD of about 2.5 kcal mol^{-1}. HF-3c, however, shows a smaller standard deviation by 1 kcal mol^{-1} than PM6-D3H4X and GFN-xTB. Because the conformational energies of this set range from 0 to 2.5 kcal mol^{-1}, the overall accuracy of all SQM methods is insufficient, which makes this set a challenging test for approximate quantum chemical methods.

The UpU46[130] set contains charged molecules and reference energies that vary from 0.5 to 16.5 kcal mol^{-1}. DFTB3-D3(BJ) and GFN-xTB perform very well for this set with MADs of only 1.4 kcal mol^{-1} and MDs of −0.8 kcal mol^{-1} and −0.1 kcal mol^{-1}. PM6-D3H4X and HF-3c are among the worst performers here together with MSINDO-D3(BJ)H+. They systematically overestimate the conformational energies and show a large spread of 2.6 and 3.0 kcal mol^{-1}. A more detailed analysis of the individual conformational energies in Figure 4b shows that conformer 36, which contains a hydrogen bond between a phosphate oxygen and a ribose hydroxyl group, presents a difficult case. Despite performing similarly to GFN-xTB on average, DFTB3-D3(BJ) predicts different energetic ordering of the conformers, exhibiting an alternative energetic minimum. This solidifies the leading position of GFN-xTB, which reproduces the trend of conformational energies with a very small MD and reasonable MAD.

The Hairpin[129] set is an extension of ACONF to much larger alkanes. The NDDO methods OM2-D3(BJ) and PM6-D3H4X are among the front runners again. The performance of the TB models, however, shows differences. The accuracy of GFN-xTB becomes better than PM6-D3H4X and approaches the one of OM2-D3(BJ), and the accuracy of DFTB3-D3(BJ) deteriorates. HF-3c similarly to DFTB3-D3(BJ) underestimates the relative conformational energies of the folded, hairpin structures by about 2−3 kcal mol^{-1}. MSINDO-D3(BJ)H+ shows large systematic errors (>7 kcal mol^{-1}). The small standard deviations of all methods, consistent with the results on the ACONF set, show that the observed deviations have systematic character and are due to imbalance of short-range repulsion and long-range dispersion interactions.

In summary, GFN-xTB performs rather well reproducing the trends of the conformational energies and providing reasonable absolute estimates. A possible application of GFN-xTB is therefore the conformational sampling of large flexible systems as part of a multilevel scheme for structure prediction.

3.4. Potential Energy Curves, Thermochemistry, and Other Properties. In the preceding sections, we demonstrated the excellent performance of GFN-xTB for a broad range of systems for the target properties, molecular structures, and noncovalent interaction energies. In this section, we discuss the performance of GFN-xTB for PESs and their characteristics, such as covalent bond energies, harmonic vibrational frequencies, and reaction barriers.

We compare the accuracy of GFN-xTB for harmonic frequencies using the systems in the ROT34,[103,104] LB12, and HMGB11[1] sets, for which we computed the zero point vibrational energies E_{ZPVE} and free energy (translational, rotational, and vibrational) contributions G_{TRV} from the nuclear degrees of freedom.

The agreement of GFN-xTB with the PBEh-3c reference is excellent for both E_{ZPVE} and G_{TRV} (see Figure 5). GFN-xTB, similar to GGAs, does not require any scaling of the harmonic frequencies to obtain good agreement with PBEh-3c, which requires scaling by 0.95. These results suggest that GFN-xTB can be used in multilevel schemes for the computation of harmonic frequencies.

We tested GFN-xTB on the ISO34[131] and ISOL22[132] sets, which contain isomerization energies of organic molecules, yielding MADs of 6.5 and 9.8 kcal mol^{-1}. The performance for ISOL22 is comparable to the dispersion-corrected GGAs BLYP-D3(BJ)[133,134] and B97-D3(BJ)[135] and inferior to dispersion-corrected hybrid functionals whose MADs are 2−3 times smaller.[1,136] The performance of GFN-xTB for ISO34 is worse compared to dispersion-corrected GGAs with an MAD

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009

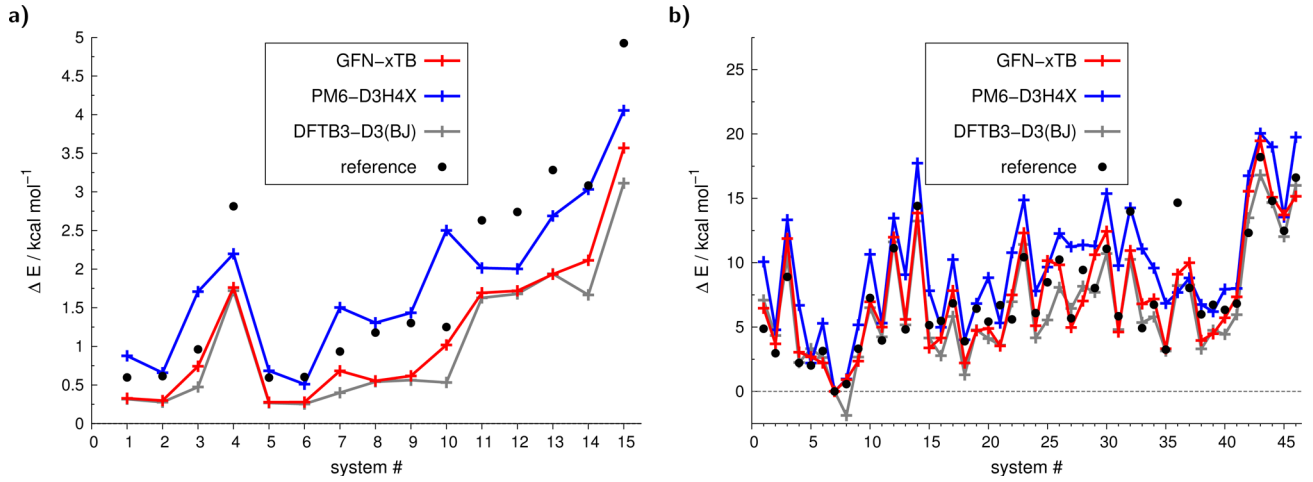


Figure 4. (a) Conformational energies of the alkane conformer (ACONF) set computed with GFN-xTB (red), PM6-D3H4X (blue), and DFTB3-D3(BJ) (gray). The reference energies from ref 125 are given as black dots. (b) Conformational energies of the diuracilphosphate (UpU46) set computed with GFN-xTB (red), PM6-D3H4X (blue), and DFTB3-D3(BJ) (gray). The reference energies from ref 130 are given as black dots.

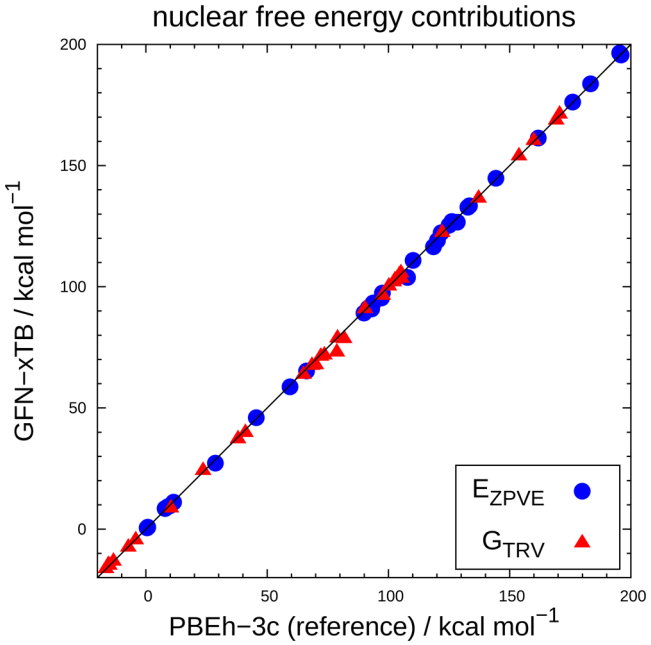


Figure 5. Comparison of zero point vibrational energies E_{ZPVE} and free energy contributions G_{TRV} from the translational, rotational, and vibrational degrees of freedom computed with GFN-xTB and PBEh-3c for the ROT34,[103,104] LB12, and HMGB11[1] systems (G_{TRV} at T = 298.15 K according to ref 110). All values are given in kcal mol^{-1}, and harmonic frequencies from PBEh-3c are scaled by a factor of 0.95.[1] For a more condensed representation, the values for the three LB12 systems 2-(1-diamantyl)[121]tetramantane (DIAD), the FLP, and in,in-bis(hydrosilane) (BHS), which are significantly larger than 200 kcal mol^{-1}, are not shown here and given in the Supporting Information.

around 2 kcal mol^{-1}.[136] In passing, we mention that PM6-D3H4X and DFTB3-D3(BJ) yield smaller MADs of 7.0 and 8.0 kcal mol^{-1}, respectively, for ISO34 and 5.1 and 3.4 kcal mol^{-1} for ISOL22 (see Supporting Information for detailed results also of other SQM methods). Thus, these methods, which were parametrized also to energies, perform better for bond energies than GFN-xTB but are still less accurate than dispersion-corrected density functionals. GFN-xTB provides good molecular geometries but in general does not yield very accurate (covalent) thermochemistry and is not recommended for routine reaction energy calculations. However, in our opinion, this holds for SQM methods in general, because accurate molecular energies are difficult to obtain even at sophisticated DFT or WFT levels, and the effect of semiempirical approximations is most severe for covalent bond energies. Hence, SQM methods should rather be used for rough estimates of covalent thermochemistry or as initial part of multilevel, large-scale screening applications. Note that this is different for noncovalent interaction energies, which are of reasonable to good quality as shown above.

For the purpose of standard molecular thermochemistry estimates, we employ a semiempirical protocol for the computation of standard heats of formation at 298 K (HOF). It introduces element-specific parameters to correct the computed atomization energy E_{at} for offsets in the atomic GFN-xTB energy levels of atom A, i.e.,

$$E'_{at} = E_{at}(\mathrm{GFN-xTB}) + \sum_{A}^{atoms} \delta E_{el}^{A} + CN_A f_{el}^{A} \quad (16)$$

Here δE_{el}^{A} is an element-specific correction for atom A and f_{el}^{A} is an additional element-specific parameter, which scales a contribution from hybridization, as taken into account via the coordination number CN_A. The value of E'_{at} is then used as usual to derive the HOF at 298 K by adding the molecular thermostatistical enthalpy including zero point vibrational contributions. The empirical parameters δE_{el}^{A} and f_{el}^{A} are determined by fitting E'_{at} to the G3/99 set[137] of reference HOF. We note in passing that also for DFTB, the use of additional element-specific fit parameters for the calculation of heats of formation has been presented,[138] while the conventional NDDO-based SQM methods do not apply such corrections.[139] Including only the CN_A-independent correction leads to an MAD for the G3 set of 12.2 kcal mol^{-1}, while the addition of the CN_A dependent term reduces the deviation to 9.8 kcal mol^{-1}. This approach is cross-validated on an independent HOF benchmark set taken from Table 3 of ref 140 (only neutral systems included) and compared to PM6 data (see Figure 6).

Both methods employ fully optimized structures. The GFN-xTB approach to HOF is further simplified by computing not

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009

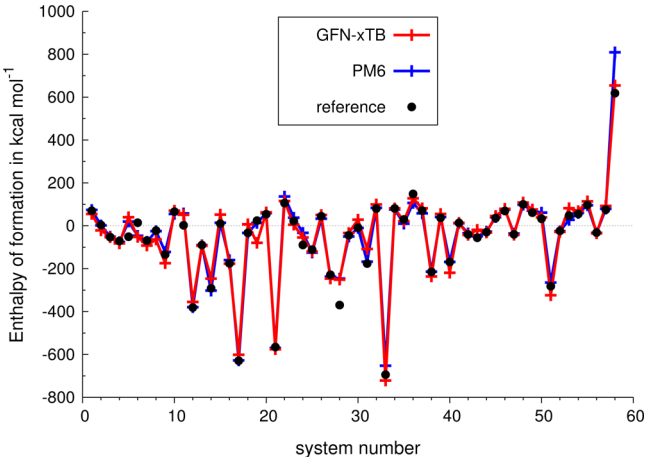


Figure 6. Enthalpies of formation for 'difficult' molecules (including elements up to chlorine) computed with GFN-xTB and PM6 along with experimental values (see ref 140). The detailed values are given in the Supporting Information.

the true GFN-xTB Hessian for the thermostatistical terms but employing Lindh's model Hessian,[69] which is also employed in the optimizer of the xtb code (see above), for this purpose. This approximation introduces additional errors of only about 1 kcal/mol on average for the G3/99 set of molecules, and it may be beneficial for large-scale screening applications (our code also allows usage of the GFN-xTB Hessian if this is preferred). As can be seen from Figure 6, GFN-xTB yields HOFs for this rather difficult set of molecules, which are of comparable quality (MAD = 24.9 kcal mol^{-1}) as the ones from PM6 (MAD = 17.2 kcal mol^{-1}). Though not designed for that purpose, the corrected GFN-xTB scheme can thus be used to obtain reasonable estimates for molecular thermochemistry. For example, the MADs obtained with GFN-xTB on the ISO34 and ISOL22 sets decrease by slightly more than 2 kcal mol^{-1}, if the atomic correction in eq 16 is included.

We study the behavior of GFN-xTB for covalent bonding to gain further insight. In Figure 7, we give the potential energy curves computed with GFN-xTB for the diatomic molecules H_2, N_2, F_2, and LiH where we also depict the positions of the minima computed at higher electronic structure level. The data demonstrate that GFN-xTB describes very well bond lengths in agreement with the results in Section 3.1, but the method overestimates significantly the covalent bond energies. In general it seems very difficult in a SQM context to fit simultaneously covalent bond lengths and bond energies by varying the fit weights of the properties because focusing the fit on bonding energies resulted in overestimated bond lengths and vice versa. Related observations have been recently reported for DFT energies and densities in ref 141. This correlation, however, does not affect the good noncovalent energies because they depend weakly on the H_0 part of the TB Hamiltonian.

As is shown in Figure 7a, GFN-xTB is able to dissociate molecules correctly. This is made possible by the finite temperature electron smearing approach, which mimics the incorporation of static correlation in an independent electron framework. The value of the electronic temperature controls the amount of electron smearing between the occupied and virtual orbital spaces, and we use it as a parameter. The curves in Figure 7a are plotted with T_{el} = 300 K, and a N_2 curve with a much larger electronic temperature of T_{el} = 4000 K is also shown in Figure 7b. It follows from the comparison in Figure 7b that Fermi smearing leaves the equilibrium part of the dissociation curve unaltered. In fact, Fermi smearing dissociates N_2 to N atoms with the correct value of spin, S = 0, and the difference in the dissociation asymptote is completely accounted for by the change in electronic entropy with temperature, $-T_{el}\Delta S_{el}$, with respect to the free atoms, each with spin S = 3/2. We include $-T_{el}\Delta S_{el}$ in the electronic energy because it enforces the correct spin and spatial symmetry and fractional occupations along the curve. Note that this only holds for a restricted treatment with identical spatial orbitals for the two spin parts. The downside is, however, that GFN-xTB in the present, restricted version cannot properly distinguish different spin states (see ref 43 for a unrestricted TB treatment). For T_{el} > 0 K and degenerate HOMO/LUMO levels, a low spin configuration is always more stable (larger S_{el}) than a high spin one.

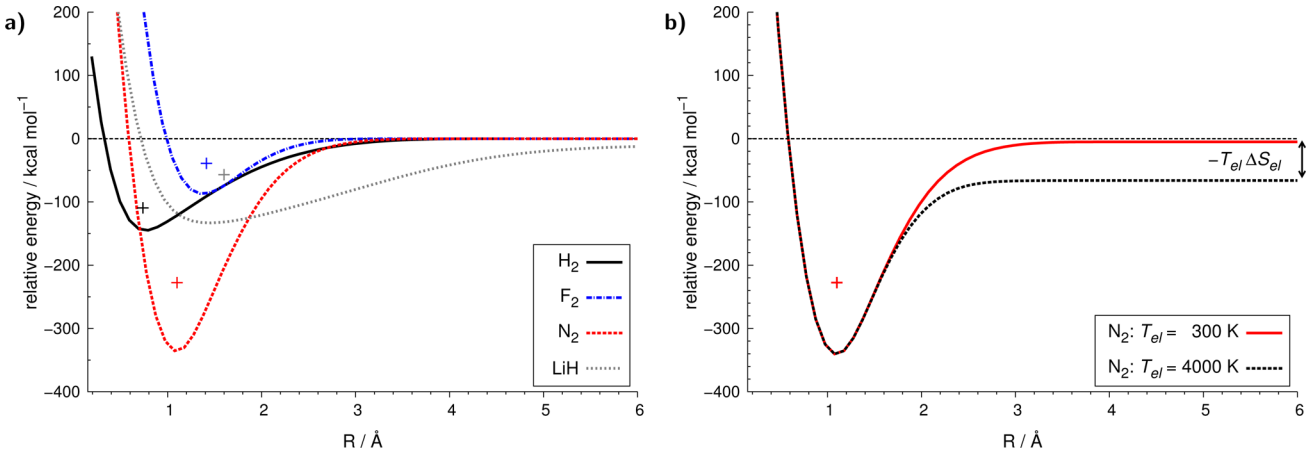


Figure 7. (a) Potential energy curves for the dissociation of H_2, F_2, N_2, and LiH computed with GFN-xTB (T_{el} = 300 K). The points mark the position of the minimum obtained from high-level calculations for F_2,[142] N_2,[142] and LiH.[143] The H_2 reference point is computed at the FCI/aug-cc-pV6Z//MP2/aug-cc-pV6Z level of theory (this work). The energies are given relative to the free atoms (S = 3/2 for nitrogen, S = 1/2 for the others). (b) The potential energy curve for N_2 computed at two different electronic temperatures T_{el} = 300 K and T_{el} = 4000 K. The difference in relative energy from zero at $R \rightarrow \infty$ results from the difference in the electronic entropy ΔS_{el} for dissociated N_2 (S = 0) and free atoms (S = 3/2).

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009

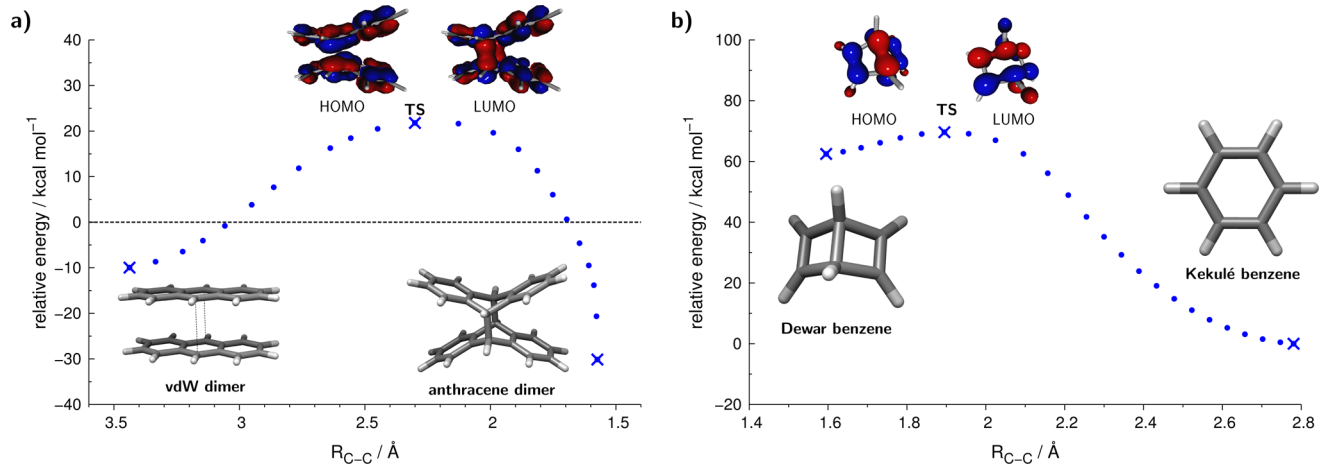


Figure 8. (a) Potential energy curve for the anthracene dimer formation (T_{el} = 4000 K). The energy is given relative to the free anthracene molecules. The frontier molecular orbitals are plotted at the transition-state (TS) geometry as well. (b) Potential energy curve for the conversion of Dewar benzene to Kekulé benzene (T_{el} = 14,000 K). The frontier molecular orbitals are plotted at the estimated transition-state (TS) geometry.

The use of Fermi smearing reduces barriers of chemical processes, especially with methods that underestimate the HOMO−LUMO gap such as GGA DFT functionals and TB methods. GFN-xTB yields HOMO−LUMO gaps that are only slightly smaller (by 0.2 eV) than the ones from GGA DFT calculations (see Table S27 in the Supporting Information). The use of the finite temperature technique allows also the investigation of reaction pathways before turning to computationally demanding multireference methods and provides stable scans of the PES without convergence problems during the SCC step. Two examples of orbital symmetry forbidden reaction pathways are given in Figure 8. GFN-xTB gives a smooth PES for the thermally activated (formally forbidden) dimerization of anthracene (Figure 8a) with a transition-state region characterized by fractionally occupied HOMO and LUMO orbitals. For the benzene reaction, the HOMO−LUMO gap is larger, and higher temperature needs to be applied to populate the relevant levels and to obtain a smooth surface around the transition-state geometry.

Since GFN-xTB overestimates bond energies, the method may, however, provide reasonable reaction energetics at larger electronic temperatures, which effectively lowers the bond energies (see Figure 7b). This will facilitate the implementation of GFN-xTB in QCEIMS[144] for the simulation of electron impact mass spectra in the future.

Finally, the dissociation potential energy curve of the ion pair [BMIM][PF_6], which is a typical ionic liquid, is shown in Figure 9. GFN-xTB produces an intermolecular potential energy curve that agrees very well with the high-level coupled cluster reference suggesting use of the approach in molecular dynamics studies of ionic liquids.

3.5. Known Problems. SQM methods compromise accuracy and computational speed. In a TB method, the sources of error mainly are self-interaction error, monopole description of the electrostatic interactions, AO basis set deficiencies, and parametrization errors. Similarly to GGA DFT functionals, TB methods tend to overestimate the delocalization of electrons and produce too small orbital energy gaps. In turn, they implicitly include static correlation effects and hence work well for metallic systems and the dissociation of covalent bonds. At variance from GGAs, the damped Coulomb law (eq 4) provides an asymptotically correct (exchange-correlation) potential. The monopole approximation is a weak part of the

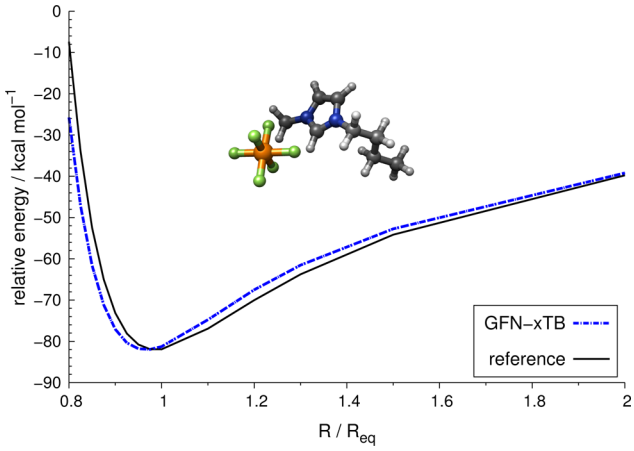


Figure 9. Intermolecular potential energy curve for the dissociation of the [BMIM][PF_6] ion pair computed with GFN-xTB (T_{el} = 300 K) on PBEh-3c geometries. The reference curve is taken from counterpoise corrected DLPNO-CCSD(T)/TightPNO/aug-cc-pVQZ single-point calculations.

approach, and the errors originating from it are difficult to estimate. From our experience, the accuracy diminishes the more polar and more directional the electrostatic interactions are. A good example are the halogen bonds, which require an atom pairwise correction for reasonable performance. Empirical evidence seems to indicate that strong directional electrostatic effects are somewhat emulated by "covalent" interactions in the H_0 part, which is reflected by relatively large sp-shell occupations in the transition-metal complexes compared to first-principles DFT results.

Approximate electrostatic interactions and/or small AO basis sets may lead sometimes to qualitatively incorrect molecular geometries. Extensive experience with GFN-xTB in our lab has singled out the following problematic systems for the new method. Cyclobutane has a folded, D_{2d} symmetric, nonplanar structure (C−C−C−C torsional angle of about 20°), which becomes planar in GFN-xTB and virtually all SQM methods tested. The related and chemically more relevant case of cyclopentane with a C_s envelope structure is less problematic. The ring C−C−C−C torsion angle, which is around 40° at PBEh-3c (or PBE0-D3/def2-TZVP) level, is described well by

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009

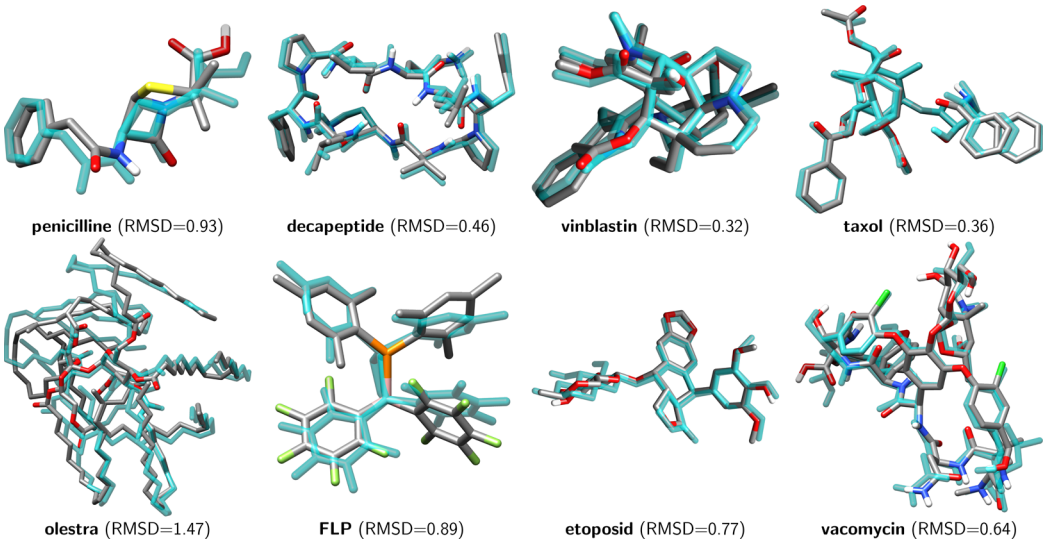


Figure 10. Overlay of eight moderately sized, organic molecules. The structures are optimized with PBEh-3c (transparent cyan) and GFN-xTB (color coded atoms). The RMSD for an all-atom best fit is given in Å. Hydrogen atoms bonded to carbon are not depicted for clarity.

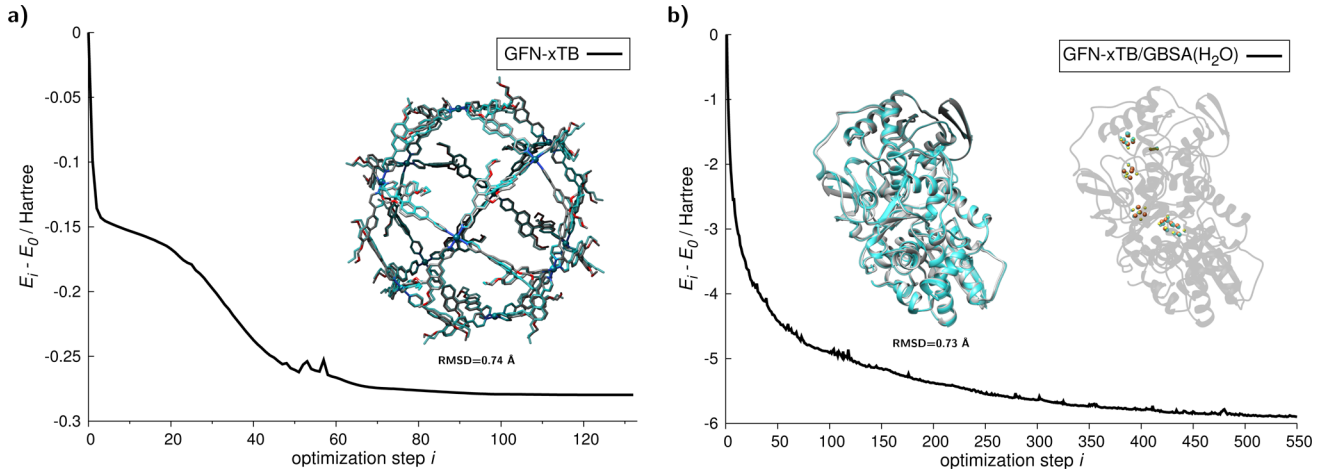


Figure 11. (a) Energy along a full GFN-xTB optimization (133 steps) of the large $[Pd_{12}L_{24}]^{24+}$ complex starting from the *O*-symmetric HF-3c structure. An overlay of the final geometry is shown (hydrogen atoms are omitted for clarity) together with the HF-3c structure (cyan) and the all atom position RMSD. (b) Energy along a GFN-xTB/GBSA(H_2O) optimization (550 steps) of the *Clostridium pasteurianum* [FeFe]-hydrogenase (PDB ID: 3C8Y) with loose convergence thresholds.[150,151] The entire protein contains 8951 atoms (charge: −23). An overlay of the X-ray (cyan) and the optimized GFN-xTB geometry is shown for the protein backbone (left) and the organometallic centers (right). The heavy atom position RMSD is given below the left overlay.

OM2-D3(BJ) (36.5°), GFN-xTB (35.5°), and DFTB3-D3(BJ) (34.1°), while MSINDO-D3(BJ)H+ and PM6-D3H4X give geometries closer to a planar structure (25.7° and 27.5°). Polyynes are common building blocks in supramolecular structures, and their C−C−C bending potentials determine the shape persistence of these complexes.[145] For 1,3-butadiyne, all methods provide reasonable C−C−C bending potentials with harmonic frequencies around 200 cm^{-1}, close to the PBEh-3c value of 240 cm^{-1}. Problems appear with almost all TB- and NDDO-type methods for C_2F_2 (or C_2OMe_2), yielding bent instead of linear minima. The error is more or less absent for PM6-D3H4X and less pronounced for GFN-xTB, which provides linear HC_2F (DFTB3-D3(BJ): bent) and a smaller bending angle for C_2F_2 (21.5° instead of 36.4° for DFTB3-D3(BJ)). This problem can only be repaired by parametrization at the expense of sacrificing accuracy for other hydrocarbons. Another typical example of self-interaction error is C_2 symmetric *cis*-1,3-butadiene, which has a C−C−C−C torsion angle of 35° at the PBE0-D3(BJ)/def2-TZVP level. This diminishes to 32° at the GGA (PBE) level and essentially to zero with GFN-xTB and DFTB3-D3(BJ). The OM2-D3(BJ), MSINDO-D3(BJ)H+, and PM6-D3H4X methods which are closer to HF in character give correctly the nonplanar structure with reasonable torsion angles (29°, 40°, and 46°, respectively).

Related to the self-interaction error problem in (meta-) GGAs and TB methods is the vanishing gap obtained from gas-phase calculations of large polynucleotide, polypeptide chains, or zwitter ions. This way, the electronic structure of these biologically very important systems is as badly described as in (semi)local DFT.[146,147] In GFN-xTB this problem is less severe when the GBSA implicit solvation model[72] is applied. For example, without GBSA, GFN-xTB produces a HOMO−LUMO gap of 0.7 eV for the α-helical $Ac(Ala)_{19}Me$ oligopeptide (20 peptide units, geometry taken from ref 148). Similar observations have been reported for GGAs.[147] GFN-xTB including implicit GBSA(water) solvation yields a much

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009

more realistic gap of 4.5 eV at no significant additional cost. We thus recommend to always use the GBSA solvation model for solvated organic and biomolecular systems and in particular as a workaround for zwitterionic systems where the SCC is otherwise not convergent. We demonstrate the applicability of GFN-xTB/GBSA for a larger, realistic case in the next section. The theory as well as more detailed tests on the GBSA approach will be presented elsewhere.[72]

3.6. Performance for Large Systems. In the previous sections, we discussed the accuracy of the GFN-xTB method for standard benchmark sets containing many small cases, and in the present section, we gauge the speed and applicability of the method and its implementation in the `xtb` code to more extended chemical systems.

We compare GFN-xTB and PBEh-3c structures from a set of relatively large and flexible organic molecules taken from ref 63 in Figure 10. To facilitate the analysis, we provide the all-atom position RMSD below each structure overlay in the same figure. GFN-xTB closely resembles the PBEh-3c structures for the more rigid molecules like taxol and the compounds with significant stabilization from hydrogen bonds like the decapeptide. We see larger deviations for flexible structures, where the PES for dihedral distortions is flat, such as the frustrated Lewis pair (FLP) with rotating aryl substituents, and olestra, which comprises long aliphatic fatty acid chains. Olestra, in particular, contains 453 atoms, which makes DFT calculations very time-consuming; one optimization step with the efficient PBEh-3c method takes about 20 min on 16 CPUs, whereas the full GFN-xTB optimization comprising 454 steps takes only 14 min on 4 CPUs. The GFN-xTB geometries are very reasonable and can be used for single-point calculations or refinements at higher levels of theory.

Systems with several hundreds or thousands of atoms are ubiquitous in biochemistry and supramolecular chemistry. To demonstrate the applicability of GFN-xTB to such systems, we consider the optimization of a large dodecanuclear organometallic polyhedron $[Pd_{12}L_{24}]^{24+}$ comprising 1622 atoms,[149] where L stands for a chiral (*P*)-6,6′-dipyridinyl-2,2′-bis-(methoxymethoxy)-1,1′-binaphthyl ligand. We have previously computed the electronic circular dichroism spectrum of this molecule using only the *O*-symmetric minimum geometry, for which we could afford a DFT optimization with the TURBOMOLE program suite.[84−86] Here we show the energy during the optimization in Figure 11 together with an overlay of the final GFN-xTB and initial HF-3c structures. The number of optimization steps is 133 pointing to a fast convergence for a system with more than 1600 atoms. Despite the flexibility of the ligands, the all-atom position RMSD between the GFN-xTB and HF-3c geometry is only 0.74 Å.

Finally, we apply GFN-xTB in combination with the GBSA(H_2O) solvation model to optimize an entire [FeFe]-hydrogenase protein complex, which comprises close to 9000 atoms. We took the initial structure (3C8Y) from the protein data bank (PDB),[150,152] removed crystal water molecules, and added hydrogen atoms[153] using the Maestro 11[154] program suite. We performed a combined GFN-xTB/GBSA(H_2O) geometry optimization with loose convergence thresholds for this protein complex. For this protein, which comprises a charge of −23, this optimization with 550 steps took two months on 32 CPUs. We provide a plot of the energy during the optimization along with an overlay of the initial and final structures in Figure 11b. The RMSD of the heavy atom positions of the obtained structure and the starting X-ray structure is only 0.73 Å, which is an indication that the ground-state PES of this system is reasonably well described with GFN-xTB/GBSA(H_2O). The embedded Fe_xS_x (x = 2,4) clusters complicate the quantum chemical description of the hydrogenase due a vanishing HOMO−LUMO gap and demand finite temperature Fermi smearing. The overlay of the inorganic centers in Figure 11b affirms the robustness and consistency of this treatment and the code. Thus, GFN-xTB/GBSA(H_2O) successfully captures the electronic structure of both the organometallic sites and the organic polypeptide.

4. CONCLUSIONS

We developed a robust and broadly applicable semiempirical quantum chemical method for the computation of structures, vibrational frequencies, and noncovalent interactions for molecules across the periodic table. It is connected to an efficient implicit solvation model (GBSA), which we intend to publish in the near future, enabling the simulation of bulk electrostatic screening and molecular surface effects occurring in condensed phases. The size of the typical target system is around 1000 atoms, and initial tests show that the method can be successfully applied to proteins with 3000 atoms or more. Although the focus is on bio/organic systems, an advantageous characteristic of the method is the robust and consistent treatment of inorganic main group and transition-metal complexes. This opens new avenues to the computation of large metallo-protein structures and exploration of the dynamics of large transition-metal complexes. We envision numerous applications in the future, including and not limited to studies of protein structure, protein−protein and protein−ligand interactions, and supramolecular complexes ranging from esoteric molecular machines to giant shape-persistent nanoarchitectures.

A key premise of the present method is its special purpose character. In our view, low-cost semiempirical QM methods cannot describe simultaneously very different chemical properties, such as structures and chemical reaction energies, and the GFN-xTB method (as the name conveys) focuses on structural properties. The efficiently computed structures and vibrational frequencies or the conformers obtained from global search procedures can (and should) be used subsequently for more accurate DFT or WFT refinements. We hope that the method can serve as a general tool in quantum chemistry and in particular recommend GFN-xTB optimized structures (and thermostatistical corrections) in a multilevel scheme together with PBEh-3c single-point energies. Large-scale molecular dynamics, screening of huge molecular spaces (libraries), parametrization of force-fields, or providing input for novel machine learning techniques are obvious other fields of application. A strength of the new approach is that the Hamiltonian contains physically interpretable, element-wise parameters, which can straightforwardly be adjusted to other properties or specific substance classes. Our own current efforts include a reparameterization for ionization potential and electron affinity calculations for electrochemical problems and the use of GFN-xTB electronic structure information in an intermolecular potential. We believe that the new method can help bridge different molecular scales by making possible the consistent computation of reasonable structures and noncovalent interactions for medium-sized (≲100 atoms) and large chemical systems (up to 10,000 atoms) at the same level of theory.

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.7b00118.

Element-specific parameters and the detailed results obtained on the considered test sets (PDF)

Cartesian coordinates of all structures, which were optimized in this study (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: grimme@thch.uni-bonn.de. Phone: +49-228/73-2351.

ORCID

Stefan Grimme: 0000-0002-5844-4371

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the DFG in the framework of the "Gottfried-Wilhelm-Leibniz" prize. P.S. would also like to kindly acknowledge the financial support of the Alexander von Humboldt foundation for a postdoctoral fellowship. The authors thank V. Ásgeirsson, C. Bauer, P. Burger, M. Bursch, M. P. Checinski, S. Ehrlich, and J. Seibert for many initial tests of the method.

■ DEDICATION

Dedicated to Professor Sigrid D. Peyerimhoff on the occasion of her 80th birthday.

■ REFERENCES

(1) Grimme, S.; Brandenburg, J. G.; Bannwarth, C.; Hansen, A. Consistent structures and interactions by density functional theory with small atomic orbital basis sets. *J. Chem. Phys.* **2015**, *143*, 054107.

(2) Brandenburg, J. G.; Caldeweyher, E.; Grimme, S. Screened exchange hybrid density functional for accurate and efficient structures and interaction energies. *Phys. Chem. Chem. Phys.* **2016**, *18*, 15519−15523.

(3) Sure, R.; Grimme, S. Corrected small basis set Hartree-Fock method for large systems. *J. Comput. Chem.* **2013**, *34*, 1672−1685.

(4) Dror, R. O.; Dirks, R. M.; Grossman, J.; Xu, H.; Shaw, D. E. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu. Rev. Biophys.* **2012**, *41*, 429−452.

(5) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *WIREs Comput. Mol. Sci.* **2013**, *3*, 198−210.

(6) Zhu, X.; Lopes, P. E. M.; MacKerell, A. D. Recent developments and applications of the CHARMM force fields. *WIREs Comput. Mol. Sci.* **2012**, *2*, 167−185.

(7) Thiel, W. Semiempirical quantum−chemical methods. *WIREs Comput. Mol. Sci.* **2014**, *4*, 145−157.

(8) Yilmazer, N. D.; Korth, M. Enhanced semiempirical QM methods for biomolecular interactions. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 169−175.

(9) Christensen, A. S.; Kubař, T.; Cui, Q.; Elstner, M. Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications. *Chem. Rev.* **2016**, *116*, 5301−5337.

(10) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *58*, 7260−7268.

(11) Yang, Y.; Yu, H.; York, D.; Cui, Q.; Elstner, M. Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method: Third-Order Expansion of the Density Functional Theory Total Energy and Introduction of a Modified Effective Coulomb Interaction. *J. Phys. Chem. A* **2007**, *111*, 10861−10873.

(12) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* **2011**, *7*, 931−948.

(13) Frauenheim, T.; Seifert, G.; Elstner, M.; Niehaus, T.; Köhler, C.; Amkreutz, M.; Sternberg, M.; Hajnal, Z.; Carlo, A. D.; Suhai, S. Atomistic simulations of complex materials: ground-state and excited-state properties. *J. Phys.: Condens. Matter* **2002**, *14*, 3015.

(14) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. A QM/MM Implementation of the Self-Consistent Charge Density Functional Tight Binding (SCC-DFTB) Method. *J. Phys. Chem. B* **2001**, *105*, 569−585.

(15) Niehaus, T. A.; Suhai, S.; Sala, F. D.; Lugli, P.; Elstner, M.; Seifert, G.; Frauenheim, T. Tight-binding approach to time-dependent density-functional response theory. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2001**, *63*, 085108.

(16) Brandenburg, J. G.; Hochheim, M.; Bredow, T.; Grimme, S. Low-Cost Quantum Chemical Methods for Non-Covalent Interactions. *J. Phys. Chem. Lett.* **2014**, *5*, 4275−4284.

(17) Elstner, M. The SCC-DFTB method and its application to biological systems. *Theor. Chem. Acc.* **2006**, *116*, 316−325.

(18) Elstner, M.; Cui, Q.; Munih, P.; Kaxiras, E.; Frauenheim, T.; Karplus, M. Modeling zinc in biomolecules with the self consistent charge-density functional tight binding (SCC-DFTB) method: Applications to structural and energetic analysis. *J. Comput. Chem.* **2003**, *24*, 565−581.

(19) Elstner, M.; Jalkanen, K. J.; Knapp-Mohammady, M.; Frauenheim, T.; Suhai, S. Energetics and structure of glycine and alanine based model peptides: Approximate SCC-DFTB, AM1 and PM3 methods in comparison with DFT, HF and MP2 calculations. *Chem. Phys.* **2001**, *263*, 203−219.

(20) Zheng, G.; Irle, S.; Morokuma, K. Performance of the DFTB method in comparison to DFT and semiempirical methods for geometries and energies of C20−C86 fullerene isomers. *Chem. Phys. Lett.* **2005**, *412*, 210−216.

(21) Sattelmeyer, K. W.; Tirado-Rives, J.; Jorgensen, W. L. Comparison of SCC-DFTB and NDDO-Based Semiempirical Molecular Orbital Methods for Organic Molecules. *J. Phys. Chem. A* **2006**, *110*, 13551−13559.

(22) Zhou, H.; Tajkhorshid, E.; Frauenheim, T.; Suhai, S.; Elstner, M. Performance of the AM1, PM3, and SCC-DFTB methods in the study of conjugated Schiff base molecules. *Chem. Phys.* **2002**, *277*, 91−103.

(23) Frauenheim, T. *DFTB+ (Density Functional based Tight Binding)*; DFTB.ORG, Universität Bremen: Bremen, Germany, 2008; http://www.dftb.org (August 9, 2014).

(24) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. Chemistry with ADF. *J. Comput. Chem.* **2001**, *22*, 931−967.

(25) *ADF DFTB*, SCM, Theoretical Chemistry, Vrije Universiteit: Amsterdam, The Netherlands, 2013; http://www.scm.com.

(26) 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;

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009

Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J *Gaussian 09 Revision E.01*; Gaussian Inc.: Wallingford, CT, 2009.
(27) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(28) Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173.
(29) Korth, M.; Pitoňák, M.; Řezáč, J.; Hobza, P. A Transferable H-Bonding Correction for Semiempirical Quantum-Chemical Methods. *J. Chem. Theory Comput.* **2010**, *6*, 344−352.
(30) Kromann, J. C.; Christensen, A. S.; Steinmann, C.; Korth, M.; Jensen, J. H. A third-generation dispersion and third-generation hydrogen bonding corrected PM6 method: PM6-D3H+. *PeerJ* **2014**, *2*, e449.
(31) Řezáč, J.; Hobza, P. Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *J. Chem. Theory Comput.* **2012**, *8*, 141−151.
(32) Brahmkshatriya, P. S.; Dobeš, P.; Fanfrlík, J.; Řezáč, J.; Paruch, K.; Bronowska, A.; Lepšík, M.; Hobza, P. Quantum Mechanical Scoring: Structural and Energetic Insights into Cyclin-Dependent Kinase 2 Inhibition by Pyrazolo[1,5-a]pyrimidines. *Curr. Comput.-Aided Drug Des.* **2013**, *9*, 118−129.
(33) Saito, T.; Kitagawa, Y.; Takano, Y. Reparameterization of PM6 Applied to Organic Diradical Molecules. *J. Phys. Chem. A* **2016**, *120*, 8750−8760.
(34) Weber, W.; Thiel, W. Orthogonalization corrections for semiempirical methods. *Theor. Chem. Acc.* **2000**, *103*, 495−506.
(35) Tuttle, T.; Thiel, W. OMx-D: semiempirical methods with orthogonalization and dispersion corrections. Implementation and biochemical application. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2159−2166.
(36) Dral, P. O.; Wu, X.; Spörkel, L.; Koslowski, A.; Weber, W.; Steiger, R.; Scholten, M.; Thiel, W. Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Theory, Implementation, and Parameters. *J. Chem. Theory Comput.* **2016**, *12*, 1082−1096.
(37) Koslowski, A.; Beck, M. E.; Thiel, W. Implementation of a general multireference configuration interaction procedure with analytic gradients in a semiempirical context using the graphical unitary group approach. *J. Comput. Chem.* **2003**, *24*, 714−726.
(38) Grimme, S. Towards First Principles Calculation of Electron Impact Mass Spectra of Molecules. *Angew. Chem., Int. Ed.* **2013**, *52*, 6306−6312.
(39) Stewart, J. J. P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **2013**, *19*, 1−32.
(40) Voityuk, A. A. INDO/X: A New Semiempirical Method for Excited States of Organic and Biological Molecules. *J. Chem. Theory Comput.* **2014**, *10*, 4950−4958.
(41) Gadaczek, I.; Krause, K.; Hintze, K. J.; Bredow, T. MSINDO-sCIS: A New Method for the Calculation of Excited States of Large Molecules. *J. Chem. Theory Comput.* **2011**, *7*, 3675−3685.
(42) Miyamoto, K.; Miller, T. F.; Manby, F. R. Fock-matrix corrections in density-functional theory, and use in embedded mean-field theory. *J. Chem. Theory Comput.* **2016**, *12*, 5811−5822.
(43) Grimme, S.; Bannwarth, C. Ultra-fast computation of electronic spectra for large systems by tight-binding based simplified Tamm-Dancoff approximation (sTDA-xTB). *J. Chem. Phys.* **2016**, *145*, 054103.
(44) Gonzalez-Lafont, A.; Truong, T. N.; Truhlar, D. G. Direct dynamics calculations with NDDO (neglect of diatomic differential overlap) molecular orbital theory with specific reaction parameters. *J. Phys. Chem.* **1991**, *95*, 4618−4627.
(45) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. Class IV charge models: A new semiempirical approach in quantum chemistry. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87−110.
(46) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.
(47) Cramer, C. J.; Truhlar, D. G. AM1-SM2 and PM3-SM3 parameterized SCF solvation models for free energies in aqueous solution. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 629−666.
(48) Bauer, C. A.; Grimme, S. How to Compute Electron Ionization Mass Spectra from First Principles. *J. Phys. Chem. A* **2016**, *120*, 3755−3766.
(49) Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals. *J. Chem. Phys.* **1969**, *51*, 2657−2664.
(50) Mermin, N. D. Thermal Properties of the Inhomogeneous Electron Gas. *Phys. Rev.* **1965**, *137*, 1441−1443.
(51) Grimme, S.; Hansen, A. A Practicable Real-Space Measure and Visualization of Static Electron-Correlation Effects. *Angew. Chem., Int. Ed.* **2015**, *54*, 12308−12313.
(52) Köhler, C.; Seifert, G.; Frauenheim, T. Density functional based calculations for Fe_n ($n \leq 32$). *Chem. Phys.* **2005**, *309*, 23−31.
(53) Gaus, M.; Jin, H.; Demapan, D.; Christensen, A. S.; Goyal, P.; Elstner, M.; Cui, Q. DFTB3 Parametrization for Copper: The Importance of Orbital Angular Momentum Dependence of Hubbard Parameters. *J. Chem. Theory Comput.* **2015**, *11*, 4205−4219.
(54) Nishimoto, K.; Mataga, N. Electronic Structure and Spectra of Some Nitrogen Heterocycles. *Z. Phys. Chem.* **1957**, *12*, 335−338.
(55) Ohno, K. Some Remarks on the Pariser-Parr-Pople Method. *Theor. Chim. Act.* **1964**, *2*, 219.
(56) Klopman, G. A Semiempirical Treatment of Molecular Structures. II. Molecular Terms and Application to Diatomic Molecules. *J. Am. Chem. Soc.* **1964**, *86*, 4550.
(57) Grimme, S. A simplified Tamm-Dancoff density functional approach for the electronic excitation spectra of very large molecules. *J. Chem. Phys.* **2013**, *138*, 244104.
(58) Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23*, 69−89.
(59) Hall, G. G. *Proc. R. Soc. London, Ser. A* **1951**, *205*, 541−552.
(60) Hülsen, M.; Weigand, A.; Dolg, M. Quasirelativistic energy-consistent 4f-in-core pseudopotentials for tetravalent lanthanide elements. *Theor. Chem. Acc.* **2009**, *122*, 23−29.
(61) Mantina, M.; Valero, R.; Cramer, C. J.; Truhlar, D. G. In *CRC Handbook of Chemistry and Physics*, 91st ed.; Haynes, W. M., Ed.; CRC Press: Boca Raton, FL, 2010; pp 9−50.
(62) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(63) Grimme, S. A General Quantum Mechanically Derived Force Field (QMDFF) for Molecules and Condensed Phase Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 4497−4514.
(64) Johnson, E. R.; Becke, A. D. A post-Hartree−Fock model of intermolecular interactions. *J. Chem. Phys.* **2005**, *123*, 024101.
(65) Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **2016**, *116*, 5105−5154.
(66) Řezáč, J.; Hobza, P. A halogen-bonding correction for the semiempirical PM6 method. *Chem. Phys. Lett.* **2011**, *506*, 286−289.
(67) Bredow, T.; Jug, K. In *Encyclopedia of Computational Chemistry (online ed.)*; von Ragué Schleyer, P., Ed.; Wiley: New York, 2004.
(68) Eckert, F.; Pulay, P.; Werner, H.-J. Ab initio geometry optimization for large molecules. *J. Comput. Chem.* **1997**, *18*, 1473−1483.
(69) Lindh, R.; Bernhardsson, A.; Karlström, G.; Malmqvist, P.-Å On the use of a Hessian model function in molecular geometry optimizations. *Chem. Phys. Lett.* **1995**, *241*, 423−428.
(70) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(71) Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H. A. Accessible surface areas as a measure of the thermodynamic parameters of

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009

hydration of peptides. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 3086−3090.
(72) Shushkov, P.; Grimme, S. *manuscript in preparation*.
(73) Please contact xtb@thch.uni-bonn.de for the program.
(74) Levenberg, K. A Method for the Solution of Certain Non-linear Problems in Least Squares. *Q. Appl. Math.* **1944**, *2*, 164−168.
(75) Marquardt, D. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431−441.
(76) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158−6170.
(77) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.
(78) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J. Chem. Theory Comput.* **2012**, *8*, 527−541.
(79) Jurečka, P.; Šponer, J.; Cerny, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985−1993.
(80) Marshall, M. S.; Burns, L. A.; Sherrill, C. D. Basis set convergence of the coupled-cluster correction, $\delta_{MP2}^{CCSD(T)}$: Best practices for benchmarking non-covalent interactions and the attendant revision of the S22, NBC10, HBC6, and HSG databases. *J. Chem. Phys.* **2011**, *135*, 194102.
(81) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427.
(82) Sedlak, R.; Janowski, T.; Pitoňák, M.; Řezáč, J.; Pulay, P.; Hobza, P. Accuracy of Quantum Chemical Methods for Large Noncovalent Complexes. *J. Chem. Theory Comput.* **2013**, *9*, 3364−3374.
(83) Kozuch, S.; Martin, J. M. L. Halogen Bonds: Benchmarks and Theoretical Analysis. *J. Chem. Theory Comput.* **2013**, *9*, 1918−1931.
(84) *TURBOMOLE V7.0 2015*; University of Karlsruhe and Forschungszentrum Karlsruhe GmbH , TURBOMOLE GmbH: Karlsruhe, Germany, 2015; http://www.turbomole.com.
(85) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic Structure Calculations on Workstation Computers: The Program System Turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165−169.
(86) Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. Turbomole. *WIREs Comput. Mol. Sci.* **2014**, *4*, 91−100.
(87) Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta.* **1977**, *44*, 129−138.
(88) Neese, F. The ORCA program system. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73−78.
(89) Vahtras, O.; Almlöf, J.; Feyereisen, M. W. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.* **1993**, *213*, 514−518.
(90) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials. *Theor. Chem. Acc.* **1997**, *97*, 119−124.
(91) Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057−1065.
(92) Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* **2013**, *9*, 338−354.
(93) Gaus, M.; Lu, X.; Elstner, M.; Cui, Q. Parameterization of DFTB3/3OB for Sulfur and Phosphorus for Chemical and Biological Applications. *J. Chem. Theory Comput.* **2014**, *10*, 1518−1537.
(94) Kubillus, M.; Kubař, T.; Gaus, M.; Řezáč, J.; Elstner, M. Parameterization of the DFTB3Method for Br, Ca, Cl, F, I, K, and Na in Organic and Biological Systems. *J. Chem. Theory Comput.* **2015**, *11*, 332−342.
(95) Stewart, J. J. P. *MOPAC2016*; Stewart Computational Chemistry: Colorado Springs, CO, 2016; http://OpenMOPAC.net (August 16, 2016).
(96) *MNDO2005*, version 7.0, MPI für Kohlenforschung: Mülheim, Germany.
(97) Oliveira, A. F.; Philipsen, P.; Heine, T. DFTB Parameters for the Periodic Table, Part 2: Energies and Energy Gradients from Hydrogen to Calcium. *J. Chem. Theory Comput.* **2015**, *11*, 5209−5218.
(98) Zheng, G.; Witek, H. A.; Bobadova-Parvanova, P.; Irle, S.; Musaev, D. G.; Prabhakar, R.; Morokuma, K.; Lundberg, M.; Elstner, M.; Köhler, C.; Frauenheim, T. Parameter Calibration of Transition-Metal Elements for the Spin-Polarized Self-Consistent-Charge Density-Functional Tight-Binding (DFTB) Method: Sc, Ti, Fe, Co, and Ni. *J. Chem. Theory Comput.* **2007**, *3*, 1349−1367.
(99) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(100) Korth, M. Third-Generation Hydrogen-Bonding Corrections for Semiempirical QM Methods and Force Fields. *J. Chem. Theory Comput.* **2010**, *6*, 3808−3816.
(101) Brandenburg, J. G.; Grimme, S. Dispersion Corrected Hartree-Fock and Density Functional Theory for Organic Crystal Structure Prediction. *Top. Curr. Chem.* **2013**, *345*, 1−23.
(102) See http://www.thch.uni-bonn.de/.
(103) Grimme, S.; Steinmetz, M. Effects of London Dispersion Correction in Density Functional Theory on the Structures of Organic Molecules in the Gas Phase. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16031−16042.
(104) Risthaus, T.; Steinmetz, M.; Grimme, S. Implementation of Nuclear Gradients of Range-Separated Hybrid Density Functionals and Benchmarking on Rotational Constants for Organic Molecules. *J. Comput. Chem.* **2014**, *35*, 1509−1516.
(105) Puzzarini, C.; Heckert, M.; Gauss, J. The accuracy of rotational constants predicted by high-level quantum-chemical calculations. I. molecules containing first-row atoms. *J. Chem. Phys.* **2008**, *128*, 194108−194108.
(106) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868. erratum *Phys. Rev. Lett.* 78, 1396 (1997).
(107) Schäfer, A.; Huber, C.; Ahlrichs, R. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J. Chem. Phys.* **1994**, *100*, 5829−5835.
(108) Bühl, M.; Kabrede, H. Geometries of Transition-Metal Complexes from Density-Functional Theory. *J. Chem. Theory Comput.* **2006**, *2*, 1282−1290.
(109) Wagner, J. P.; Schreiner, P. R. London Dispersion in Molecular Chemistry−Reconsidering Steric Effects. *Angew. Chem., Int. Ed.* **2015**, *54*, 12274−12296.
(110) Grimme, S. Supramolecular binding thermodynamics by dispersion corrected density functional theory. *Chem. - Eur. J.* **2012**, *18*, 9955−9964.
(111) Sure, R.; Grimme, S. Comprehensive Benchmark of Association (Free) Energies of Realistic Host−Guest Complexes. *J. Chem. Theory Comput.* **2015**, *11*, 3785−3801.
(112) Bryantsev, V. S.; Diallo, M. S.; van Duin, A. C. T.; Goddard, W. A. Evaluation of B3LYP, X3LYP, and M06-Class Density Functionals for Predicting the Binding Energies of Neutral, Protonated, and Deprotonated Water Clusters. *J. Chem. Theory Comput.* **2009**, *5*, 1016−1026.
(113) Anacker, T.; Friedrich, J. New accurate benchmark energies for large water clusters: DFT is better than expected. *J. Comput. Chem.* **2014**, *35*, 634−643.
(114) Antony, J.; Grimme, S.; Liakos, D. G.; Neese, F. Protein-Ligand Interaction Energies with Dispersion Corrected Density Functional Theory and High-Level Wave Function Based Methods. *J. Phys. Chem. A* **2011**, *115*, 11210−11220.
(115) Zhao, Y.; Truhlar, D. G. Benchmark Databases for Nonbonded Interactions and Their Use To Test Density Functional Theory. *J. Chem. Theory Comput.* **2005**, *1*, 415−432.

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009

(116) Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 4285−4292.
(117) Vogiatzis, K. D.; Klopper, W.; Friedrich, J. Non-covalent Interactions of CO_2 with Functional Groups of Metal−Organic Frameworks from a CCSD(T) Scheme Applicable to Large Systems. *J. Chem. Theory Comput.* **2015**, *11*, 1574−1584.
(118) Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *J. Chem. Phys.* **2013**, *139*, 134101.
(119) Weigend, F.; Furche, F.; Ahlrichs, R. Gaussian basis sets of quadruple zeta quality for atoms H to Kr. *J. Chem. Phys.* **2003**, *119*, 12753−12762.
(120) Valdes, H.; Pluhackova, K.; Pitonák, M.; Řezáč, J.; Hobza, P. Benchmark database on isolated small peptides containing an aromatic side chain: comparison between wave function and density functional theory methods and empirical force field. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2747−2757.
(121) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618−622.
(122) Dunning, T. H., Jr. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(123) Kendall, R. A.; Dunning, T. H., Jr; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796−6806.
(124) Brandenburg, J. G.; Grimme, S. Organic crystal polymorphism: a benchmark for dispersion-corrected mean-field electronic structure methods. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *72*, 502−513.
(125) Gruzman, D.; Karton, A.; Martin, J. M. L. Performance of Ab Initio and Density Functional Methods for Conformational Equilibria of $C_nH_{2n.2}$ Alkane Isomers (n = 4−8). *J. Phys. Chem. A* **2009**, *113*, 11974−11983.
(126) Wilke, J. J.; Lind, M. C.; Schaefer, H. F.; Császár, A. G.; Allen, W. D. Conformers of Gaseous Cysteine. *J. Chem. Theory Comput.* **2009**, *5*, 1511−1523.
(127) Řeha, D.; Valdés, H.; Vondrášek, J.; Hobza, P.; Abu-Riziq, A.; Crews, B.; de Vries, M. S. Structure and IR Spectrum of Phenylalanyl−Glycyl-Glycine Tripetide in the Gas−Phase: IR/UV Experiments, Ab Initio Quantum Chemical Calculations, and Molecular Dynamic Simulations. *Chem. - Eur. J.* **2005**, *11*, 6803−6817.
(128) Kozuch, S.; Martin, J. M. L. Spin-component-scaled double hybrids: An extensive search for the best fifth-rung functionals blending DFT and perturbation theory. *J. Comput. Chem.* **2013**, *34*, 2327−2344.
(129) Liakos, D. G.; Neese, F. Domain Based Pair Natural Orbital Coupled Cluster Studies on Linear and Folded Alkane Chains. *J. Chem. Theory Comput.* **2015**, *11*, 2137−2143.
(130) Kruse, H.; Mladek, A.; Gkionis, K.; Hansen, A.; Grimme, S.; Sponer, J. Quantum Chemical Benchmark Study on 46 RNA Backbone Families Using a Dinucleotide Unit. *J. Chem. Theory Comput.* **2015**, *11*, 4972−4991.
(131) Grimme, S.; Steinmetz, M.; Korth, M. How to Compute Isomerization Energies of Organic Molecules with Quantum Chemical Methods. *J. Org. Chem.* **2007**, *72*, 2118−2126.
(132) Huenerbein, R.; Schirmer, B.; Moellmann, J.; Grimme, S. Effects of London dispersion on the isomerization reactions of large organic molecules: a density functional benchmark study. *Phys. Chem. Chem. Phys.* **2010**, *12*, 6940−6948.
(133) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behaviour. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098−3100.
(134) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.
(135) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006**, *27*, 1787−1799.
(136) Goerigk, L.; Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670−6688.
(137) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Assessment of Gaussian-3 and density functional theories for a larger experimental test set. *J. Chem. Phys.* **2000**, *112*, 7374−7383.
(138) Sattelmeyer, K. W.; Tirado-Rives, J.; Jorgensen, W. L. Comparison of SCC-DFTB and NDDO-Based Semiempirical Molecular Orbital Methods for Organic Molecules. *J. Phys. Chem. A* **2006**, *110*, 13551−13559.
(139) Otte, N.; Scholten, M.; Thiel, W. Looking at Self-Consistent-Charge Density Functional Tight Binding from a Semiempirical Perspective. *J. Phys. Chem. A* **2007**, *111*, 5751−5755.
(140) Grimme, S. Accurate Calculation of the Heats of Formation for Large Main Group Compounds with Spin-Component Scaled MP2Methods. *J. Phys. Chem. A* **2005**, *109*, 3067−3077.
(141) Medvedev, M. G.; Bushmarinov, I. S.; Sun, J.; Perdew, J. P.; Lyssenko, K. A. Density functional theory is straying from the path toward the exact functional. *Science* **2017**, *355*, 49−52.
(142) Bytautas, L.; Ruedenberg, K. Correlation energy extrapolation by intrinsic scaling. IV. Accurate binding energies of the homonuclear diatomic molecules carbon, nitrogen, oxygen, and fluorine. *J. Chem. Phys.* **2005**, *122*, 154110.
(143) Maniero, A. M.; Acioli, P. H. Full configuration interaction pseudopotential determination of the ground-state potential energy curves of Li_2 and LiH. *Int. J. Quantum Chem.* **2005**, *103*, 711−717.
(144) Ásgeirsson, V.; Bauer, C. A.; Grimme, S. *submitted*.
(145) Schweez, C.; Shushkov, P.; Grimme, S.; Höger, S. Synthesis and Dynamics of Nanosized Phenylene−Ethynylene−Butadiynylene Rotaxanes and the Role of Shape Persistence. *Angew. Chem., Int. Ed.* **2016**, *55*, 3328−3333.
(146) Grimme, S.; Hujo, W.; Kirchner, B. Performance of dispersion-corrected density functional theory for the interactions in ionic liquids. *Phys. Chem. Chem. Phys.* **2012**, *14*, 4875−4883.
(147) Rudberg, E.; Rubensson, E. H.; Sałek, P. Kohn−Sham Density Functional Theory Electronic Structure Calculations with Linearly Scaling Computational Time and Memory Usage. *J. Chem. Theory Comput.* **2011**, *7*, 340−350.
(148) Bannwarth, C.; Grimme, S. Electronic Circular Dichroism of Highly Conjugated π-Systems: Breakdown of the Tamm-Dancoff/Configuration Interaction Singles Approximation. *J. Phys. Chem. A* **2015**, *119*, 3653−3662.
(149) Gütz, C.; Hovorka, R.; Klein, C.; Jiang, Q.-Q.; Bannwarth, C.; Engeser, M.; Schmuck, C.; Assenmacher, W.; Mader, W.; Topić, F.; Rissanen, K.; Grimme, S.; Lützen, A. Enantiomerically Pure [M6L12] or [M12L24] Polyhedra from Flexible Bis(Pyridine) Ligands. *Angew. Chem., Int. Ed.* **2014**, *53*, 1693−1698.
(150) http://www.rcsb.org/pdb.
(151) Pandey, A. S.; Harris, T. V.; Giles, L. J.; Peters, J. W.; Szilagyi, R. K. Dithiomethylether as a Ligand in the Hydrogenase H-Cluster. *J. Am. Chem. Soc.* **2008**, *130*, 4533−4540.
(152) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(153) Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 221−234.
(154) *MS Jaguar*, 2016−4; Schrödinger LLC, New York, NY, 2016.

DOI: 10.1021/acs.jctc.7b00118
J. Chem. Theory Comput. 2017, 13, 1989−2009