GFN2-xTB – an Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions

Christoph Bannwarth,[*,†,‡] Sebastian Ehlert,[†] and Stefan Grimme[*,†]

†*Mulliken Center for Theoretical Chemistry, Universität Bonn, Beringstr. 4, 53115 Bonn, Germany*

‡*New address: Department of Chemistry, Stanford University, Stanford, CA 94305, United States of America.*

E-mail: christoph.bannwarth@stanford.edu; grimme@thch.uni-bonn.de

Phone: +49-228/73-2351

Abstract

Keywords: density functional tight-binding, semiempirical molecular orbital theory, multipole electrostatics, non-covalent interactions

Abstract

An extended semiempirical tight-binding model is presented, which is primarily designed for the fast calculation of structures and non-covalent interactions energies for molecular systems with roughly 1000 atoms. The essential novelty in this so-called GFN2-xTB method is the inclusion of anisotropic second order density fluctuation effects via short-range damped interactions of cumulative atomic multipole moments. Without noticeable increase in the computational demands, this results in a less empirical and overall more physically sound method, which does not require any classical halogen or hydrogen bonding corrections and which relies solely on global and element-specific parameters (available up to radon, $Z = 86$). Moreover, the atomic partial charge dependent D4 London dispersion model is incorporated self-consistently, which can be naturally obtained in a tight-binding picture from second order density fluctuations. Fully analytical and numerically precise gradients (nuclear forces) are implemented. The accuracy of the method is benchmarked for a wide variety of systems and compared with other semiempirical methods. Along with excellent performance for the "target" properties, we also find lower errors for "off-target" properties such as barrier heights and molecular dipole moments. High computational efficiency along with the improved physics compared to it precursor GFN-xTB makes this method well-suited to explore the conformational space of molecular systems. Significant improvements are futhermore observed for various benchmark sets, which are prototypical for biomolecular systems in aqueous solution.

1 Introduction

The accurate description of large molecular systems particularly in the condensed phase still remains one of the main challenges faced in theoretical chemistry.[1,2] While Kohn-Sham density functional theory (DFT) can routinely provide gas-phase (or continuum embedded) structures and energies for roughly a few hundred atoms, long-time molecular dynamics simulations or conformational sampling are still not feasible in a reasonably sized atomic orbital (AO) basis set for systems of this size, e.g., within a day of computation time on a standard desktop computer. Recent efforts of our[3–6] and other groups[7–10] have thus been directed towards the development of low-cost DFT and Hartree-Fock methods to allow such calculations.

However, many interesting problems in biochemistry, materials science, supramolecular and macromolecular chemistry would require methods that can handle several thousands of atoms, which is beyond the scope even for the aforementioned low-cost methods. Though classical force-fields (FFs) can routinely be used for nanosecond timescale molecular dynamics simulations,[11] their limitations are manifold and they are not suited for general use. Though faster computers and more importantly parallel computing architectures along with accordingly adjusted codes have become available in recent years,[12–25] there is still need for inherently simple, reasonable accurate, and efficient electronic structure methods, which are applicable to large systems also without specialized hardware.

Semiempirical quantum mechanical (SQM) methods[26–28] are the physically motivated approach to bridge the gap between *ab initio* quantum mechanical (QM) and FF methods. The foundation of SQM methods typically is a valence-only minimal basis self-consistent field method, either derived from Hartree-Fock theory or Kohn-Sham DFT. Since a QM description for the valence electrons is used in SQM models, their scope of applications is much broader than that of FFs. At the same time, they are faster by at least two orders of magnitude compared to *ab initio* QM methods, due to drastic integral approximations.[28] These, however, come at the price of a significantly reduced accuracy and robustness. Nevertheless, if parameters are adjusted carefully, SQM methods can be used sufficiently well for screening purposes of a desired property.[29–34] Nowadays, frequently used sophisticated Hartree-Fock-based zero-differential overlap (ZDO) methods include PM6[35–40] and

OM2.[41–44] While the former mostly finds use in the computation of ground state structures and energies,[45–48] the latter is still extensively used in excited state dynamics studies.[49–53] In the past twenty years, density functional tight binding (DFTB) methods have found much attention.[54–61] Here, the Kohn-Sham DFT energy is expanded in terms of density fluctuations $\delta\rho$ relative to a superposition of atomic reference densities. The highest-level variant (DFTB3) employs a self-consistent charge treatment including up to third order density fluctuation terms[55,56] and has been parametrized for a number of chemical elements.[58–62]

In particular, the element pair-specific parametrization used in all aforementioned methods has complicated the parametrization procedure and only PM6 covers large parts of the periodic table of elements (70 elements). Recently, we have presented a DFTB3 variant, termed GFN-xTB, which mostly follows a global and element-specific parameters only strategy and is parametrized to all elements through radon.[34] The original purpose of the method and main target for the parameter optimization has been the computation of molecular **g**eometries, vibrational **f**requencies, and **n**on-covalent interaction energies. As such, it has successfully been used in structure optimizations of organometallic complexes[63,64] and structural sampling.[65–67] Apart from that, the method performed well in high temperature molecular dynamics simulations of electron impact mass spectra[68] and the electronic structure information of GFN-xTB can furthermore serve as input to generate intermolecular FFs.[69] Of particular significance is its contribution to a newly developed protocol for the automated computation of nuclear magnetic resonance (NMR) spectra. Therein, structural sampling with GFN-xTB is employed to generate the conformer-rotamer-ensemble, which in turn defines the detailed shape of an NMR spectrum. While the GFN-xTB could successfully identify the relevant structures for less polar molecules, polar and strongly hydrogen-bonded systems like sugars were not described as good. This motivated the present work. Here, we attempt to improve upon the GFN-xTB Hamiltonian in a physically sound manner. The theory will be described below, but in a nutshell, the method, which will be termed GFN2-xTB from now on, has the following characteristics:

- The GFN2-xTB basis set consists of a minimal valence basis set of atom centered, contracted Gaussian functions, which approximate Slater functions (STO-mG).[70] Polarization functions for most main group elements (typically second row or higher) are employed, which are

particularly important to describe hypervalent states. Different from GFN-xTB, hydrogen is only assigned a single 1s function.

- The GFN2-xTB Hamiltonian closely resembles that of GFN-xTB or of the well-known DFTB3 method. However, GFN2-xTB represents the first tight-binding method to include electrostatic interactions and exchange-correlation effects up to second order in the multipole expansion. Furthermore the simultaneously developed density-dependent D4 dispersion model in an self-consistent formulation is an inherent part of the method. Different from either GFN-xTB or DFTB3, GFN2-xTB does not employ other classical FF-type corrections, e.g., to describe hydrogen or halogen bonds. These are reasonably well described within the multipole-extended electrostatics.

- Different from other semiempirical methods, GFN2-xTB strictly follows a global and element-specific parameter strategy. No element pair-specific parameters are employed.

- As for the predecessor, properties around energetic minima, such as geometries, vibrational frequencies, and non-covalent interactions are the target quantities. Already in our work for the GFN-xTB scheme, we have identified that the preference for geometries (instead of covalent bond energies) in the fit procedure resulted in systematically overestimated covalent bond energies. We have not deviated from this strategy, as covalent bond energies are not of primary interest for this method, and furthermore, the errors are very systematic as in GFN-xTB. This way, the errors will be less random and more useful, once the systematic errors have been removed – see Ref. 45 for such a correction scheme which in combination with GFN-xTB outperforms any other semiempirical method of comparable complexity. By keeping the focus on geometries as for GFN-xTB, the GFN2-xTB method would likewise be well-suited for applications in high (electronic) temperature molecular dynamics simulations to compute electron impact mass spectra.[68] However, due to the more sophisticated physical interaction terms, we find that some "off-target" properties related to the electronic structure also improve.

After a detailed description of the GFN2-xTB Hamiltonian, technical details of the calculations are given. Then the method is benchmarked on standard sets for molecular structures, non-covalent

energies, and conformational energies. The performance for barrier heights and molecular dipole moments is also assessed. Other semiempirical methods are used for comparison, namely GFN-xTB,[34] PM6-D3H4X,[35,38,39] and DFTB3-D3(BJ).[56,58,71]

2 Theory

Most previous derivations of density functional tight-binding (DFTB) are formally based on a (semi-)local density approximation (LDA) of DFT.[28,54–56] Though non-local exchange variants have been presented,[72–74] the electron correlation functional has also been local in those cases. This however neglects long-range correlation effects, which are responsible for the important dispersion interactions. Typically, corrections for the latter are then added in an a posteriori fashion.[71,75] Without explicitly referring to a specific functional approximation, we will assume as starting point a general density functional approximation, which includes a non-local (NL) VV10-type[76] correlation contribution. The total Kohn-Sham energy expression is then given by (atomic units are used throughout)

$$E\left[\rho\right] = \int \rho(\mathbf{r}) \left[T[\rho(\mathbf{r})] + V_n(\mathbf{r}) + \varepsilon_{\mathrm{XC}}^{\mathrm{LDA}}[\rho(\mathbf{r})] + \frac{1}{2} \int \left(\frac{1}{|\mathbf{r}-\mathbf{r}'|} + \Phi_{\mathrm{C}}^{\mathrm{NL}}(\mathbf{r},\mathbf{r}') \right) \rho(\mathbf{r}') \mathrm{d}\mathbf{r}' \right] \mathrm{d}\mathbf{r} + E_{nn} \,. \tag{1}$$

Here, $T[\rho(\mathbf{r})]$ is the kinetic energy per particle and $V_n(\mathbf{r})$ the (external) potential due to the nuclei. In writing so, we are using the short-hand notation for $\int \rho(\mathbf{r}) T[\rho(\mathbf{r})] \mathrm{d}\mathbf{r} = -\sum_i n_i \left\langle \psi_i \left| \frac{1}{2}\nabla_i^2 \right| \psi_i \right\rangle$, with ψ_i being the molecular orbital and n_i the corresponding occupation number. The last three terms in the integral over d$\mathbf{r}$ arise from the electronic contributions to the mean-field potential: the (semi-)local exchange-correlation (XC), the Coulomb, and the non-local correlation energy, respectively. The non-local correlation kernel $\Phi_{\mathrm{C}}^{\mathrm{NL}}(\mathbf{r},\mathbf{r}')$ captures long-range correlation effects. E_{nn} is the classical nuclear repulsion energy. In density functional tight-binding (DFTB) theory, the total energy is expanded in terms of density fluctuations around a superposition of (neutral) atomic reference densities $\rho_0 = \sum_A \rho_{A_0}$:[28,54]

$$E\left[\rho\right] = E^{(0)}\left[\rho_0\right] + E^{(1)}\left[\rho_0, \delta\rho\right] + E^{(2)}\left[\rho_0, (\delta\rho)^2\right] + E^{(3)}\left[\rho_0, (\delta\rho)^3\right] + \cdots \tag{2}$$

A fixed reference density ρ_0 is employed and the calculation of the electronic structure is done in terms of the density fluctuations $\delta\rho$. The most widely used variants truncate this expansion after the third order term.[34,55,56] The same is true for the GFN2-xTB approach presented here. The first approximation made in tight-binding methods is the assumption that the density fluctuations are restricted to the valence orbital space, while the core electron density remains frozen. The individual approximations for the energy contributions to different orders in $\delta\rho$ have been described before,[28,54,56,72,73] but we briefly outline the origin of newly introduced terms in GFN2-xTB.

2.1 Contributions relevant to the GFN2-xTB energy

2.1.1 Zeroth order terms

In previous derivations of DFTB, the zeroth order term is reduced to a classical repulsion term.[28,54–56] Due to the different starting point (DFT-NL), an additional term will arise at first order:

$$E^{(0)}\left[\rho_0\right] \approx \underbrace{E^{(0)}_{\mathrm{rep}} + E^{(0)}_{\mathrm{disp}}}_{\text{Lennard-Jones-type term}} + \sum_A \left(E^{(0)}_{A,\mathrm{core}} + E^{(0)}_{A,\mathrm{valence}} \right) . \tag{3}$$

Here, the total energy in Eq. 1 is partitioned into same-center and off-center energy terms. All one-electron and two-electron terms in Eq. 1 involving particles at the same atom reduce to a single number, the atomic energy E^A, which can furthermore be partitioned into a core and a valence contribution (see last two terms in Eq. 3). Since the reference densities refer to neutral, spherically symmetric atoms, the summed Coulomb interactions ($e-e$, $n-n$, $e-n$) between distant atoms vanish at this level of approximation. However, in regions with overlapping atomic densities, repulsive forces due to exchange-correlation and charge penetration effects are present. Similar to previous works on DFTB, this term is modeled by a classical repulsion term.[28,54,56,72,73] Different from DFTB but in line with our previous method GFN-xTB,[34] no element pair-specific parametrization will be employed for this term. Due to the presence of a NL correlation functional in the energy expression in Eq. 1, a pairwise London dispersion contribution is also present at zeroth order (see Supporting Information, Section 1.5). The sum of the atomic energies is a constant for a given system, and the total tight-binding energy is given relative to the sum of atomic core energies,

or equivalently they are set to zero, i.e., $E^{(0)}_{A,\text{core}} := 0$. $E^{(0)}_{A,\text{valence}}$ is the sum of the valence orbital energies, which represent eigensolutions of the Hamiltonian for the free atom.

$$E^{(0)}_{A,\text{valence}} = \sum_{l \in A} \sum_{\kappa \in l} P^0_{\kappa\kappa} H^l_A \tag{4}$$

The zeroth order density matrix $\mathbf{P}^0$ is diagonal, as the atomic orbitals with energy H^l_A are assumed to be eigenstates of the free atom. E_rep and $E^{(0)}_\text{disp}$ will be described by classical expressions between "clamped" atoms in GFN2-xTB (see Eqs. 9 and 32). It can be seen that the zeroth order energy in the $\delta\rho$ expansion is thus related to the well-known Lennard-Jones energy, and can be a reasonable approximation to treat non-covalent interactions of noble gas atoms. It does not contribute to the tight-binding electronic energy, which depends only on the fluctuations $\delta\rho$.

2.1.2 First order terms

For the choice of spherical atomic reference densities ρ_{A_0}, the description of covalent bonds in tight-binding theories becomes possible at first order in the $\delta\rho$ expansion. This typically achieved by changes of in the occupation $\delta\mathbf{P}$ of the atomic energy levels by invoking the well-known extended Hückel approximation. Since $\delta\rho$ is assumed to be non-zero only in the valence electron space (see above), the latter will be restricted to the valence orbitals. Due to the density fluctuations to first order, the atoms can obtain a net charge and are no longer neutral. This has no effect on the interatomic electrostatic (ES) interactions, since the electrostatic potential from the remaining neutral atoms (with ρ_{A_0}) is still zero. However, since starting from the DFT-NL expression in Eq. 1 in GFN2-xTB, the energy to first order in $\delta\rho$ is augmented with a first order dispersion contribution:

$$E^{(1)}\left[\rho_0, \delta\rho\right] \;\approx\; E^{(1)}_\text{disp} \;+\; \sum_A E^{(1)}_{A,\text{valence}} \tag{5}$$

The last term describes first order changes in the occupation and energies of the electronic valence levels. If combined with Eq. 4, this is essentially equivalent to extended Hückel theory (EHT), i.e., $E_\text{EHT} = \sum_A \left(E^{(0)}_{A,\text{valence}} + E^{(1)}_{A,\text{valence}}\right)$. The explicit expression to approximate this term in GFN2-xTB will be given in Section 2.2. Due to the non-vanishing zeroth order London dispersion

potential of the other charge-neutral atoms, the dispersion energy changes at first order in $\delta\rho$. The expansion/contraction of the atomic density will increase/decrease the magnitude of the respective interaction. In GFN2-xTB these and second order effects (see below) will be taken into account within the self-consistent D4 dispersion model (see Section 2.2.6).[77,78]

2.1.3 Second order terms

At second order in $\delta\rho$, energy contributions arise, which require a self-consistent tight-binding calculation.

$$E^{(2)}\left[\rho_0, (\delta\rho)^2\right] \approx E^{(2)}_{\mathrm{ES}} + E^{(2)}_{\mathrm{XC}} + E^{(2)}_{\mathrm{disp}}\,. \tag{6}$$

At second order, interatomic electrostatic and one-center exchange-correlation terms occur. In DFTB schemes, these are generally condensed in a Mataga-Nishimoto-Ohno-Klopman damped Coulomb interaction between atomic or shell charge monopoles.[79–81] The same will be done in GFN2-xTB (see Section 2.2). For the first time in a tight-binding scheme, however, we will go beyond the monopole approximation for both $E^{(2)}_{\mathrm{ES}}$ and $E^{(2)}_{\mathrm{XC}}$ (see Section 2.2.5) and include anisotropic effects up to second order in the multipole expansion. Isotropic second order $\delta\rho$ effects in the dispersion energy are also included within the self-consistent D4 dispersion model in GFN2-xTB.

2.1.4 Third order terms

Basically all energy contributions of the Kohn-Sham energy expression in Eq. 1 that require a self-consistent solution at the DFT level of theory will also contribute to energy corrections of higher order (≥ 2) in $\delta\rho$. As in previous works on DFTB and GFN-xTB,[34,55] we will neglect all third order terms but an isotropic on-site term, which mostly originates from short-ranged Coulomb and XC effects.

$$E^{(3)}\left[\rho_0, (\delta\rho)^3\right] \approx E^{(3)}_{\mathrm{ES}} + E^{(3)}_{\mathrm{XC}} \approx \frac{1}{3}\sum_A \sum_{l\in A} \Gamma_{A,l}\, q^3_{A,l} \tag{7}$$

Different from those previous works, however, the third order term is expressed in terms of partial shell charges (Mulliken approximation) in GFN2-xTB.

2.2 The GFN2-xTB method

Having related the individual energy components to different orders in $\delta\rho$, we will outline the specific contributions to the GFN2-xTB energy in the following. The total GFN2-xTB energy expression is given by

$$E_{\text{GFN2-xTB}} = E_{\text{rep}} + E_{\text{disp}} + E_{\text{EHT}} + E_{\text{IES+IXC}} + E_{\text{AES}} + E_{\text{AXC}} + G_{\text{Fermi}}\,. \tag{8}$$

The newly introduced abbreviations in the subscripts indicate the isotropic electrostatic (IES) and isotropic XC (IXC), and likewise the anisotropic electrostatic (AES) and anisotropic XC (AXC) energies, respectively. Before going into details, it is important to note that no halogen- or hydrogen-bond corrections are included in GFN2-xTB. The description of these interactions is already improved by the AES energy E_{AES}, which is described in Section 2.2.5.

2.2.1 The classical repulsion energy

For the repulsion energy in Eq. 8, we employ an atom pairwise potential similar to the one proposed in Refs. 34,82.

$$E_{\text{rep}} = \sum_{AB} \frac{Y_A^{\text{eff}} Y_B^{\text{eff}}}{R_{AB}} e^{-(\alpha_A \alpha_B)^{\frac{1}{2}} (R_{AB})^{k_{\text{rep}}}}\,, \tag{9}$$

where Y_A^{eff} and Y_B^{eff} define the magnitude of the repulsive interaction. Like α_A and α_B, they are element-specific parameters. k_{rep} is a parameter, which is equal to one if both atoms are either H or He, and equal to 3/2 otherwise. The different k_{rep} parameter for the very light element pairs turned out to be beneficial for torsion barriers in alkanes, without sacrificing the accuracy for non-covalent complexes. The repulsion energy in GFN2-xTB is described classically and is independent of changes in the electronic structure. It should be mentioned that compared to GFN-xTB,[34] the Y^{eff} values correlate less with the atomic number in GFN2-xTB.

2.2.2 Choice of electronic wavefunction and finite temperature treatment

As in the predecessor GFN-xTB, we are working with a formally spin-restricted wavefunction throughout and no spin density dependent terms are present. Hence, α and β orbitals always have

identical spatial parts and orbital energies but possibly different occupation. In order to handle static correlation (nearly degenerate states) via fractional orbital occupations, a finite temperature treatment is used. The last term G_{Fermi} in Eq. 8 formally refers to the entropic contribution of an electronic free energy at finite electronic temperature T_{el} due to Fermi smearing.[83] This term is necessary to provide a variational solution for fractional occupations and is given by

$$G_{\mathrm{Fermi}} = k_{\mathrm{B}} T_{\mathrm{el}} \sum_{\sigma=\alpha,\beta} \sum_{i} \left[n_{i\sigma} \ln\left(n_{i\sigma}\right) + \left(1 - n_{i\sigma}\right) \ln\left(1 - n_{i\sigma}\right) \right] . \tag{10}$$

k_{B} is Boltzmann's constant and $n_{i\sigma}$ refers to the (fractional) occupation number of the spin-MO $\psi_{i\sigma}$. These are given by

$$n_{i\sigma} = \frac{1}{\exp[(\epsilon_i - \epsilon_{\mathrm{F}}^{\sigma})/(k_{\mathrm{B}} T_{\mathrm{el}})] + 1} . \tag{11}$$

ϵ_i is the orbital energy of the orbital ψ_i and $\epsilon_{\mathrm{F}}^{\sigma} = 0.5(\epsilon_{\mathrm{HOMO}}^{\sigma} + \epsilon_{\mathrm{LUMO}}^{\sigma})$ is the Fermi level for the respective orbital space (α or β). T_{el} the electronic temperature, which by default is equal to 300 K as in GFN-xTB.[34]

The occupation n_i for the spatial molecular orbital ψ_i (which is the same for $\psi_{i\alpha}$ and $\psi_{i\beta}$) is given as

$$n_i = n_{i\alpha} + n_{i\beta} . \tag{12}$$

It should be stressed that this finite temperature treatment predominantly has the purpose of enabling fractional orbital occupations in static correlation cases.

As usual, the spatial molecular orbitals (MOs) ψ_i are expressed as linear combinations of atom-centered orbitals (LCAO)[84,85],

$$\psi_i = \sum_{\kappa}^{N_{AO}} c_{\kappa i} \phi_{\kappa}(\zeta_{\kappa}, \text{STO-}m\text{G}) . \tag{13}$$

Following Stewart's Gaussian expansions,[70] ϕ_{κ} refer to contracted Gaussian atomic orbitals, which are used to approximate a spherical Slater-type orbital with exponent ζ_{κ}. The number of primitives m varies between 3 and 6 – explicit numbers of primitives are given in the supporting information (SI). In GFN2-xTB, a minimal spherical valence basis set is employed, whereas most heavier main

group elements ($Z > 9$) are provided with a polarization function. AOs located on the same atom are always orthogonal to each other in GFN2-xTB, thus formally forming a basis of eigenstates of the free atom. The basis set employed is given in Table 1, whereas the Slater exponents are listed in the SI. As in GFN-xTB, the "f-in-core" approximation is employed for lanthanides. Variational

Table 1: Slater-type AO basis sets employed for the different elements. n denotes the principal quantum number of the valence shell of the element.

element	basis functions
H	ns
He	ns,$(n+1)$p
group 1, Be–F, Zn, Cd, Hg–Po	$n$sp
Ne	$n$sp,$(n+1)$d
group 2, 13–18	$n$spd
transition metals and lanthanides	nd,$(n+1)$sp

minimization of the energy expression in Eq. 8 with respect to the linear coefficients $c_{\kappa i}$ in Eq. 13 leads to the general eigenvalue problem, which is likewise encountered in Hartree-Fock and Kohn-Sham density functional theory

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\epsilon}\,. \tag{14}$$

The elements of the tight-binding Hamiltonian or "Fock" matrix $\mathbf{F}$ will be given after presentation of the individual energy terms. $\mathbf{S}$ is the overlap matrix and $\mathbf{C}$ is the LCAO-MO coefficient matrix. $\boldsymbol{\epsilon}$ is a diagonal matrix containing the orbital energies.

2.2.3 The extended Hückel-type energy

The extended Hückel contribution is the crucial ingredient to describe covalent bonds in tight-binding methods.

$$E_{\mathrm{EHT}} = \sum_i n_i \langle\psi_i|\hat{H}_0|\psi_i\rangle = \sum_\kappa \sum_\lambda \sum_i n_i c_{\kappa i} c_{\lambda i} \langle\phi_\lambda|\hat{H}_0|\phi_\kappa\rangle \equiv \sum_\kappa \sum_\lambda P_{\kappa\lambda} H_{\lambda\kappa}\,. \tag{15}$$

Here, the density matrix element $P_{\kappa\lambda} = P^0_{\kappa\lambda} + \delta P_{\kappa\lambda}$. Due to its origin in the first order $\delta\rho$ fluctuation term, the operator $\hat{H}_0$ is formally a one-electron operator, which should however provide the zeroth

order energy (Eq. 4) for neutral non-interacting atoms. The corresponding off-diagonal matrix elements $H_{\kappa\lambda}$ are thus approximated in the following way:

$$H_{\kappa\lambda} = k_{ll'}\frac{1}{2}(H_{\kappa\kappa} + H_{\lambda\lambda})S_{\kappa\lambda}\left(\frac{2\sqrt{\zeta_\kappa\zeta_\lambda}}{\zeta_\kappa + \zeta_\lambda}\right)^{\frac{1}{2}}(1 + k_{EN}\Delta EN_{AB}^2)\Pi(R_{AB,ll'})\ , \quad (\kappa \in l \in A, \lambda \in l' \in B) \tag{16}$$

$k_{ll'}$ is the effective scaling factor common to all EHT methods. Its value depends on the angular momentum quantum number of the interacting AOs (see Table 2 for explicit values). Similar to GFN-xTB,[34] the atomic energy levels, i.e., the diagonal elements of the EHT Hamiltonian matrix, are made flexible by being proportional to the coordination number,

$$H_{\kappa\kappa} = H_A^l - H_{CN_A}^l CN'_A, \quad (\kappa \in l \in A\,) \tag{17}$$

with

$$CN'_A = \sum_{B\neq A}^{N_{atoms}} \left(1 + e^{-10(4(R_{A,cov}+R_{B,cov})/3R_{AB}-1)}\right)^{-1}\left(1 + e^{-20(4(R_{A,cov}+R_{B,cov}+2)/3R_{AB}-1)}\right)^{-1} \tag{18}$$

Here, H_A^l and $H_{CN_A}^l$ are element-specific parameters. Compared to the D3 coordination number,[86] which has been used in GFN-xTB,[34] CN'_A is smoother and slightly more long-ranged. The covalent radii of the atoms $R_{A,cov}/R_{B,cov}$ are the rescaled radii from Ref. 87 as used in the D3[86] and D4[77] (see below) model. $S_{\kappa\lambda}$ is the overlap between the two AOs. The last three terms in the product on the right hand side of Eq. 16 modify the magnitude of the interaction in certain situations. ζ_κ and ζ_λ are the Slater exponents of the two AOs. This factor is unity for $\zeta_\kappa = \zeta_\lambda$ and otherwise reduces the magnitude of the $H_{\kappa\lambda}$ matrix element. Similarly, the second last term reduces the "covalent" interaction for two atoms with different electronegativities. While the latter term exclusively adjusts non-polar vs. polar (or ionic) binding, the former term adds more flexibility for orbitals with different compactness, and thus, is also relevant in homonuclear situations. The explicit dependence on orbital exponents introduces effects present for HF/DFT in the kinetic energy one-electron integrals. As in GFN-xTB,[34] the distance and shell dependent polynomial $\Pi(R_{AB,ll'})$ is based on element-specific parameters $k_{A,l}^{poly}/k_{B,l'}^{poly}$, which are fitted, while the $R_{A,cov'}/R_{B,cov'}$ are taken from

Table 2: Global empirical parameters of the GFN2-xTB method. The parameters are either dimensionless or in atomic units.

parameter	value
k_{ss}	1.85
k_{pp},k_{dd}	2.23
k_{sp}	2.04[a]
k_{sd},k_{pd}	2.00
k_{rep}	1.5[b]
K_s	1.0
K_p	0.5
K_d	0.25
k_{EN}	0.02
multipole parameters	
Δ_{val}	1.2
R_{max}	5.0
a_3	3.0
a_5	4.0
dispersion parameters	
a_1	0.52
a_2	5.0
s_6	1.0
s_8	2.7
s_9	5.0

[a] Obtained as $k_{sp} = 0.5(k_{ss} + k_{pp})$
[b] $k_{\mathrm{rep}} = 1.0$ for H/He pairs.

Ref. 88.

$$\Pi(R_{AB,ll'}) = \left(1 + k_{A,l}^{poly}\left(\frac{R_{AB}}{R_{cov',AB}}\right)^{\frac{1}{2}}\right)\left(1 + k_{B,l'}^{poly}\left(\frac{R_{AB}}{R_{cov',AB}}\right)^{\frac{1}{2}}\right) \tag{19}$$

The key purpose of this term is the distance-dependent adjustment of the EHT-type interaction, which provides a better balance between short- (covalent) and long-ranged (non-covalent) effects. The EHT-type term in GFN2-xTB is mostly responsible for covalent binding. Via the coordination number dependence of the valence energy levels, these obtain additional flexibility beyond the $\delta\rho$-based expansion up to first order. To some extent, the fitted element- and shell-specific parameters for H_A^l and $H_{CN_A}^l$ can implicitly account for the formally neglected first and second order on-site $\delta\rho$ effects. Noteworthy is the importance of that term for atoms that can become hypervalent: the energy level for the d-polarization functions are typically high in energy for small CN'_A. Then, these

functions do neither interfere strongly with the valence functions in "standard" covalent binding nor do they significantly affect non-covalent interactions in a way which is reminiscent to the well-known basis set superposition error (BSSE) in AO-based *ab-initio* calculations. For large CN'_A values, however, this d-level is lowered in energy, which then enables a much better description of hypervalency.

2.2.4 The isotropic electrostatic and exchange-correlation energy

For charged and polar systems, the density ρ deviates from the reference density ρ_0. In that case, the net partial charges on the individual atoms are non-zero. In GFN2-xTB, the isotropic electrostatic and XC terms are treated with shell-wise partitioned Mulliken partial charges (cf. Ref. 89). The following contribution to the energy is then obtained:

$$E_{\text{IES+IXC}} = \frac{1}{2}\sum_{A,B}\sum_{l\in A}\sum_{l'\in B} q_{A,l}q_{B,l'}\gamma_{AB,ll'} + \frac{1}{3}\sum_{A}\sum_{l\in A}\Gamma_{A,l}q_{A,l}^3\,. \tag{20}$$

Here, the first term on the right-hand side is derived from the second order energy, while the last term is due to third order density fluctuations. $q_{A,l}$ refers to an isotropic monopole charge of the l shell on atom A. The distance dependence of the Coulomb interaction within the first term is described by a generalized form of the well-established Mataga-Nishimoto-Ohno-Klopman [79–81,90] formula:

$$\gamma_{AB,ll'} = \left(\frac{1}{R_{AB}^2 + \eta^{-2}}\right)^{\frac{1}{2}} \tag{21}$$

Here, R_{AB} is the interatomic distance and η is the average of the effective chemical hardnesses of the two shells l and l' on the atoms A and B:

$$\eta = \frac{1}{2}\left(\eta_{A,l} + \eta_{B,l'}\right) = \frac{1}{2}\left((1+\kappa_A^l)\eta_A + (1+\kappa_B^{l'})\eta_B\right) \tag{22}$$

η_A and η_B are treated as element-specific fit parameters. κ_A^l and $\kappa_B^{l'}$ are fitted element-specific scaling factors for the individual shells (note that $\kappa_A^l = 0$ for $l = 0$). This way, electrostatic interactions between distant atoms and on-site changes in the isotropic electrostatic and XC energy are treated in a seamless manner. The third order term is restricted to shell-wise on-site terms. As

discussed in the literature on DFTB,[56] this term can help to stabilize charged atomic states and partially remedy shortcomings from missing, e.g., diffuse functions in the AO basis.

The shell-wise parameter $\Gamma_{A,l}$ is obtained from the element-specific parameter Γ_A and the global, shell-specific parameters K_l.

$$\Gamma_{A,l} = K_l \, \Gamma_A \tag{23}$$

Γ_A is formally related to $\partial\eta_A/\partial q_A|_{q_A=0}$, but is a fitted parameter in GFN2-xTB. At variance with GFN-xTB, we use a shell-wise, though parameter-economic treatment by treating K_l as global parameters. The shell-wise treatment appeared to be beneficial for some transition metals without diminishing the accuracy for main group elements.

The shell-wise treatment requires the definition of reference valence shell occupations. For the occupation of elements of group 1, 2, 12, 13, 16, 17, and 18, we follow the aufbau principle, whereas for transition metals a modified aufbau principle of the form $n\mathrm{d}^{x-2}(n+1)\mathrm{s}^1(n+1)\mathrm{p}^1$ (x denotes the group) is used. Lighter elements of group 14 and 15 are handled slightly different to better reflect the occupations in bonded atoms. For this purpose, carbon is treated with a reference valence shell occupation of $2\mathrm{s}^1 2\mathrm{p}^3$, whereas fractional reference occupations of type $n\mathrm{s}^{1.5} n\mathrm{p}^{x-1.5}$ (x denotes the group) are used for N, Si, Ge, P, and As. The remaining elements of group 14 and 15 follow the standard aufbau principle. All element-specific parameters are given in the SI.

2.2.5 The multipole-extended electrostatic and exchange-correlation energy

The approximate expression for the ES and XC energy commonly employed in tight-binding theory is derived from the first two terms of the second order energy (see Eq. 6):[56]

$$E^{(2)}_{\mathrm{ES}} + E^{(2)}_{\mathrm{XC}} = \frac{1}{2} \iint \left(\frac{1}{r_{ij}} + \left. \frac{\partial^2 E_{\mathrm{XC}}}{\partial\rho(\mathbf{r}_i)\partial\rho(\mathbf{r}_j)} \right|_{\rho=\rho_0} \right) \delta\rho(\mathbf{r}_i)\delta\rho(\mathbf{r}_j)\mathrm{d}\mathbf{r}_i\mathrm{d}\mathbf{r}_j \tag{24}$$

Anisotropic electrostatic interactions In Figure 1, it is schematically shown how Eq. 24 is typically approximated by purely isotropic energy terms in DFTB.[54,56] γ_{AB} is the interatomic Coulomb interaction, which is damped to a finite value at short-range (see Eq. 21). This short-range damping then also includes effects due to the second order changes in the semi-local XC energy $E^{(2)}_{\mathrm{XC}}$. Apart from different partitioning schemes (atomic or shell-wise, see Section 2.2.4),

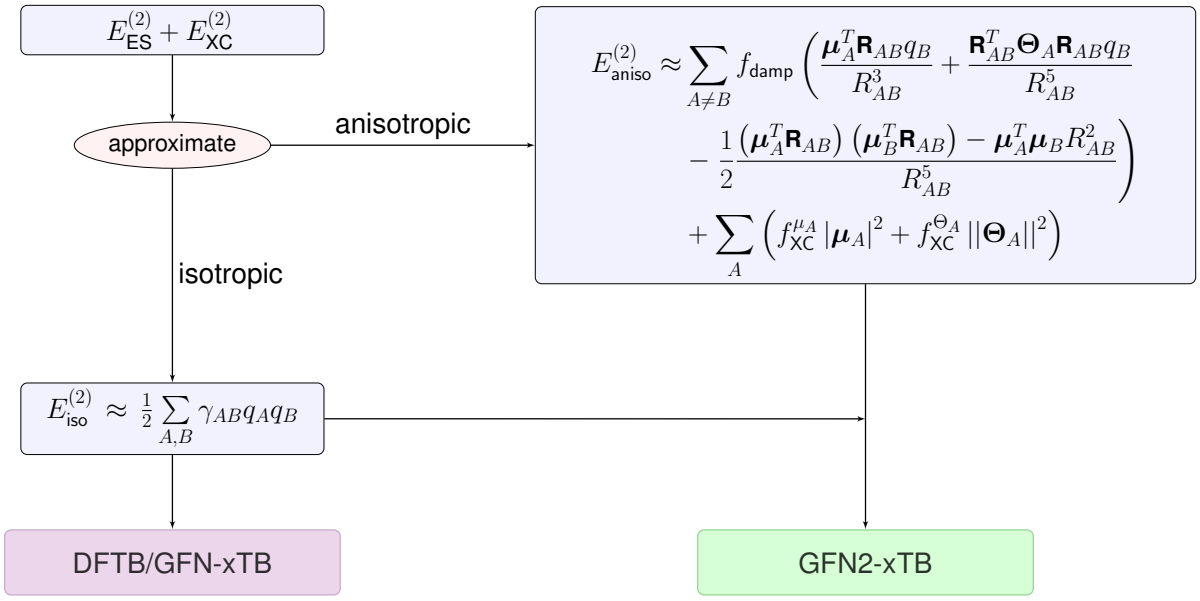


Figure 1: Schematic overview of the employed approximations for the second order ES and XC energy in tight-binding theory. While the isotopic approximation is well-established,[34,54,56] the full inclusion of anisotropic effects up to second order in the multipole expansion is novel. For simplicity, an atomic charge partitioning is used throughout in this scheme.

the same functional form for the second order ES/XC energy is used in DFTB,[54,56] GFN-xTB,[34] and the GFN2-xTB method presented here. The basic possibility of including higher multipole ES interactions in DFTB has been suggested in Ref. 91. Nevertheless, only the first order charge-dipole term has been presented[91] and no implementation in a functioning DFTB method has been reported so far. In GFN2-xTB, we pioneer in going beyond this monopole approximation for both, ES and XC terms including all terms up to second order in the multipole expansion. The newly incorporated terms are schematically shown in Figure 1 and outlined below (for a derivation, see SI). The AES energy in GFN2-xTB is given by

$$E_{\mathrm{AES}} = E_{q\mu} + E_{q\Theta} + E_{\mu\mu} \tag{25a}$$

$$= \frac{1}{2} \sum_{A,B} \left\{ f_3(R_{AB}) \left[q_A \left(\boldsymbol{\mu}_B^T \mathbf{R}_{BA} \right) + q_B \left(\boldsymbol{\mu}_A^T \mathbf{R}_{AB} \right) \right] \right. \tag{25b}$$

$$+ f_5(R_{AB}) \left[q_A \mathbf{R}_{AB}^T \boldsymbol{\Theta}_B \mathbf{R}_{AB} + q_B \mathbf{R}_{AB}^T \boldsymbol{\Theta}_A \mathbf{R}_{AB} \right. \tag{25c}$$

$$\left. \left. - 3 \left(\boldsymbol{\mu}_A^T \mathbf{R}_{AB} \right) \left(\boldsymbol{\mu}_B^T \mathbf{R}_{AB} \right) + \left(\boldsymbol{\mu}_A^T \boldsymbol{\mu}_B \right) R_{AB}^2 \right] \right\} \ . \tag{25d}$$

Here, $\boldsymbol{\mu}_A$ is the cumulative atomic dipole moment of atom A and $\boldsymbol{\Theta}_A$ is the corresponding traceless quadrupole moment

$$\Theta_A^{\alpha\beta} = \frac{3}{2}\theta_A^{\alpha\beta} - \frac{\delta_{\alpha\beta}}{2}\left(\theta_A^{xx} + \theta_A^{yy} + \theta_A^{zz}\right) . \tag{26}$$

The cumulative atomic multipole moments (CAMM)[92] up to second order are computed from:

$$q_A = Z_A - \sum_{\kappa\in A}\sum_{\lambda} P_{\kappa\lambda} \underbrace{\langle \phi_\lambda \,|\, \phi_\kappa \rangle}_{S_{\lambda\kappa}} \tag{27a}$$

$$\mu_A^\alpha = \sum_{\kappa\in A}\sum_{\lambda} P_{\kappa\lambda} \left(\alpha_A S_{\lambda\kappa} - \underbrace{\langle \phi_\lambda \,|\alpha_i|\, \phi_\kappa \rangle}_{D^\alpha_{\lambda\kappa}} \right) \tag{27b}$$

$$\theta_A^{\alpha\beta} = \sum_{\kappa\in A}\sum_{\lambda} P_{\kappa\lambda} \left(\alpha_A D^\beta_{\lambda\kappa} + \beta_A D^\alpha_{\lambda\kappa} - \alpha_A\beta_A S_{\lambda\kappa} - \underbrace{\langle \phi_\lambda \,|\alpha_i\beta_i|\, \phi_\kappa \rangle}_{Q^{\alpha\beta}_{\lambda\kappa}} \right) \tag{27c}$$

α and β are Cartesian components. $D^\alpha_{\lambda\kappa}$ and $Q^{\alpha\beta}_{\lambda\kappa}$ are the electric dipole and quadrupole moment integrals between the AOs ϕ_κ and ϕ_λ. The expressions for the CAMMs directly originate from the multipole expansion (in Cartesian coordinates) and guarantee that the respective overall molecular moments are correctly preserved. These expressions give the atomic contribution (Mulliken approximation) of the particular atom to the overall multipole moment. The CAMMs are defined as such that their respective origin is located at the particular atom, i.e., they are origin-independent, as enforced by the "shift" contributions from lower order moments (Eqs. 27b and 27c). In Eq. 25, we have gone up to second order in the multipole expansion of the Coulomb energy, thus all terms that decay with R_{AB}^{-3} or slower are included (see Eq. 28). The monopole-monopole term, which is the term of lowest order (see SI), has already been included in the shell-wise isotropic ES energy described in Section 2.2.4. The terms containing higher order multipoles should improve the description of the anisotropic electron density around the atoms. We chose to employ an atomic partitioning in GFN2-xTB for these terms. To avoid divergence for the AES energy (Eq. 25), we damp the corresponding terms at short distances. The distance dependence including damping is given by

$$f_n(R_{AB}) = \frac{f_{\text{damp}}(a_n, R_{AB})}{R_{AB}^n} = \frac{1}{R_{AB}^n} \cdot \frac{1}{1 + 6\left(\frac{R_0^{AB}}{R_{AB}}\right)^{a_n}} \tag{28}$$

The damping function is related to the zero damping function in the original D3 dispersion model.[86] a_n are adjusted global parameters, whereas $R_0^{AB} = 0.5\left(R_0^{A\prime} + R_0^{B\prime}\right)$ determines the damping of the AES interaction. $R_0^{A\prime}$ is made dependent on the D3 coordination number for many lighter elements.

$$R_0^{A\prime} = R_0^A + \frac{R_{\max} - R_0^A}{1 + \exp[-4(CN_A - N_{\text{val}} - \Delta_{\text{val}})]} \tag{29}$$

$\Delta_{\text{val}} = 1.2$ and $R_{\max} = 5.0$ bohrs. Aside from those light elements (see SI), $R_0^{A\prime} = 5.0$ bohrs for all elements. This flexible $R_0^{A\prime}$ value reduces the strength of the AES interactions for strongly coordinated atoms and was found to increase the robustness of the SCF convergence for inorganic clusters. Primarily, the AES terms are intended to improve the non-covalent interactions between the outer, i.e., less coordinated atoms. This way, no extra hydrogen or halogen bond corrections nor any element-specific bond adaptations are required.

Anisotropic XC energy While the basic idea of including AES terms in a TB model has been suggested before,[91] no such extension for second order XC effects was mentioned so far. In the context of excitation energies, the approach of Dominguez et al. to include INDO-like terms in DFTB is somewhat related, though its basis is rather founded in a semiempirical hybrid-like density functional. Here, we take the second term of Eq. 24 as starting point. The second order XC energy contribution takes the form of a static XC kernel. In the local density approximation, this term reduces to a pure same-site energy in a tight-binding scheme. Going up to second order multipolar terms, $E_{\text{XC}}^{(2)}$ can then be simplified to (see SI for the derivation):

$$E_{\text{XC}}^{(2)} \approx \sum_A \left(\underbrace{f_{\text{XC}}^{q_A} q_A^2}_{\text{isotropic XC}} + \underbrace{f_{\text{XC}}^{\mu_A} \left|\boldsymbol{\mu}_A\right|^2 + f_{\text{XC}}^{\Theta_A} \left|\left|\boldsymbol{\Theta}_A\right|\right|^2}_{\text{anisotropic XC}} \right) \tag{30}$$

The isotropic monopolar term is already included in the shell-wise isotropic XC energy (see Section 2.2.4). To our knowledge, the other terms, are proposed here for the first time in a tight-binding context. They form the anisotropic XC energy in GFN2-xTB:

$$E_{\text{AXC}} = \sum_A \left(f_{\text{XC}}^{\mu_A} \left|\boldsymbol{\mu}_A\right|^2 + f_{\text{XC}}^{\Theta_A} \left|\left|\boldsymbol{\Theta}_A\right|\right|^2 \right) \tag{31}$$

Again, $\boldsymbol{\mu}_A$ and $\boldsymbol{\Theta}_A$ are the cumulative atomic dipole and traceless quadrupole moments, which have already been introduced in Eq. 27 and 26. $f_{\mathrm{XC}}^{\mu_A}$ and $f_{\mathrm{XC}}^{\Theta_A}$ are fitted element-specific parameters. Formally, these terms are supposed to capture changes in the atomic XC energy, which result from anisotropic density distributions (polarization). To some extent, they may also alleviate shortcomings of the small AO basis set (e.g., insufficient polarization functions).

2.2.6 The density-dependent dispersion energy

In GFN2-xTB, we treat dispersion interactions by means of a self-consistent variant of the recently published D4 dispersion model.[77,78] In typical semiempirical mean-field methods, London dispersion interactions are generally treated by means of post-SCF corrections (see Refs. 28 and 75 for reviews). Though the widely employed D3 dispersion model takes environmental effects via the geometric coordination number into account, electronic structure effects are missing. In a tight-binding context, the D3 dispersion energy should be regarded as a zeroth order term, i.e., it corresponds to $E_{\mathrm{disp}}^{(0)}$. Here, we go beyond this model and include effects up to second order within the self-consistent formulation of the D4 dispersion model:[78]

$$\begin{aligned} E_{\mathrm{disp}} =& E_{\mathrm{disp}}^{(0)} + E_{\mathrm{disp}}^{(1)} + E_{\mathrm{disp}}^{(2)} \\ \approx& -\sum_{A>B}\sum_{n=6,8} s_n \frac{C_n^{AB}(q_A, CN_{\mathrm{cov}}^A, q_B, CN_{\mathrm{cov}}^B)}{R_{AB}^n} f_n^{\mathrm{damp,BJ}}(R_{AB}) \\ & - s_9 \sum_{A>B>C} \frac{(3\cos(\theta_{ABC})\cos(\theta_{BCA})\cos(\theta_{CAB})+1) C_9^{ABC}(CN_{\mathrm{cov}}^A, CN_{\mathrm{cov}}^B, CN_{\mathrm{cov}}^C)}{(R_{AB}R_{AC}R_{BC})^3} \\ & \times f_9^{\mathrm{damp,zero}}(R_{AB}, R_{AC}, R_{BC}) \end{aligned} \tag{32}$$

The last term is the charge-independent three-body (also called Axilrod-Teller-Muto or ATM) dispersion term,[93,94] which is added to incorporate the dominant part of the many-body dispersion energy. Different from the charge-dependent two-body term, this three-body contribution does not affect the electronic energy.

The damping functions $f_n^{\mathrm{damp,BJ}}$ and $f_9^{\mathrm{damp,zero}}$ in Eq. 32 have been defined in Refs. 95 and 86, respectively. It is, however, important to note that in the D4 model, the BJ-type cutoff radii[95] are used in both $f_n^{\mathrm{damp,BJ}}$ and $f_9^{\mathrm{damp,zero}}$ (see below). While the C_8^{AB} are calculated recursively[86]

from the lowest order C_6^{AB} coefficients, the latter are computed from a numerical Casimir-Polder integration.

$$C_6^{AB} = \frac{3}{\pi} \sum_j w_j \overline{\alpha}_A(i\omega_j, q_A, CN_{\text{cov}}^A) \overline{\alpha}_B(i\omega_j, q_B, CN_{\text{cov}}^B) \tag{33}$$

w_j are the integration weights, which are derived from a trapezoidal partitioning between the grid points j ($j \in [1, 23]$). The isotropically averaged, dynamic dipole-dipole polarizabilites $\overline{\alpha}_A$ at the j^{th} imaginary frequency $i\omega_j$ are obtained from the self-consistent D4 model, i.e., they are depending on the covalent coordination number,[78] and are also charge dependent. The method thus relies on precomputed atomic polarizabilities at a certain molecular geometry, i.e., with the atom having a GFN2-xTB computed atomic partial charge of $q_{A,r}$ and a covalent coordination number $CN_{\text{cov}}^{A,r}$ (the index r indicates values for the reference structures). Similar to D3, a Gaussian-weighting scheme based on the covalent coordination number CN_{cov}^A via the W_A^r terms (see Ref. 77) is employed.

$$\overline{\alpha}_A(i\omega_j, q_A, CN_{\text{cov}}^A) = \sum_r^{N_{A,\text{ref}}} \xi_A^r(q_A, q_{A,r}) \overline{\alpha}_{A,r}(i\omega_j, q_{A,r}, CN_{\text{cov}}^{A,r}) W_A^r(CN_{\text{cov}}^A, CN_{\text{cov}}^{A,r}) \tag{34}$$

The Gaussian-weighting for each reference system is given by

$$\begin{aligned} W_A^r(CN_{\text{cov}}^A, CN_{\text{cov}}^{A,r}) = \sum_{j=1}^{N_{\text{gauss}}} \frac{1}{\mathcal{N}} \exp\left[-6j \cdot \left(CN_{\text{cov}}^A - CN_{\text{cov}}^{A,r}\right)^2\right] \\ \text{with} \quad \sum_r^{N_{A,\text{ref}}} W_A^r(CN_{\text{cov}}^A, CN_{\text{cov}}^{A,r}) = 1 \end{aligned} \tag{35}$$

where $\mathcal{N}$ is a normalization constant. The number of Gaussian functions per reference system N_{gauss} is mostly one, but is equal to three for $CN_{\text{cov}}^{A,r} = 0$ and reference systems with similar coordination number (see Ref. 77 for details) The charge-dependency is included via the empirical scaling function ξ_A^r.

$$\xi_A^r(q_A, q_{A,r}) = \exp\left[3\left\{1 - \exp\left[4\eta_A\left(1 - \frac{Z_A^{\text{eff}} + q_{A,r}}{Z_A^{\text{eff}} + q_A}\right)\right]\right\}\right] \tag{36}$$

Where η_A is the chemical hardness taken from Ref. 96. Z_A^{eff} is the effective nuclear charge of atom A, which has been determined by substracting the number of core electrons represented by the

def2-ECPs in the time-dependent DFT reference calculations (see Ref. 77 for details). Due to the charge dependency, the pairwise dispersion energy in GFN2-xTB enters the electronic energy and is self-consistently optimized. A similar expression is used in the standard form of the DFT-D4 method,[77] but therein, the partial charges are obtained by purely geometrical means. The rational damping function used in the DFT-D4 model is given by

$$f_n^{\mathrm{damp,BJ}}(R_{AB}) = \frac{R_{AB}^n}{R_{AB}^n + (a_1 \cdot R_{AB}^{\mathrm{crit.}} + a_2)^6} \qquad \text{with} \quad R_{AB}^{\mathrm{crit.}} = \sqrt{\frac{C_8^{AB}}{C_6^{AB}}} \,. \tag{37}$$

The zero damping function for the ATM dispersion is defined slightly different to the previous implementations of DFT-D3, namely the factor 4/3 is dropped and the cutoff radii are calculated consistently to the two-body terms.

$$f_9^{\mathrm{damp,zero}}(R_{AB}, R_{AC}, R_{BC}) = \left(1 + 6\left(\sqrt[3]{\frac{R_{AB}^{\mathrm{crit.}} R_{BC}^{\mathrm{crit.}} R_{CA}^{\mathrm{crit.}}}{R_{AB} R_{BC} R_{CA}}}\right)^{16}\right)^{-1} \tag{38}$$

2.3 The GFN2-xTB Hamiltonian matrix

As mentioned before, GFN2-xTB includes energy terms of second and third order in $\delta\rho$. Therefore, the energy expression in Eq. 8 needs to be solved self-consistently. To compute the density matrix, the Roothaan-Hall-type eigenvalue problem (Eq. 14) is then solved.

As for the total energy, the matrix elements of the GFN2-xTB Hamiltonian can be decomposed into individual contributions

$$F_{\kappa\lambda} = H_{\kappa\lambda} + F_{\kappa\lambda}^{\mathrm{IES+IXC}} + F_{\kappa\lambda}^{\mathrm{AES}} + F_{\kappa\lambda}^{\mathrm{AXC}} + F_{\kappa\lambda}^{\mathrm{D4}} \,. \tag{39}$$

Due to the analogy to Hartree-Fock, we will denote this matrix simply as TB-Fock matrix in the following. The extended Hückel matrix elements $H_{\kappa\lambda}$ have already been described in Eq. 16. The general derivation for the isotropic ES and XC contributions to the TB-Fock matrix has been shown

in Ref. 56. In GFN2-xTB, these are given by

$$F_{\kappa\lambda}^{\mathrm{IES+IXC}} = -\frac{1}{2}S_{\kappa\lambda}\sum_{C}\sum_{l''}(\gamma_{AC,ll''}+\gamma_{BC,l'l''})q_{C,l''} - \frac{1}{2}S_{\kappa\lambda}(q_{A,l}^2\Gamma_{A,l}+q_{B,l'}^2\Gamma_{B,l'}) \quad (\kappa \in l(A), \lambda \in l'(B))\,. \tag{40}$$

where indices κ and λ denote the AOs with corresponding angular momenta l and l' and the second sum runs over the atoms C and their shells l''.

The last three terms in Eq. 39 are new, and their derivation can be found in the SI. $F_{\kappa\lambda}^{\mathrm{AES}}$ and $F_{\kappa\lambda}^{\mathrm{AXC}}$ both involve terms that include electric dipole and quadrupole, as well as overlap integrals. In a condensed notation, they can be expressed as

$$F_{\kappa\lambda}^{\mathrm{AES}} + F_{\kappa\lambda}^{\mathrm{AXC}} = \frac{1}{2}S_{\kappa\lambda}\left[V_S(\mathbf{R}_B) + V_S(\mathbf{R}_C)\right] \tag{41a}$$

$$+\frac{1}{2}\mathbf{D}_{\kappa\lambda}^T\left[\mathbf{V}_D(\mathbf{R}_B) + \mathbf{V}_D(\mathbf{R}_C)\right] \tag{41b}$$

$$+\frac{1}{2}\sum_{\alpha,\beta}Q_{\kappa\lambda}^{\alpha\beta}\left[V_Q^{\alpha\beta}(\mathbf{R}_B) + V_Q^{\alpha\beta}(\mathbf{R}_C)\right]\,,\ \forall\ \kappa \in B, \lambda \in C\,. \tag{41c}$$

Here, the respective integral (overlap, dipole, and quadrupole) proportional potential terms are given as

$$\begin{aligned} V_S(\mathbf{R}_C) = &\sum_A \big\{\mathbf{R}_C^T\left[f_5(R_{AC})\boldsymbol{\mu}_A R_{AC}^2 - \mathbf{R}_{AC}3f_5(R_{AC})\left(\boldsymbol{\mu}_A^T\mathbf{R}_{AC}\right) - f_3(R_{AC})q_A\mathbf{R}_{AC}\right] \\ &- f_5(R_{AC})\mathbf{R}_{AC}^T\boldsymbol{\Theta}_A\mathbf{R}_{AC} - f_3(R_{AC})\boldsymbol{\mu}_A^T\mathbf{R}_{AC} + q_A f_5(R_{AC})\frac{1}{2}\mathbf{R}_C^2\mathbf{R}_{AC}^2 \\ &- \frac{3}{2}q_A f_5(R_{AC})\sum_{\alpha,\beta}\alpha_{AB}\beta_{AB}\alpha_C\beta_C\big\} \\ &+ 2f_{XC}^{\mu_C}\mathbf{R}_C^T\boldsymbol{\mu}_C - f_{XC}^{\Theta_C}\mathbf{R}_C^T\left[3\boldsymbol{\Theta}_C - \mathrm{Tr}(\boldsymbol{\Theta}_C)\,\mathbf{I}\right]\mathbf{R}_C \end{aligned} \tag{42}$$

$$\begin{aligned} \mathbf{V}_D(\mathbf{R}_C) = &\sum_A \Big[\mathbf{R}_{AC}3f_5(R_{AC})\left(\boldsymbol{\mu}_A^T\mathbf{R}_{AC}\right) - f_5(R_{AC})\boldsymbol{\mu}_A R_{AC}^2 + f_3(R_{AC})q_A\mathbf{R}_{AC} \\ &- q_A f_5(R_{AC})\mathbf{R}_C R_{AC}^2 + 3q_A f_5(R_{AC})\mathbf{R}_{AC}\sum_{\alpha}\alpha_C\alpha_{AC}\Big] \\ &- 2f_{XC}^{\mu_C}\boldsymbol{\mu}_C + 2f_{XC}^{\Theta_C}\left[3\boldsymbol{\Theta}_C - \mathrm{Tr}(\boldsymbol{\Theta}_C)\,\mathbf{I}\right]\mathbf{R}_C \end{aligned} \tag{43}$$

$$
\begin{aligned}
V_Q^{\alpha\beta}(\mathbf{R}_C) = &- \sum_A q_A f_5(R_{AC}) \left[\frac{3}{2} \alpha_{AC} \beta_{AC} - \frac{1}{2} R_{AB}^2 \right] \\
&- f_{XC}^{\Theta_C} \left[3\Theta_C^{\alpha\beta} - \delta_{\alpha\beta} \sum_\alpha \Theta_C^{\alpha\alpha} \right]
\end{aligned}
\tag{44}
$$

The last line in each of the previous equations describes the AXC potential, whereas the remaining terms define the AES potential experienced at position $\mathbf{R}_C$. It is stressed that the electric multipole moment integrals are given with origin at $\mathbf{O}= (000)^T$. Hence, $V_S(\mathbf{R}_C)$ and $\mathbf{V}_D(\mathbf{R}_C)$ also include higher order potential terms due to the "shift" terms in the CAMM definition (see Eq. 27). Analytical first nuclear derivatives are not shown here, but are derived and given in the SI. However it is noted here, that the expressions for the analytical gradients become much simpler if the multipole integrals are given with origin at the respective atomic position.

The TB-Fock matrix contribution from self-consistent D4 is given by

$$
F_{\kappa\lambda}^{\mathrm{D4}} = -\frac{1}{2} S_{\kappa\lambda} (d_A + d_B), \forall \kappa \in A, \lambda \in B \tag{45}
$$

where d_A is given by (we are dropping the dependency on q and CN_{cov} for brevity)

$$
d_A = \sum_r^{N_{A,\mathrm{ref}}} \frac{\partial \xi_A^r}{\partial q_A} \sum_B \sum_s^{N_{B,\mathrm{ref}}} \sum_{n=6,8} W_A^r W_B^s \xi_B^s \cdot s_n \frac{C_n^{AB,\mathrm{ref}}}{R_{AB}^n} f_n^{\mathrm{damp,BJ}}(R_{AB}) \ . \tag{46}
$$

Here, the dispersion coefficient for two reference atoms $C_n^{AB,\mathrm{ref}}$ is evaluated at the reference points, i.e., for $q_A = q_r$, $q_B = q_s$, $CN_{\mathrm{cov}}^A = CN_{\mathrm{cov}}^r$, and $CN_{\mathrm{cov}}^B = CN_{\mathrm{cov}}^s$.

2.4 Technical details

The global parameters in Table 2 and the element-specific parameters (see SI) have been determined by minimizing the root-mean-square deviation (RMSD) between reference and GFN2-xTB-computed data. The procedure is the same as in GFN-xTB an relies on the Levenberg-Marquardt algorithm[97,98].

The global parameters have been determined along with the element-specific parameters for the

elements H, C, N, and O. Then the element-specific parameters for the other elements were subsequently determined keeping all existing parameters fixed. For the lanthanides, only the parameters for Ce and Lu were freely fitted, while a linear interpolation with the nuclear charge Z has been used for the other elements.

In general, the reference data, which was employed in the parameterization of GFN-xTB, has been extended and used for fitting. The data consisted of molecular equilibrium and non-equilibrium structures (energies and forces), harmonic force constants for up to medium sized systems ($<$ 30 atoms), and non-covalent interaction energies and structures (mainly potential energy curves and a few full optimizations). No atomic partial charges or molecular dipole moments have been used in the fit.

A mixed level of theory for the reference data is used. If systems from standard benchmark sets have been used, basis set extrapolated CCSD(T) energies are typically used, while forces and structures were mostly computed by the more efficient, though sufficiently accurate composite methods PBEh-3c[4] or B97-3c.[6] We used the TURBOMOLE suite of programs[99–101] (version 7.0) to conduct most of the ground state DFT reference calculations and geometry optimizations. The exchange-correlation functional integration grid $m4$ and the SCF convergence criterion ($10^{-7}\,E_h$) along with the resolution of the identity (RI) integral approximation[102–104] has been used.

The GFN2-xTB method parameters have first been determined with a D3-variant for the dispersion energy. Then the D4 parameters and reference values (i.e., $q_{A,\mathrm{r}}$) were determined while keeping all other parameters fixed.

Calculations for comparison with other semiempirical methods were conducted with the `DFTB+`[105] (DFTB3[56] with the 3OB parametrization[58–60]), and `MOPAC16`[106] (PM6-D3H4X[35,38,39]) codes. The DFTB3 method was used in combination with the 3OB parametrization[58–60] and D3(BJ)[71,86,95] dispersion correction. PM6-D3H4X, i.e., with zero-damped D3 dispersion, as implemented in `MOPAC16` is used.[38,86] Our standalone `dftd3` code[107] was used for the calculations of the D3(BJ) corrections. The Fermi smearing technique ($T_{el} = 300\,\mathrm{K}$) has also been employed in the DFTB3 calculations.

In the following, we will use the abbreviations: mean deviation (MD), mean absolute deviation (MAD), standard deviation (SD), mean relative deviation (MRD), mean unsigned relative deviation (MURD), standard relative deviation (SRD), maximum unsigned deviation (MAX), maximum

unsigned relative deviation (MAXR), and regularized relative root mean square error (RMSE).

3 Results and Discussion

3.1 Molecular structures

The ROT34 benchmark set[108,109] has become an established set to assess the performance of quantum chemical methods to compute gas phase structures of organic molecules. The quantity to be compared is the spectroscopically accessible rotational constant B_0, which can quantum chemically be corrected for nuclear vibrational effects to B_e. This "clamped" nuclei rotational constant can then be compared directly to the local Born-Oppenheimer minimum energy structure of the investigated method. As long as conformational changes can be excluded, smaller rotational constants then typically indicate elongated covalent bonds relative to the reference. Since the isoamyl-acetate molecule in the ROT34 set was found to be problematic w.r.t. conformational changes for many semiempirical methods, it is excluded in this work. The data for the other SQM methods is taken from Ref. 34, whereas the detailed results for GFN2-xTB are given in the SI. It has been discussed before that the tight-binding methods show significantly smaller standard relative deviations (SRDs) compared to the ZDO methods OM2-D3(BJ) and PM6-D3H4X.[34] GFN2-xTB ranks second from the considered methods and is outperformed only slightly by its predecessor GFN-xTB. This likely results from the fact that the relative weight of geometries in the fitting procedure has been larger for the GFN-xTB than for the GFN2-xTB method. Furthermore, it should be noted that although GFN-xTB is mostly constructed from global and element-specific parameters, there still are a few element pair-specific "fine-tuning" parameters (e.g., between N–H or H–H pairs). GFN2-xTB relies solely on element-specific and global parameters, thus its performance for these organic structures can be regarded as excellent. In Figure 3, the MADs for three structure benchmark sets are shown, which are more difficult for quantum chemical methods. The LB12[4] consists of 12 molecules with an "unusually" long bond between two atoms. HMGB11[4] contains heavy main group and TMC32[110] contains 50 bond distances in a total of 32 transition metal complexes. GFN-xTB and GFN2-xTB clearly outperform PM6-D3H4X for this purpose.

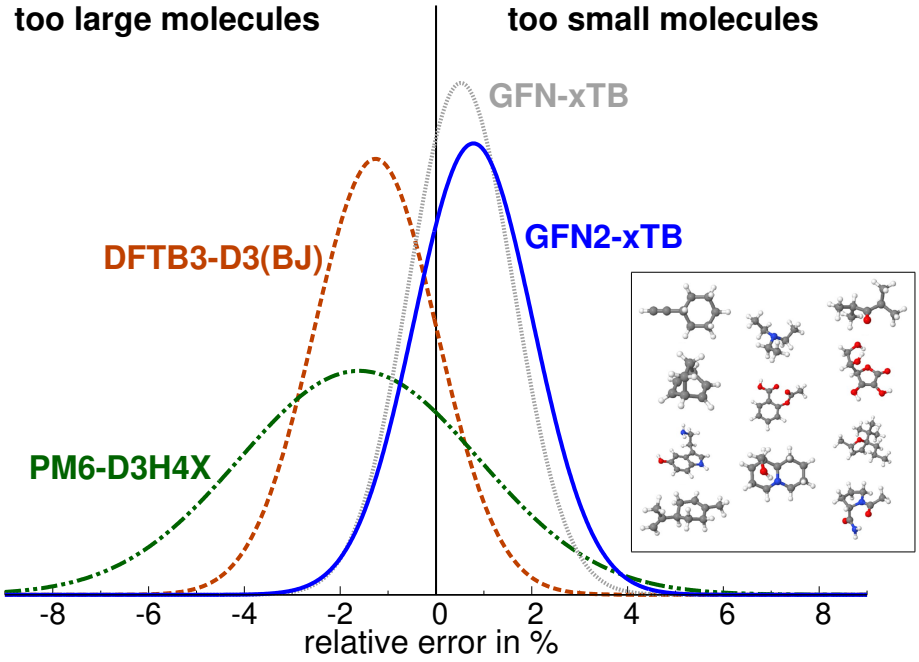


Figure 2: Normal distribution plots for the relative errors in the computed equilibrium rotational constants B_e for the ROT34 benchmark, with system 2 (isoamyl-acetate) excluded. The GFN-xTB (MRD=0.52%, SRD=1.10%), DFTB-D3(BJ) (MRD=−1.26%, SRD=1.28%), and PM6-D3H4X (MRD=−1.60%, SRD=2.50%) results are taken from Ref. 34. The values for GFN2-xTB are computed in this work (MRD=0.78%, SRD=1.24%).

In particular, it becomes apparent that GFN2-xTB reproduces the LB12 and HMGB11 bond lengths particularly well. This reflects the consistent element-specific parametrization in GFN2-xTB. It should, however, be noted that there exists one outlier in the LB12 set for GFN2-xTB (385 pm instead of 286 pm for the S_8^{2+} system), which is excluded in the statistical analysis. Presumably, this overestimated bond length is caused by the strong net charge of the system in combination with the higher order, but truncated multipole expansion (cf. GFN-xTB, which shows a bond distance underestimated by about 55 pm). This system is also difficult for many density functional approximations showing similar large deviations from the reference.[4] For the transition metal complexes (right hand side of Figure 3), the three semiempirical methods perform more similarly with GFN-xTB performing best and GFN2-xTB ranking second. For all sets considered here, it is observed that – compared to GFN-xTB – the magnitude of the MD is reduced for GFN2-xTB (see SI and Ref. 34).

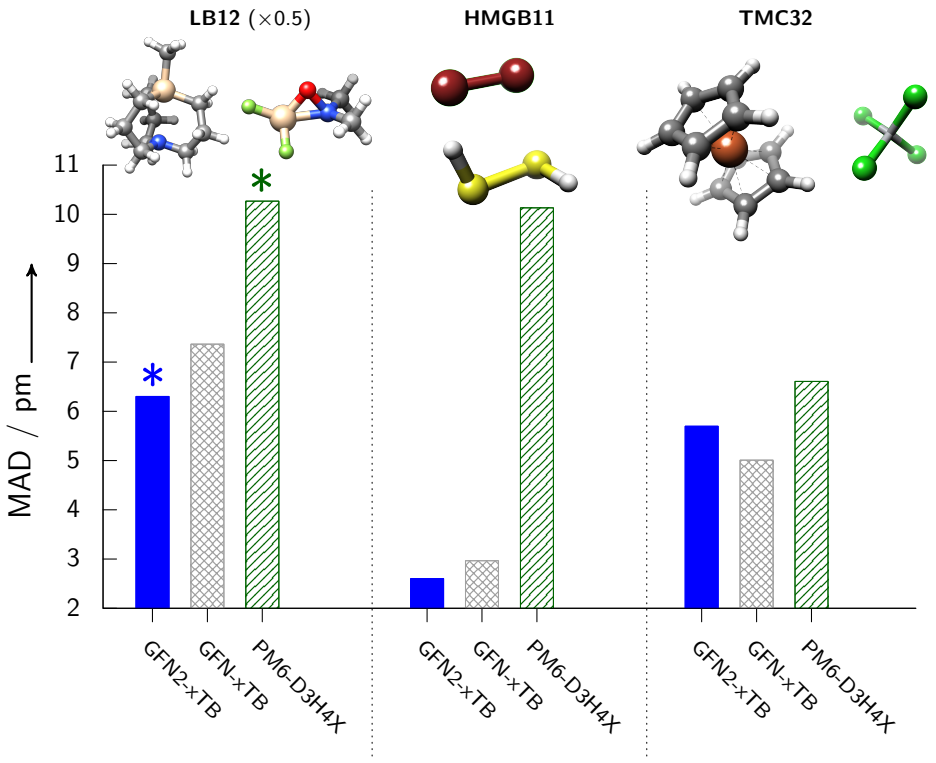


Figure 3: MADs in pm for bond lengths computed with GFN2-xTB, GFN-xTB, and PM6-D3H4X (see SI for detailed results). The LB12 and HMGB11 sets are taken from Ref. 4. For LB12, the MADs are scaled be factor of 0.5. The transition metal containing systems HAPPOD and KAMDOR in LB12 have been discarded for PM6-D3H4X. For GFN2-xTB, the S_8^{2+} system in LB12 is neglected.

In Table 3, the statistical data for structures with focus on non-covalent interactions are given. Here, the center-of-mass (CMA) distance deviation for fully optimized complexes from the S22[111] set are given. Furthermore, the relative deviations for the extrapolated CMA distances for the S66x8,[112] S22x5,[113] and X40x10[114] sets are given. These are determined by cubic spline interpolations of the respective interactions at the different distances, which has previously been done already for the S66x8 set.[4,34] Regarding the first two sets (fully optimized S22 and S66x8), GFN2-xTB performs much like GFN-xTB, but shows slightly stronger underestimation for the CMA distances of S66x8. Furthermore, the deviations appear to be a bit more systematic for both sets as indicated by the smaller standard relative deviation (SRD) for S66x8 and standard deviation (SD) for S22. All of the SQM methods considered perform comparably well with DFTB3-D3(BJ) showing the largest deviations for the S66x8 set.

Table 3: Comparison of structures of non-covalently bound systems. The CMA distances of the S22[111] complexes are obtained from a free optimization of the complex geometries. The deviations are given in pm. The CMA distances of the S66[112] S22x5,[113] and X40x10[114] complexes are derived from cubic spline interpolations of energies computed on the differently separated structures. Here, the relative errors in % are given.

	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)
		S22[111] (CMA distance in pm)		
MD:	−5	−5[a]	3[a]	−11[a]
MAD:	14	15[a]	14[a]	14[a]
SD:	16	18[a]	21[a]	16[a]
MAX:	32	50[a]	59[a]	45[a]
		S66x8[112] (CMA distance in %)		
MRD:	−0.82	−0.63[a]	0.58[a]	−2.87[a]
MURD:	1.99	1.96[a]	2.01[a]	3.15[a]
SRD:	2.58	2.68[a]	2.57[a]	2.84[a]
MAXR:	6.91	6.95[a]	6.74[a]	10.93[a]
		S22x5[113] (CMA distance in %)		
MRD:	−1.40	−0.76	0.03	−3.09
MURD:	2.76	2.49	1.97	3.51
SRD:	3.11	3.32	2.40	3.59
MAXR:	6.85	6.78	4.44	10.70
		X40x10[114] (CMA distance in %)		
MRD:	−0.17	−1.72	−1.39	−2.16
MURD:	2.59	3.47	5.25	3.67
SRD:	3.22	3.98	7.40	4.04
MAXR:	7.82	8.85	20.09	11.53

[a] Data taken from Ref. 34.

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, MAX=maximum absolute deviation.

The results obtained here are quite important also as an assessment for the balance of the non-covalent interaction terms. If repulsion, but in particular dispersion and the (anisotropic) ES terms are not described in a balanced way, in particular the free optimization of the S22 structures should yield larger deviations for the CMA distances. PM6-D3H4X has been parametrized to the S66x8 energies, and as expected, shows good performance for this set, though not better than GFN2-xTB. Its performance for S22x5 is remarkably good, for which it shows the smallest deviations of the tested methods, followed by GFN-xTB and GFN2-xTB. For X40x10, GFN2-xTB clearly

outperforms the other methods, and marks a significant improvement over GFN-xTB. Hence on average, GFN2-xTB is the best of all SQM methods considered here, which indicates that the Hamiltonian terms responsible for non-covalent interactions are well-behaved.

3.2 Non-covalent interaction energies

In order to investigate the performance for non-covalent interaction energies, we consider benchmark sets from the large GMTKN55 benchmark data set.[115] The MADs for the non-covalent interaction

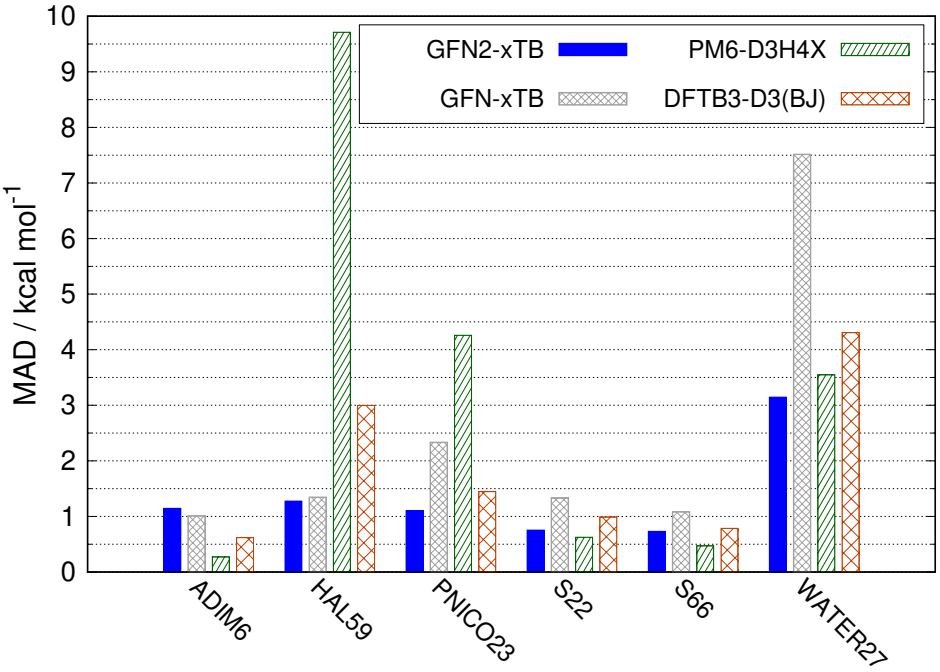


Figure 4: Mean absolute deviations (MADs) in kcal mol^{-1} for the non-covalent association energies of different benchmark sets from the GMTKN55 database[115].

energies are collected in Figures 4 and 6. Except for the weak interactions in the alkane dimer set ADIM6,[115] GFN2-xTB always ranks first or second among the considered methods. The benefit of including higher multipole moments is already observable for the often considered (also for fitting) S22[111,116] and S66[112] sets. Here, the GFN2-xTB method outperforms GFN-xTB and DFTB-D3(BJ) without specific Hamiltonian terms to treat hydrogen bonds. Only the PM6-D3H4X method shows a vanishingly lower MAD in both sets, which is expected as these sets had been used to adjust the D3 and H4 parameters.[38]

However, the advantages of including higher multipole electrostatic terms become particularly obvious when looking at the HAL59[114,115,117] and PNICO23[115,118] benchmarks. In these benchmarks, the anisotropic electron density of the bonded halogen and pnicogen atoms is very important. Though, the monopole-based tight-binding methods can perform well for either of the sets due their parametrization or specific halogen bond corrections (in GFN-xTB), a consistently good description is only observed for GFN2-xTB. This demonstrates the more physical behavior of GFN2-xTB, which shows a better performance without special correction terms. PM6-D3H4X performs bad for both benchmark sets, particularly for HAL59 in which seven systems yield an error > 10 kcal mol^{-1}. Noteworthy is the low MAD of GFN2-xTB for the WATER27 set, which is obtained without any additional H-bond specific correction term. In fact, the MAD is even lower than for well-performing generalized gradient density functional approximations (GGAs) like BLYP-D3(BJ) (MAD of 4.1 kcal mol^{-1})[115] or some hybrid functionals (e.g., PBE0-D3(BJ) with an MAD of 5.9 kcal mol^{-1})[115].

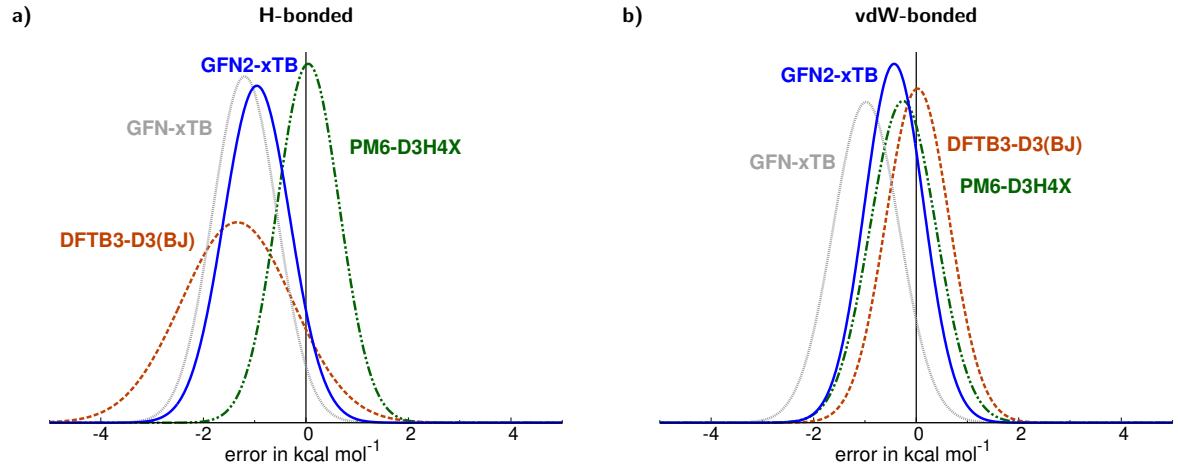


Figure 5: Normal distribution plots for the errors in the computed interaction energies the S66 benchmark.[112] Plot a) refers to the hydrogen bonded (first 23) systems, whereas plot b) refers to van-der-Waals-type bonded systems.

In Figure 5, the results for the S66 benchmark set are processed in more detail. This figure shows the error distribution plots subdivided for the hydrogen bonds (left hand side) and predominantly van-der-Waals-type interacting systems (right hand side). The MD for GFN2-xTB is significantly improved compared to the monopole-based GFN-xTB for the latter, which indicates the impor-

tance of AES for such systems. DFTB3-D3(BJ) also works well there – probably due to the pair-specifically parametrized repulsion potentials – but shows larger deviations for the hydrogen bonded systems. Different from that, GFN2-xTB produces consistently small errors ($|\mathrm{MD}| < 1$ kcal mol^{-1} for both subsets). PM6-D3H4X also performs extremely well for both subsets.

In the following, other benchmark sets for non-covalent interactions from the GMTKN55 database are considered (see Figure 6). These sets consist of more difficult systems (charged) or elements

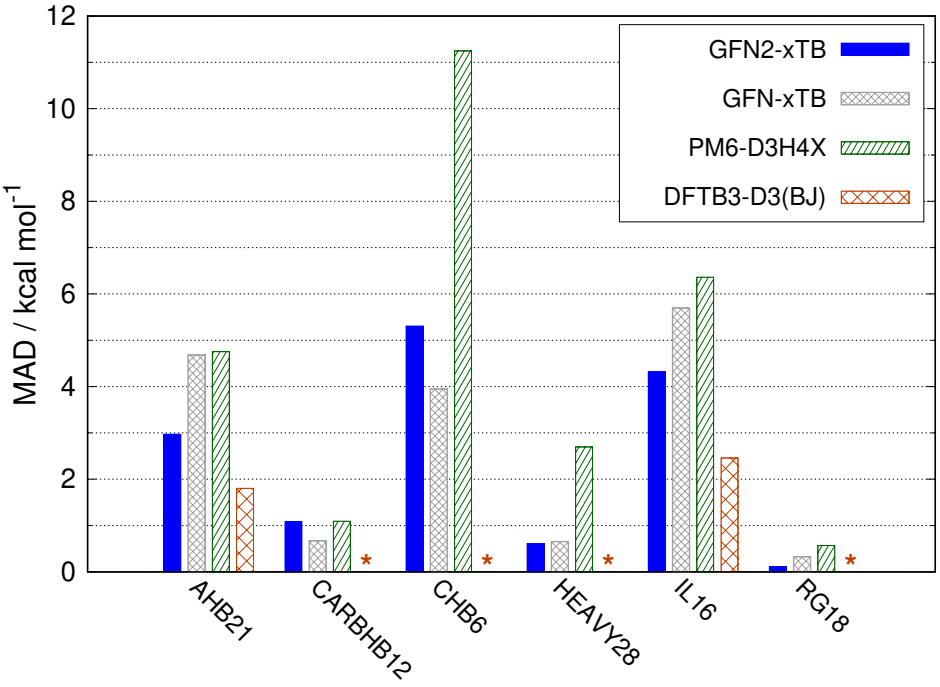


Figure 6: Mean absolute deviations (MADs) in kcal mol^{-1} for the non-covalent association energies of different benchmark sets from the GMTKN55 database.[115] These sets contain charged systems or heavy main group elements for which DFTB3 does not have parameters in the 3OB parametrization[58–60] (indicated by an asterisk).

other than from the first and second row. Though GFN-xTB performs slightly better for the cationic hydrogen bonded systems in CHB6 and the carbene-hydrogen bonded systems in CARBHB12, GFN2-xTB overall performs best and no outlier is observed, also for the heavy main group and noble gas elements. While DFTB3-D3(BJ) shows low MADs for the ionic liquids (IL16) and anionic hydrogen bonds (AHB21), the lack of parameters severely limits the applicability of the method. PM6-D3H4X on the other hand consistently shows bigger MADs than GFN-xTB, as well as GFN2-xTB.

Having demonstrated the good performance of GFN2-xTB for non-covalent interactions of small systems including different elements and interaction types, we next turn our attention to larger systems. For this purpose, the S30L set[119] is considered, which consists of 30 large non-covalently bound neutral and charged complexes. The results for GFN2-xTB are compared with PM6-D3H4X and DFTB3-D3(BJ) in Figure 7. GFN-xTB performs similar to DFTB3-D3(BJ) and is excluded from this figure for clarity (see SI for the GFN-xTB results). It is seen that GFN2-xTB closely

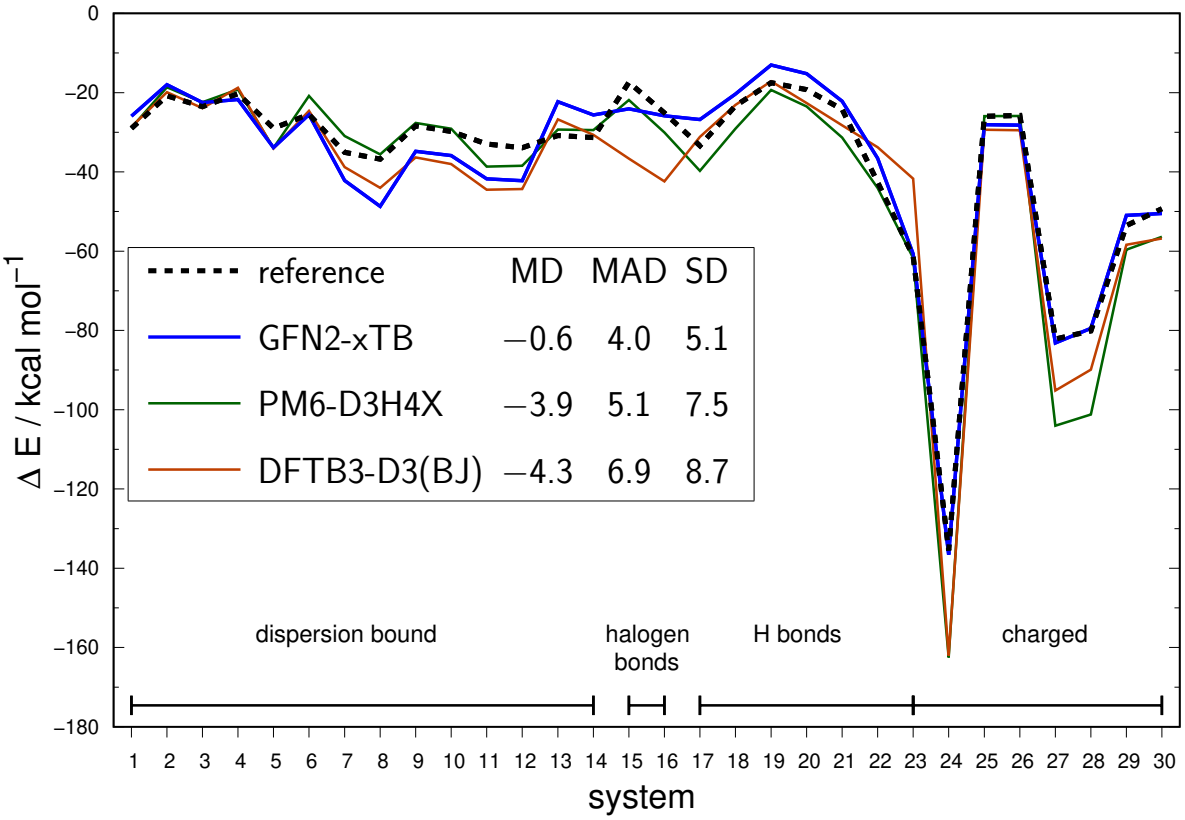


Figure 7: Association energies of the S30L[119] benchmark set computed with GFN2-xTB, PM6-D3H4X, and DFTB3-D3(BJ). The values and statistical data is given in kcal mol^{-1}.

resembles the reference values, i.e., domain-based pair natural orbital coupled cluster with singles, doubles, and perturbative triples – short DLPNO-CCSD(T)[120] – extrapolated to the complete basis set (CBS) limit. Overall, it shows the smallest (in magnitude) MD, MAD, and SD of all methods considered and is roughly on a par with some dispersion-corrected density functionals in a large quadruple-ζ basis set (e.g., B3LYP-D3(BJ)/def2-QZVP).[119] The nearly non-detectable

deviation for the charged systems is striking and even the largest association energy of about -135.5 kcal mol^{-1} (system 24) is reproduced to within <1% deviation, while the other semiempirical methods show errors of about 20%. This is a reassuring result and reflects the well-described electrostatic and polarization interactions in GFN2-xTB. Actually, the largest deviations are observed for the van-der-Waals-dominated complexes of conjugated π-systems (systems 7–12). GFN2-xTB overestimates the magnitude for the association energy of these complexes. This behavior partially results from non-additivity dispersion effects and the ATM term, which is also included in GFN2-xTB (see Section 2.2.6), only partially compensates for this. This has already been observed and discussed for DFT-D3 methods (see the comment in reference 93 of Ref. 75), and hence, does not represent a weakness specific to the GFN2-xTB Hamiltonian.

3.3 Conformational energies

For the description of conformational energies, the balance between covalent and (intramolecular) non-covalent forces is very important. The proper energy ranking of conformers is essential for SQM methods, which can efficiently be used to sample the conformational space of chemically important, moderately sized systems (<100-200 atoms). As mentioned before, determining the thermostatistically populated conformer-rotamer-ensemble for the calculation of spin-coupled nuclear-magnetic resonance spectra has been a driving force for the development of the GFN2-xTB method, and the AES term in particular. In Figure 8, MAD bar plots for the conformational energies of different benchmark sets from the GMTKN55[115] are given (see SI for detailed values). It can be seen that the GFN2-xTB method ranks best in five out of eight sets, namely for the ACONF[121], Amino20x4,[122], ICONF,[115] PCONF21,[115,123,124] and SCONF.[115,125] Particularly positive is the considerable improvement over GFN-xTB for more polar and hydrogen-bonded systems in Amino20x4, PCONF21, and SCONF. This suggests a generally better performance for sugars and polypeptide systems. Furthermore, the GFN2-xTB MAD of 1.6 kcal mol^{-1} for the ICONF stands out, as DFTB3-D3(BJ), which ranks second, has an MAD that is larger by almost 1 kcal mol^{-1}. Also for the BUT14DIOL[115,126] and MCONF[115,127] sets, the GFN2-xTB method performs quite similar to the other SQM methods. The only outlier with an MAD of 2.9 kcal mol^{-1} is the UPU23 set,[115,128] which is about twice as large as the MADs of either GFN-xTB or DFTB3-

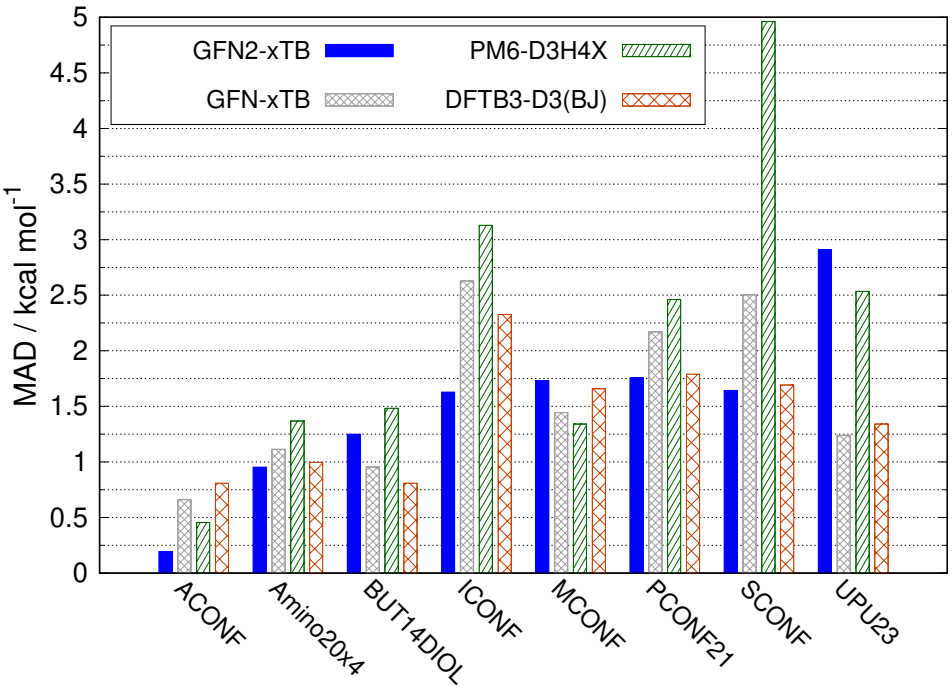


Figure 8: Mean absolute deviations (MADs) in kcal mol^{-1} for conformational energy benchmark sets. The structures and reference data are taken from the GMTKN55 data base.[115] Detailed values are listed in the SI.

D3(BJ). The latter both perform particularly well for this set, whereas PM6-D3H4X is only slightly (< 0.5 kcal mol^{-1}) better than GFN2-xTB. All in all, GFN2-xTB shows the best performance for the conformer sets considered here.

Recently, a set of glucose and maltose conformers has been compiled.[129] These sugar conformers are particularly challenging due to the many differently hydrogen bonded conformers. In Figure 9, conformational energies of both sets are plotted against the high level reference data (CBS extrapolated DLPNO-CCSD(T) values).[129] First of all, it is observed that, on average, all SQM methods underestimate the conformational energies in particular of high-lying conformers. In agreement with the aforementioned results for the SCONF set, GFN2-xTB outperforms the other methods in all statistical measures considered. Remarkable are the small SD values, which only DFTB3-D3(BJ) is nearly on par with. The results for these sugar conformers are encouraging in that GFN2-xTB will be able to provide more reliable conformer-rotamer ensembles than GFN-xTB even for polar and hydrogen bonded systems.

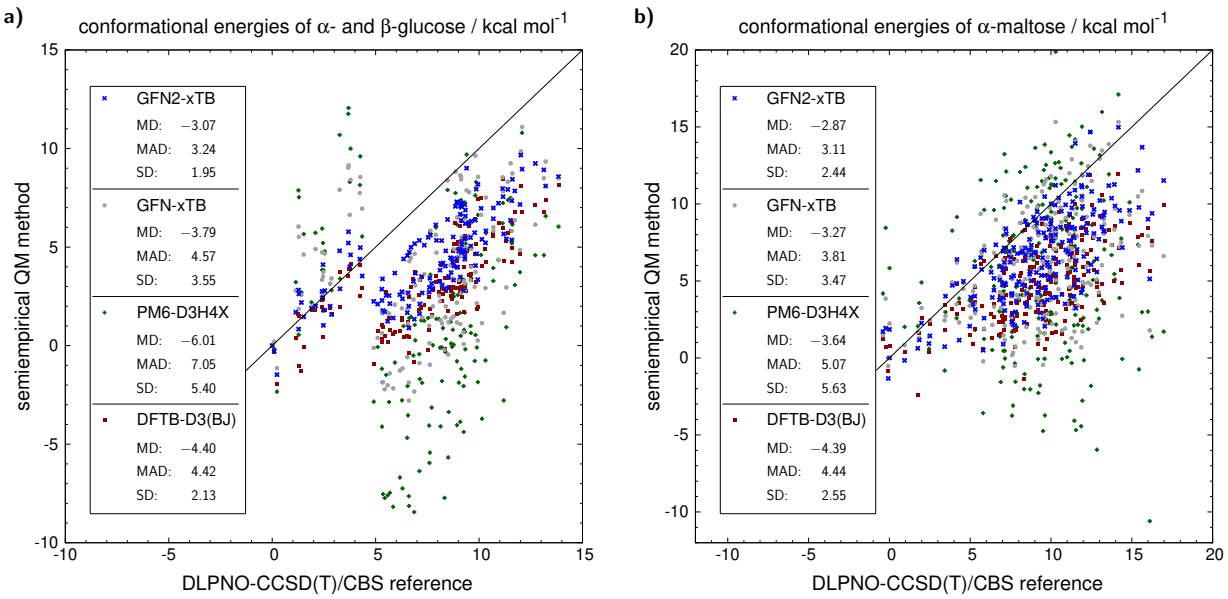


Figure 9: Correlation plots for **a)** the conformational energies of 80 α-glucose and 76 β-glucose conformers and **b)** the conformational energies of 205 α-maltose conformers. The energies are given in kcal mol^{-1} and are computed with different SQM methods. The structures and reference conformational energies are taken from Ref. 129.

3.4 Rotational and vibrational free energy computations

Due to computational efficiency in the calculation of gradients and numerical Hessians, a major application of GFN2-xTB will likely be the computation of free energy corrections in thermochemical studies. This has already been pointed out for GFN-xTB[34] and we will crosscheck the performance of GFN2-xTB against GFN-xTB and PBEh-3c (frequencies scaled by 0.95).[4] The latter is a a hybrid density functional composite method, which is computationally considerably more demanding than the semiempirical tight binding methods (roughly two orders of magnitude). As reference method, another composite method called B97-3c,[6] is used, which employs a GGA functional in a triple-ζ basis set. As a "real-life" measure, differences for reactions in the summed rotational and vibrational free energies termed ΔG_{RRHO} at $T = 298.15$ K are considered. While the rotational part is a measure for the differences in optimized structures, the vibrational contribution contains information about the PES curvature around the minima. We use the systems from the following sets, which are part of the GMTKN55 benchmark set database:[115] AL2X6, DARC, HEAVYSB11, ISOL24, TAUT15, ALK8, G2RC, and ISO34. Furthermore the recently proposed

MOR41 set for closed-shell metal organic reactions is included.[130] The detailed results are listed in the SI. It should be noted that translational free energy contribution, which is independent of the electronic structure method used, is purposely excluded, since the number of reactants differs from the number of products for many of the reactions considered. This way, the magnitudes for ΔG_{RRHO} are more similar across the different sets. Furthermore for all methods, the harmonic oscillator/free rotor interpolation from Ref 131 is applied for harmonic frequencies with magnitudes smaller than 50 cm^{-1}. The results are plotted in Figure 10. The idea about this comparison is that

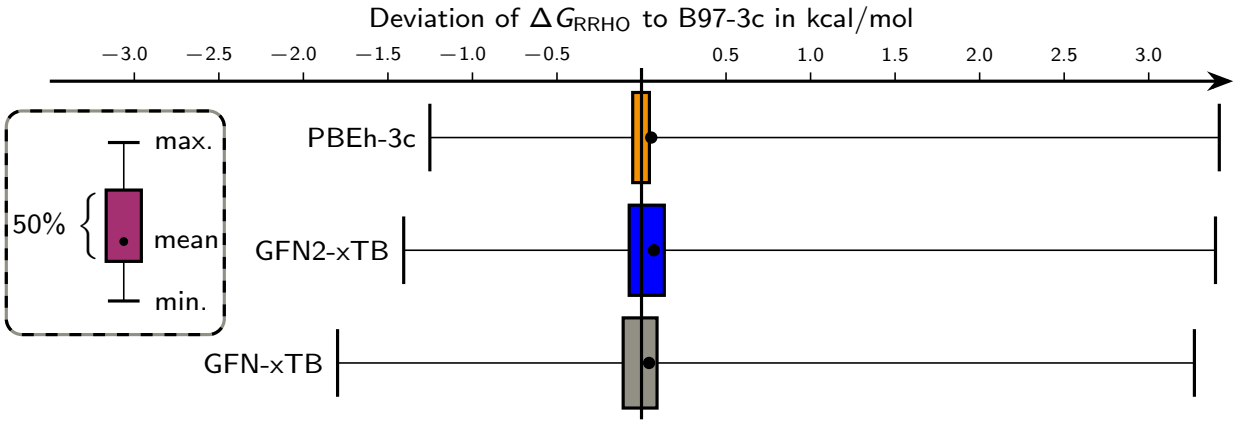


Figure 10: Deviation plot for the rotational and vibrational reaction free energy computed with the tested methods compared to B97-3c.[6] The MD as well as maximum and minimum deviation is shown for each method. The boxes show the range in which the smallest half of the errors is found. For PBEh-3c, the frequencies have been scaled by 0.95.

both composite "3c" methods represent comparably accurate methods and it will depend on the system, which one performs better. As expected, GFN-xTB shows slighty larger errors and only has a somewhat larger spread of errors. Similar is true also for the GFN2-xTB method. In fact the 50% cases with smallest deviations for both tight binding methods are found within a range that is about twice as large compared to PBEh-3c. However in absolute numbers, this error is on a 0.1 kcal mol^{-1} scale and hence practically irrelevant, given that some deviations to B97-3c are much larger for the tight binding methods as well as for PBEh-3c. Therefore, GFN2-xTB, just like GFN-xTB, should be a reasonable method of choice to routinely compute harmonic frequencies for subsequent thermostatistical treatments (including transition metals complexes). This is particularly important, since such calculations with *ab initio* and DFT methods quickly become the

computational bottleneck in typical workflows.

3.5 Other properties

In the previous section, we have assessed GFN2-xTB for well-established benchmark sets that have also been used to study DFT approximations. Though not training sets, these benchmark sets predominantly coincided with the target properties of the GFN2-xTB method. In this section, GFN2-xTB is tested for some off-target properties. As in Ref. 34, we investigate the performance of GFN2-xTB for covalently bonded diatomics in Figure 11. Similar to GFN-xTB, we observe a

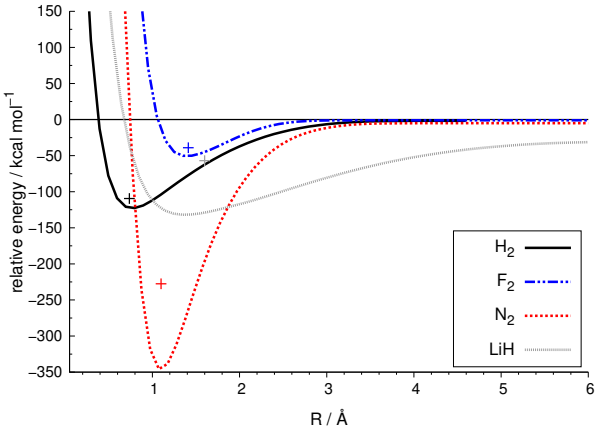


Figure 11: Potential energy curves computed with GFN2-xTB for the dissociation of H_2, F_2, N_2, and LiH. The electronic temperature treatment (T_{el} = 300 K) allows the homolytic dissociation without a multi-reference treatment. The points mark the minimum energy positions from high-level calculations.[34,132,133] The energies are given relative to the free atoms (S=3/2 for nitrogen, S=1/2 for the others).

systematic overestimation of typical bond dissociation energies, while the minimum positions are reproduced rather well (high level reference data are marked with crosses). This agrees well with the observations made in Section 3.1, i.e., that GFN2-xTB on average is on a par with GFN-xTB in the description of covalently bonded molecular structures. At the same time, GFN2-xTB shares the property of overestimating covalent bond energies. However, a slight difference compared to GFN-xTB (cf. Ref. 34) is observable for the non-polar, single-bonded diatomics (H_2 and F_2). Here the overestimation is slightly less pronounced than with GFN-xTB, while the triple bond energy of N_2 and the polar LiH bond energy show about the same magnitude as GFN-xTB. It is to be

determined in the future, how this will affect, e.g., the simulation of mass spectra or reaction enthalpies (within the correction scheme applied in Ref. 45).

Next, we turn our attention to some of the kinetics-oriented benchmark sets of the GMTKN55[115] database. In Figure 12, the MADs for different barrier heights are presented. For three sets,

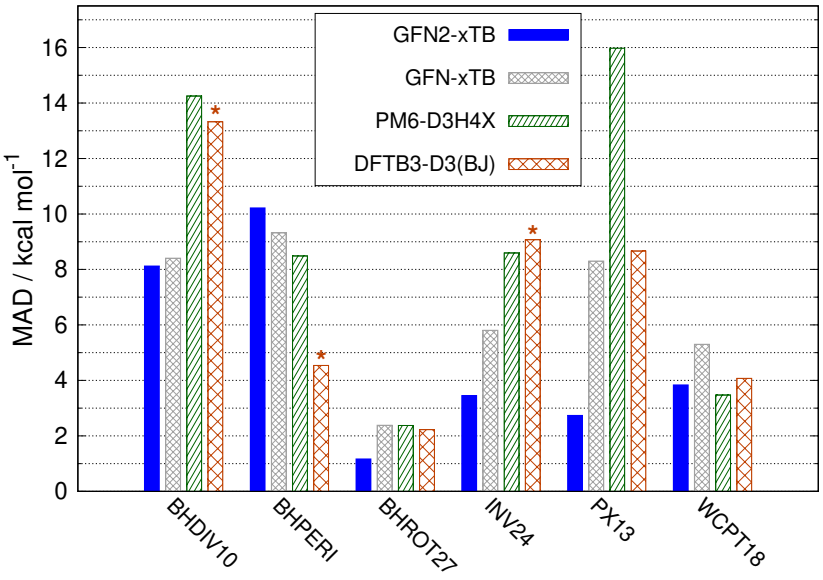


Figure 12: Mean absolute deviations (MADs) for reaction barriers (in kcal mol^{-1}) computed with different semiempirical methods. Due to missing parameters for silicon, DFTB3 calculations are not possible for two systems in BHDIV10[115] and one system in the BHPERI[134] set. Furthermore, one extreme outlier (PCl_3) in the INV24[135] was found. In both cases, these systems are removed from the statistical analysis for DFTB3-D3(BJ).

systems are neglected in the statistical analysis for DFTB3-D3(BJ), i.e., the Si-containing systems in BHDIV10 and BHPERI[134] (missing parameters) and also a severe outlier in the INV24[135] set. Here, the planar PCl_3 transition state structure yielded a preposterously large repulsion energy ($> 10^4$ Hartree).

Along all sets considered, the GFN2-xTB performs best. It shows the lowest MAD for the BHDIV10,[115] BHROT27,[115] INV24,[135], and PX13,[115,136] sets and the second lowest for the WCPT18.[115,137] Only for the barrier heights of pericyclic reactions (BHPERI),[134] GFN2-xTB is outperformed by the other SQM methods, though GFN-xTB and PM6-D3H4X are only better by about 1–1.5 kcal mol^{-1}. The performance of GFN2-xTB for all sets in Fig. 12 is remarkable, given that none of these (or similar systems) were used fitting. In particular for the proton transfer sets PX13

and WCPT18, as well as the single bond rotation set BHROT27, GFN2-xTB shows MADs, which are comparable to or better than those of the hybrid functional PBE0.[115] This – along with the low MAD for the Amino20x4,[122], PCONF21,[115,123,124] and WATER27[138] sets (see above) – indicates that GFN2-xTB may be well-suited to study biomolecular systems in aqueous solution.

As a last test, we investigate the ability to reproduce the permanent electric dipole moments of small molecules which is relevant for obtaining good long-range non-covalent interactions. Such a set with purely theoretical reference values has recently been compiled by Halt and Head-Gordon.[139]

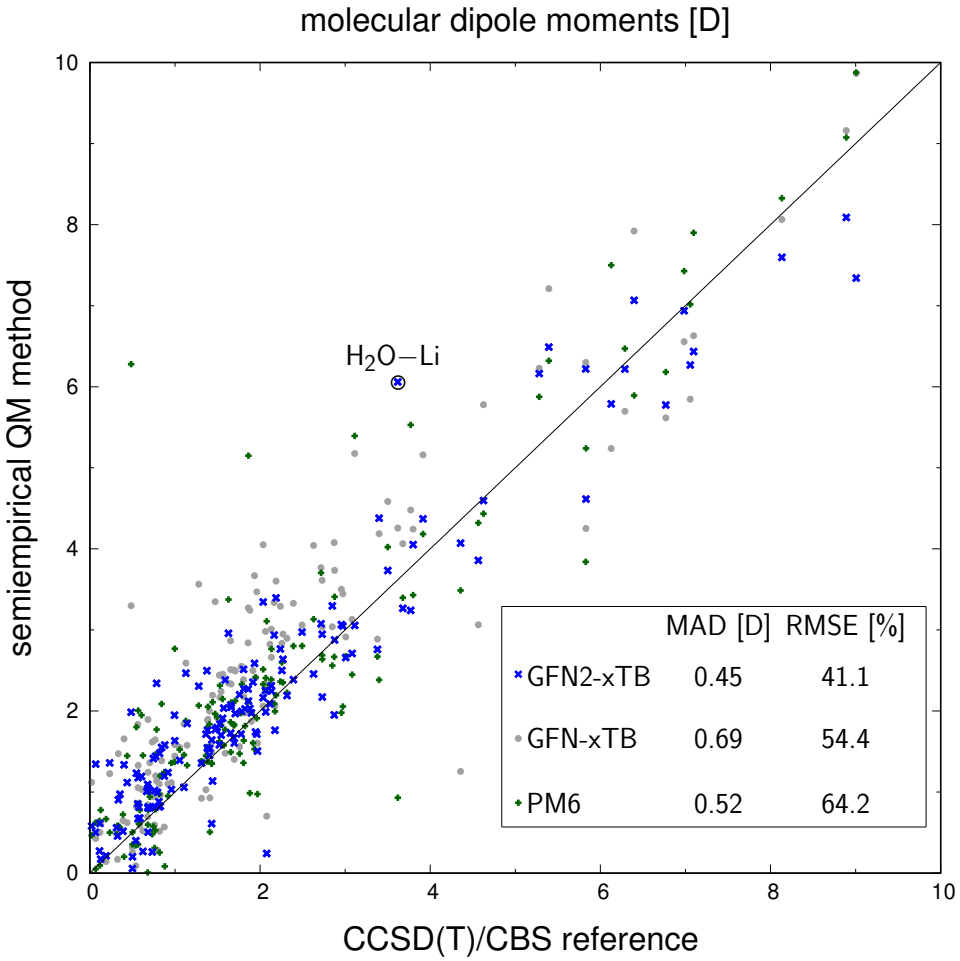


Figure 13: Permanent molecular dipole moments computed for open, as well as closed shell systems with GFN2-xTB, GFN-xTB, and PM6-D3H4X. The benchmark set (structures and reference values) are taken from from Ref. 139. Detailed values are listed in the SI. The root mean square error (in %) is regularized with a value of 1 Debye.

In Figure 13, the correlation plot for the computed molecular dipole moments is shown. Due

to missing parameters for many elements, DFTB3 is excluded here. While dipole moments have been used in the parametrization procedure of PM6,[35] none were used in the fit of GFN-xTB and GFN2-xTB. Nevertheless, PM6 does not show significantly better agreement with the high level reference. In fact, GFN2-xTB shows a lower MAD, as well as a lower regularized root mean square error (RMSE). As in Ref. 139, a value of 1 Debye is used for regularization. The RMSE is close to the one for MP2 (37.5%) as presented in the original work.[139] This is an encouraging finding, given that the set covers 14 different elements and almost half of the systems have an open-shell ground state. Our findings provide further support for the improved physics in the GFN2-xTB Hamiltonian, as well as for our parametrization strategy.

Finally, it should be mentioned that in comparison to GFN-xTB, no increase in the computational time is expected for typical organic and biomolecular systems. As such, a single-point energy calculation for crambin (641 atoms) takes 30 s with GFN2-xTB, but instead the same calculation requires 50 s with GFN-xTB (single-core run on a laptop with a 1.6 GHz Intel Core i5 CPU and 8 GB RAM). Though the additional integral evaluation (dipole and quadrupole one-electron integrals) in GFN2-xTB is more elaborate, the rate-determining step in both methods is the diagonalization of the tight-binding Hamiltonian matrix. Due to the extra s-function for hydrogen,[34] the matrix dimension in GFN-xTB is significantly larger for typical organic and biomolecular systems (by a factor of about 1.5), and hence the computation time is even reduced for GFN2-xTB compared to GFN-xTB.

The method is implemented in the standalone `xtb` code, which can be requested from the authors.[140]

4 Conclusions

We developed a broadly applicable semiempirical quantum mechanical method, termed GFN2-xTB, which represents the first tight-binding method to include electrostatic and exchange-correlation Hamiltonian terms beyond the monopole approximation. The method is free from any hydrogen or halogen bond specific corrections, which are a standard add-on in other contemporary semiempirical schemes. Furthermore, the self-consistent D4 dispersion model is an inherent part of the GFN2-xTB method and allows to efficiently incorporate electronic structure effects on the two-body dispersion

energy. The GFN2-xTB method relies strictly on element-specific and global parameters and is parametrized for all elements up to radon ($Z = 86$). Like for its predecessor, GFN-xTB, the parameters were fitted to yield reasonable structures, vibrational frequencies and non-covalent interactions for molecules across the periodic table. The main focus of this method are organic, organometallic, and biochemical systems on the order of a few thousand atoms. In particular, the greatly improved non-covalent interactions will likely trigger structural searches and studies of conformational and protein-ligand studies in the near future.

Apart from these, the improved electrostatics and more consistent parametrization procedure has provided a method, which better reproduces the electronic density compared to other semiempirical methods. The method may thus qualify to be used in docking procedures by providing reasonable electrostatic potentials.

However, as a difficult-to-quantify drawback compared to GFN-xTB we finally note sometimes less robust SCF convergence in particular for metallic systems or polar inorganic clusters probably caused by the short-range part of the AES potential. Further work to improve the GFN family of methods for this field of application is in progress.

Acknowledgments

This work was supported by the DFG in the framework of the "Gottfried Wilhelm Leibniz Prize" awarded to S.G. The authors would like to thank Prof. Dr. Thomas Bredow, Dr. Jan Gerit Brandenburg, and Dr. Andreas Hansen for helpful discussions during the development of the GFN2-xTB method. Furthermore, the authors are grateful to Eike Caldeweyher for providing routines, which are used in the self-consistent D4 treatment.

Associated Content

The derivations for the newly developed AES and AXC, as well as the self-consistent D4 dispersion energy expressions, along with the respective TB-Fock matrix contributions, and nuclear gradients are shown in the supporting information. Element-specific parameters and the detailed results

obtained on the considered test sets are given here as well.

References

(1) Grimme, S.; Schreiner, P. R. Computational Chemistry: The Fate of Current Methods and Future Challenges. *Angew. Chem. Int. Ed.* **2017**, *57*, 4170–4176.

(2) Houk, K. N.; Liu, F. Holy Grails for Computational Organic Chemistry and Biochemistry. *Acc. Chem. Res.* **2017**, *50*, 539–543.

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

(4) 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.

(5) 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.

(6) Brandenburg, J. G.; Bannwarth, C.; Hansen, A.; Grimme, S. B97-3c: A revised low-cost variant of the B97-D density functional method. *J. Chem. Phys.* **2018**, *148*, 064104.

(7) Otero-de-la Roza, A.; DiLabio, G. A. Transferable Atom-Centered Potentials for the Correction of Basis Set Incompleteness Errors in Density-Functional Theory. *J. Chem. Theory Comput.* **2017**, *13*, 3505–3524.

(8) Witte, J.; Neaton, J. B.; Head-Gordon, M. Effective empirical corrections for basis set superposition error in the def2-SVPD basis: gCP and DFT-C. *J. Chem. Phys.* **2017**, *146*, 234105.

(9) Hostaš, J.; Řezáč, J. Accurate DFT-D3 Calculations in a Small Basis Set. *J. Chem. Theory Comput.* **2017**, *13*, 3575–3585.

(10) Kulik, H. J.; Seelam, N.; Mar, B. D.; Martinez, T. J. Adapting DFT+U for the Chemically Motivated Correction of Minimal Basis Set Incompleteness. *J. Phys. Chem. A* **2016**, *120*, 5939–5949.

(11) Salsbury, F. R. Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. *Curr. Opin. Pharmacol.* **2010**, *10*, 738–744.

(12) Kussmann, J.; Ochsenfeld, C. Pre-selective screening for matrix elements in linear-scaling exact exchange calculations. *J. Chem. Phys.* **2013**, *138*, 134114.

(13) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 2619–2628.

(14) Luehr, N.; Ufimtsev, I. S.; Martinez, T. J. Dynamic Precision for Electron Repulsion Integral Evaluation on Graphical Processing Units (GPUs). *J. Chem. Theory Comput.* **2011**, *7*, 949–954.

(15) Schütt, O.; Messmer, P.; Hutter, J.; VandeVondele, J. *Electronic Structure Calculations on Graphics Processing Units*; John Wiley & Sons, Ltd, 2016; pp 173–190.

(16) Yasuda, K.; Maruoka, H. Efficient calculation of two-electron integrals for high angular basis functions. *Int. J. Quantum Chem.* **2014**, *114*, 543–552.

(17) Asadchev, A.; Gordon, M. S. New Multithreaded Hybrid CPU/GPU Approach to Hartree–Fock. *J. Chem. Theory Comput.* **2012**, *8*, 4166–4176.

(18) Yasuda, K. Accelerating Density Functional Calculations with Graphics Processing Unit. *J. Chem. Theory Comput.* **2008**, *4*, 1230–1236.

(19) Maia, J. D. C.; Urquiza Carvalho, G. A.; Mangueira, C. P.; Santana, S. R.; Cabral, L. A. F.; Rocha, G. B. GPU Linear Algebra Libraries and GPGPU Programming for Accelerating MOPAC Semiempirical Quantum Chemistry Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3072–3081.

(20) Wu, X.; Koslowski, A.; Thiel, W. Semiempirical Quantum Chemical Calculations Accelerated on a Hybrid Multicore CPU–GPU Computing Platform. *J. Chem. Theory Comput.* **2012**, *8*, 2272–2281.

(21) Kussmann, J.; Ochsenfeld, C. Hybrid CPU/GPU Integral Engine for Strong-Scaling Ab Initio Methods. *J. Chem. Theory Comput.* **2017**, *13*, 3153–3159.

(22) Kalinowski, J.; Wennmohs, F.; Neese, F. Arbitrary Angular Momentum Electron Repulsion Integrals with Graphical Processing Units: Application to the Resolution of Identity Hartree–Fock Method. *J. Chem. Theory Comput.* **2017**, *13*, 3160–3170.

(23) van Schoot, H.; Visscher, L. *Electronic Structure Calculations on Graphics Processing Units*; John Wiley & Sons, Ltd, 2016; pp 101–114.

(24) Ufimtsev, I. S.; Martínez, T. J. Graphical Processing Units for Quantum Chemistry. *Comput. Sci. Eng.* **2008**, *10*, 26–34.

(25) Jakowski, J.; Irle, S.; Morokuma, K. In *GPU Computing Gems Emerald Edition*; Hwu, W.-m. W., Ed.; Applications of GPU Computing Series; Morgan Kaufmann: Boston, 2011; pp 59 – 73.

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

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

(28) 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.

(29) 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.

(30) 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. Aid. Mol. Des.* **1995**, *9*, 87–110.

(31) 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.

(32) Cramer, C. J.; Truhlar, D. G. AM1-SM2 and PM3-SM3 parameterized SCF solvation models for free energies in aqueous solution. *J. Comput. Aid. Mol. Des.* **1992**, *6*, 629–666.

(33) 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.

(34) Grimme, S.; Bannwarth, C.; Shushkov, P. 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 ($Z = 1 - 86$). *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.

(35) 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.

(36) 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.

(37) 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.

(38) Řezáč, J.; Hobza, P. Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *J. Chem. Theory Comput.* **2012**, *8*, 141–151.

(39) S. Brahmkshatriya, P.; 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.-Aid. Drug.* **2013**, *9*, 118–129.

(40) Saito, T.; Kitagawa, Y.; Takano, Y. Reparameterization of PM6 Applied to Organic Diradical Molecules. *J. Phys. Chem. A* **2016**, *120*, 8750–8760.

(41) Weber, W.; Thiel, W. Orthogonalization corrections for semiempirical methods. *Theor. Chem. Acc.* **2000**, *103*, 495–506.

(42) 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.

(43) 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.

(44) 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.

(45) Kromann, J.; Welford, A.; Christensen, A. S.; Jensen, J. Random Versus Systematic Errors in Reaction Enthalpies Computed Using Semi-empirical and Minimal Basis Set Methods. **2018**,

(46) Bikadi, Z.; Hazai, E. Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. *J. Cheminform.* **2009**, *1*, 15.

(47) Stewart, J. J. P. Application of the PM6 method to modeling proteins. *J. Mol. Model.* **2009**, *15*, 765–805.

(48) Miriyala, V. M.; Řezáč, J. Testing Semiempirical Quantum Mechanical Methods on a Data Set of Interaction Energies Mapping Repulsive Contacts in Organic Molecules. *J. Phys. Chem. A* **2018**, *122*, 2801–2808.

(49) Kazaryan, A.; Lan, Z.; Schäfer, L. V.; Thiel, W.; Filatov, M. Surface Hopping Excited-State Dynamics Study of the Photoisomerization of a Light-Driven Fluorene Molecular Rotary Motor. *J. Chem. Theory Comput.* **2011**, *7*, 2189–2199.

(50) Spörkel, L.; Cui, G.; Thiel, W. Photodynamics of Schiff Base Salicylideneaniline: Trajectory Surface-Hopping Simulations. *J. Phys. Chem. A* **2013**, *117*, 4574–4583.

(51) Spörkel, L.; Cui, G.; Koslowski, A.; Thiel, W. Nonequilibrium H/D Isotope Effects from Trajectory-Based Nonadiabatic Dynamics. *J. Phys. Chem. A* **2014**, *118*, 152–157.

(52) Spörkel, L.; Jankowska, J.; Thiel, W. Photoswitching of Salicylidene Methylamine: A Theoretical Photodynamics Study. *J. Phys. Chem. B* **2015**, *119*, 2702–2710.

(53) Dokukina, I.; Marian, C. M.; Weingart, O. New Perspectives on an Old Issue: A Comparative MS-CASPT2 and OM2-MRCI Study of Polyenes and Protonated Schiff Bases. *Photochem. Photobiol.* **2017**, *93*, 1345–1355.

(54) 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* **1998**, *58*, 7260–7268.

(55) 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.

(56) 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.

(57) 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* **2001**, *63*, 085108.

(58) Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* **2013**, *9*, 338–354.

(59) 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.

(60) Kubillus, M.; Kubař, T.; Gaus, M.; Řezáč, J.; Elstner, M. Parameterization of the DFTB3 Method for Br, Ca, Cl, F, I, K, and Na in Organic and Biological Systems. *J. Chem. Theory Comput.* **2015**, *11*, 332–342.

(61) 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.

(62) Vujović, M.; Huynh, M.; Steiner, S.; Garcia-Fernandez, P.; Elstner, M.; Cui, Q.; Gruden, M. Exploring the applicability of density functional tight binding to transition metal ions. Parameterization for nickel with the spin-polarized DFTB3 model. *J. Comp. Chem.* **2018**,

(63) Bursch, M.; Hansen, A.; Grimme, S. Fast and Reasonable Geometry Optimization of Lanthanoid Complexes with an Extended Tight Binding Quantum Chemical Method. *Inorg. Chem.* **2017**, *56*, 12485–12491.

(64) Struch, N.; Bannwarth, C.; Ronson, T. K.; Lorenz, Y.; Mienert, B.; Wagner, N.; Engeser, M.; Bill, E.; Puttreddy, R.; Rissanen, K.; Beck, J.; Grimme, S.; Nitschke, J. R.; Lützen, A. An Octanuclear Metallosupramolecular Cage Designed To Exhibit Spin-Crossover Behavior. *Angew. Chem. Int. Ed.* **2017**, *56*, 4930–4935.

(65) Seibert, J.; Bannwarth, C.; Grimme, S. Biomolecular Structure Information from High-Speed Quantum Mechanical Electronic Spectra Calculation. *J. Am. Chem. Soc.* **2017**, *139*, 11682–11685.

(66) Pracht, P.; Bauer, C. A.; Grimme, S. Automated and efficient quantum chemical determination and energetic ranking of molecular protonation sites. *J. Comput. Chem.* **2017**, *38*, 2618–2631.

(67) Grimme, S.; Bannwarth, C.; Dohm, S.; Hansen, A.; Pisarek, J.; Pracht, P.; Seibert, J.; Neese, F. Fully Automated Quantum-Chemistry-Based Computation of Spin–Spin-Coupled Nuclear Magnetic Resonance Spectra. *Angew. Chem. Int. Ed.* **2017**, *56*, 14763–14769.

(68) Asgeirsson, V.; Bauer, C. A.; Grimme, S. Quantum chemical calculation of electron ionization mass spectra for general organic and inorganic molecules. *Chem. Sci.* **2017**, *8*, 4879–4895.

(69) Grimme, S.; Bannwarth, C.; Caldeweyher, E.; Pisarek, J.; Hansen, A. A general intermolecular force field based on tight-binding quantum chemical calculations. *J. Chem. Phys.* **2017**, *147*, 161708.

(70) 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.

(71) Brandenburg, J. G.; Grimme, S. Dispersion Corrected Hartree-Fock and Density Functional Theory for Organic Crystal Structure Prediction. *Top. Curr. Chem.* **2014**, *345*, 1–23.

(72) Humeniuk, A.; Mitrić, R. Long-range correction for tight-binding TD-DFT. *J. Chem. Phys.* **2015**, *143*, 134120.

(73) Lutsker, V.; Aradi, B.; Niehaus, T. A. Implementation and benchmark of a long-range corrected functional in the density functional based tight-binding method. *J. Chem. Phys.* **2015**, *143*, 184107.

(74) Kranz, J. J.; Elstner, M.; Aradi, B.; Frauenheim, T.; Lutsker, V.; Garcia, A. D.; Niehaus, T. A. Time-Dependent Extension of the Long-Range Corrected Density Functional Based Tight-Binding Method. *J. Chem. Theory Comput.* **2017**, *13*, 1737–1747.

(75) Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **2016**, *116*, 5105–5154.

(76) Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals density functional: the simpler the better. *J. Chem. Phys.* **2010**, *133*, 244103.

(77) Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Bannwarth, C.; Grimme, S. *manuscript in preparation*

(78) Caldeweyher, E.; Bannwarth, C.; Grimme, S. Extension of the D3 dispersion coefficient model. *J. Chem. Phys.* **2017**, *147*, 034112.

(79) Nishimoto, K.; Mataga, N. Electronic Structure and Spectra of Some Nitrogen Heterocycles. *Z. Phys. Chem.* **1957**, *12*, 335–338.

(80) Ohno, K. Some Remarks on the Pariser-Parr-Pople Method. *Theor. Chim. Act.* **1964**, *2*, 219.

(81) Klopman, G. A Semiempirical Treatment of Molecular Structures. II. Molecular Terms and Application to Diatomic Molecules. *J. Am. Chem. Soc.* **1964**, *86*, 4450.

(82) Grimme, S. A General Quantum Mechanically Derived Force Field (QMDFF) for Molecules and Condensed Phase Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 4497–4514.

(83) Mermin, N. D. Thermal Properties of the Inhomogeneous Electron Gas. *Phys. Rev. A* **1965**, *137*, 1441–1443.

(84) Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23*, 69–89.

(85) Hall, G. G. The Molecular Orbital Theory of Chemical Valency. VIII. A Method of Calculating Ionization Potentials. **1951**, *205*, 541–552.

(86) 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.

(87) Pyykkö, P.; Atsumi, M. Molecular Single-Bond Covalent Radii for Elements 1–118. *Chem. Eur. J. 15*, 186–197.

(88) Mantina, M.; Valero, R.; Cramer, C. J.; Truhlar, D. G. In *CRC Handbook of Chemistry and Physics, 91nd edition*; Haynes, W. M., Ed.; CRC Press: Boca Raton, FL, 2010; pp 9–49–9–50.

(89) Köhler, C.; Seifert, G.; Frauenheim, T. Density functional based calculations for Fe_n ($n \leq 32$). *Chem. Phys.* **2005**, *309*, 23 – 31.

(90) Grimme, S. A simplified Tamm-Dancoff density functional approach for the electronic excitation spectra of very large molecules. *J. Chem. Phys.* **2013**, *138*, 244104.

(91) Zoltán, B.; Bálint, A. Possible improvements to the self-consistent-charges density-functional tight-binding method within the second order. *Phys. Status Solidi B* **2012**, *249*, 259–269.

(92) Köster, A. M.; Leboeuf, M.; Salahub, D. R. In *Molecular Electrostatic Potentials*; Murray, J. S., Sen, K., Eds.; Theoretical and Computational Chemistry; Elsevier, 1996; Vol. 3; pp 105 – 142.

(93) Axilrod, B. M.; Teller, E. Interaction of the van der Waals Type Between Three Atoms. *J. Chem. Phys.* **1943**, *11*, 299–300.

(94) Muto, Y. *Proc. Phys. Math. Soc. Jpn.* **1943**, *17*, 629.

(95) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

(96) Ghosh, D. C.; Islam, N. Semiempirical evaluation of the global hardness of the atoms of 103 elements of the periodic table using the most probable radii as their size descriptors. *International Journal of Quantum Chemistry 110*, 1206–1213.

(97) Levenberg, K. A Method for the Solution of Certain Non-linear Problems in Least Squares. *Q. Appl. Math.* **1944**, *2*, 164–168.

(98) Marquardt, D. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431–441.

(99) TURBOMOLE V7.0 2015, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from `http://www.turbomole.com`.

(100) 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.

(101) Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. Turbomole. *WIREs Comput. Mol. Sci.* **2014**, *4*, 91–100.

(102) Vahtras, O.; Almlöf, J.; Feyereisen, M. W. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.* **1993**, *213*, 514–518.

(103) 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.

(104) Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.

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

(106) Stewart, J. J. P. *MOPAC2016*; Stewart Computational Chemistry: Colorado Springs, CO, USA, 2016; http://OpenMOPAC.net (August 16, 2016).

(107) See `http://www.thch.uni-bonn.de/`.

(108) 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.

(109) 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.

(110) Bühl, M.; Kabrede, H. Geometries of Transition-Metal Complexes from Density-Functional Theory. *J. Chem. Theory Comput.* **2006**, *2*, 1282–1290.

(111) 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.

(112) 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.

(113) Gráfová, L.; Pitoňák, M.; Řezáč, J.; Hobza, P. Comparative Study of Selected Wave Function and Density Functional Methods for Noncovalent Interaction Energy Calculations Using the Extended S22 Data Set. *J. Chem. Theory Comput.* **2010**, *6*, 2365–2376.

(114) Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 4285–4292.

(115) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184–32215.

(116) Marshall, M. S.; Burns, L. A.; Sherrill, C. D. Basis set convergence of the coupled-cluster correction, $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$: 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.

(117) Kozuch, S.; Martin, J. M. L. Halogen Bonds: Benchmarks and Theoretical Analysis. *J. Chem. Theory Comput.* **2013**, *9*, 1918–1931.

(118) Setiawan, D.; Kraka, E.; Cremer, D. Strength of the Pnicogen Bond in Complexes Involving Group Va Elements N, P, and As. *J. Phys. Chem. A* **2015**, *119*, 1642–1656.

(119) Sure, R.; Grimme, S. Comprehensive Benchmark of Association (Free) Energies of Realistic Host–Guest Complexes. *J. Chem. Theory Comput.* **2015**, *11*, 3785–3801.

(120) 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.

(121) 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.

(122) Kesharwani, M. K.; Karton, A.; Martin, J. M. L. Benchmark ab Initio Conformational Energies for the Proteinogenic Amino Acids through Explicitly Correlated Methods. Assessment of Density Functional Methods. *J. Chem. Theory Comput.* **2016**, *12*, 444–454.

(123) 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.

(124) Goerigk, L.; Karton, A.; Martin, J. M. L.; Radom, L. Accurate quantum chemical energies for tetrapeptide conformations: why MP2 data with an insufficient basis set should be handled with caution. *Phys. Chem. Chem. Phys.* **2013**, *15*, 7028–7031.

(125) Csonka, G. I.; French, A. D.; Johnson, G. P.; Stortz, C. A. Evaluation of Density Functionals and Basis Sets for Carbohydrates. *J. Chem. Theory Comput.* **2009**, *5*, 679–692.

(126) Kozuch, S.; Bachrach, S. M.; Martin, J. M. Conformational Equilibria in Butane-1,4-diol: A Benchmark of a Prototypical System with Strong Intramolecular H-bonds. *J. Phys. Chem. A* **2014**, *118*, 293–303.

(127) 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.

(128) 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.

(129) Marianski, M.; Supady, A.; Ingram, T.; Schneider, M.; Baldauf, C. Assessing the Accuracy of Across-the-Scale Methods for Predicting Carbohydrate Conformational Energies for the Examples of Glucose and α-Maltose. *J. Chem. Theory Comput.* **2016**, *12*, 6157–6168.

(130) Dohm, S.; Hansen, A.; Steinmetz, M.; Grimme, S.; Checinski, M. P. Comprehensive Thermochemical Benchmark Set of Realistic Closed-Shell Metal Organic Reactions. *J. Chem. Theory Comput.* **2018**, *14*, 2596–2608.

(131) Grimme, S. Supramolecular binding thermodynamics by dispersion corrected density functional theory. *Chem. Eur. J.* **2012**, *18*, 9955–9964.

(132) 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.

(133) 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.

(134) Karton, A.; Goerigk, L. Accurate reaction barrier heights of pericyclic reactions: Surprisingly large deviations for the CBS-QB3 composite method and their consequences in DFT benchmark studies. *J. Comput. Chem. 36*, 622–632.

(135) Goerigk, L.; Sharma, R. The INV24 test set: how well do quantum-chemical methods describe inversion and racemization barriers? *Can. J. Chem.* **2016**, *94*, 1133–1143.

(136) Karton, A.; O'Reilly, R. J.; Chan, B.; Radom, L. Determination of Barrier Heights for Proton Exchange in Small Water, Ammonia, and Hydrogen Fluoride Clusters with G4(MP2)-Type, MPn, and SCS-MPn Procedures–A Caveat. *J. Chem. Theory Comput.* **2012**, *8*, 3128–3136.

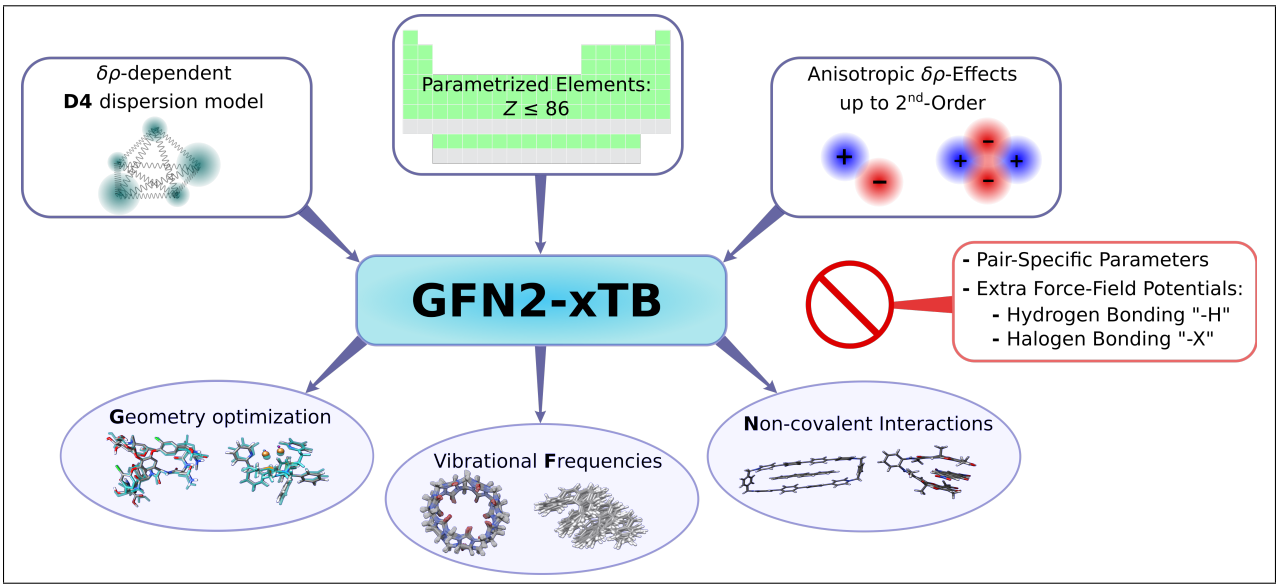
(137) Karton, A.; O'Reilly, R. J.; Radom, L. Assessment of Theoretical Procedures for Calculating Barrier Heights for a Diverse Set of Water-Catalyzed Proton-Transfer Reactions. *J. Phys. Chem. A* **2012**, *116*, 4211–4221.

(138) 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.

(139) Hait, D.; Head-Gordon, M. How Accurate Is Density Functional Theory at Predicting Dipole Moments? An Assessment Using a New Database of 200 Benchmark Values. *J. Chem. Theory Comput.* **2018**, *14*, 1969–1981.

(140) `xtb` standalone code (version 5.8.1). Please contact xtb@thch.uni-bonn.de for the program.

Graphical TOC Entry



Supporting Information: GFN2-xTB – an Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions

Christoph Bannwarth,[*,†,‡] Sebastian Ehlert,[†] and Stefan Grimme[*,†]

†*Mulliken Center for Theoretical Chemistry, Universität Bonn, Beringstr. 4, 53115 Bonn, Germany*

‡*New address: Department of Chemistry, Stanford University, Stanford, CA 94305, United States of America.*

E-mail: christoph.bannwarth@stanford.edu; grimme@thch.uni-bonn.de

Phone: +49-228/73-2351

Contents

List of Tables

1 Electrostatic and exchange-correlation energy contribution for second order density fluctuations

If no integration ranges are given, the integration goes from $-\infty$ to $+\infty$. In density functional tight-binding, the electrostatic and exchange-correlation contribution to the total energy resulting from second order density fluctuations is given as.[S1]

$$E_{\mathrm{ES}}^{(2)} + E_{\mathrm{XC}}^{(2)} = \frac{1}{2} \iint \left(\frac{1}{r_{ij}} + \left. \frac{\partial^2 E_{\mathrm{XC}}}{\partial\rho(\mathbf{r}_i)\partial\rho(\mathbf{r}_j)} \right|_{\rho=\rho_0} \right) \delta\rho(\mathbf{r}_i)\delta\rho(\mathbf{r}_j)\mathrm{d}\mathbf{r}_i\mathrm{d}\mathbf{r}_j \tag{1}$$

The two terms will be approximated individually in the following.

1.1 Anisotropic electrostatics

1.1.1 Multipole expansion in two electronic variables

We will first assume that the Coulomb forces on the same atomic site can be neglected or effectively be absorbed in the XC contribution. Then, we will assume distant atoms and factorize the integral of the Coulomb operator:

$$\lim_{R_{AB}\to\infty} E_{\mathrm{ES}}^{(2)} \approx \frac{1}{2} \sum_{i\in A, j\in B} \frac{\int \rho_A(\mathbf{r}_i)\mathrm{d}\mathbf{r}_i \int \rho_B(\mathbf{r}_j)\mathrm{d}\mathbf{r}_j}{r_{ij}} \,. \tag{2}$$

Here, the particle i is in proximity to the site (atomic center) A with $\mathbf{r}_i = \mathbf{R}_A + \mathbf{r}_{Ai}$ and particle j is close to corresponding center B with $\mathbf{r}_j = \mathbf{R}_B + \mathbf{r}_{Bj}$, The Coulomb operator can

be expanded in a Cartesian multipole expansion to (showing all terms up to second order)

$$
\begin{aligned}
\frac{1}{r_{ij}} \approx & \left.\frac{1}{r_{ij}}\right|_{\mathbf{r}_{Ai},\mathbf{r}_{Bj}=\mathbf{0}} + \sum_{\alpha=x,y,z}\left[(\alpha_i-\alpha_A)\frac{\partial}{\partial\alpha_i}+(\alpha_j-\alpha_B)\frac{\partial}{\partial\alpha_j}\right]\left.\frac{1}{r_{ij}}\right|_{\mathbf{r}_{Ai},\mathbf{r}_{Bj}=\mathbf{0}} \\
& + \frac{1}{2}\sum_{\alpha,\beta=x,y,z}\left[(\alpha_i-\alpha_A)\frac{\partial}{\partial\alpha_i}+(\alpha_j-\alpha_B)\frac{\partial}{\partial\alpha_j}\right]\left[(\beta_i-\beta_A)\frac{\partial}{\partial\beta_i}+(\beta_j-\beta_B)\frac{\partial}{\partial\beta_j}\right]\left.\frac{1}{r_{ij}}\right|_{\mathbf{r}_{Ai},\mathbf{r}_{Bj}=\mathbf{0}} \\
& + \cdots \\
= & \left.\frac{1}{r_{ij}}\right|_{\mathbf{r}_{Ai},\mathbf{r}_{Bj}=\mathbf{0}} + \sum_{\alpha=x,y,z}\left.\frac{(\alpha_i-\alpha_A)\,\alpha_{ij}+(\alpha_j-\alpha_B)\,\alpha_{ji}}{r_{ij}^3}\right|_{\mathbf{r}_{Ai},\mathbf{r}_{Bj}=\mathbf{0}} \\
& + \frac{1}{2}\sum_{\alpha,\beta=x,y,z}\left[\frac{(\alpha_i-\alpha_A)(\beta_i-\beta_A)(3\alpha_{ij}\beta_{ij}-\delta_{\alpha\beta}r_{ij}^2)+(\alpha_j-\alpha_B)(\beta_i-\beta_A)(3\alpha_{ji}\beta_{ij}+\delta_{\alpha\beta}r_{ij}^2)}{r_{ij}^5}\right. \\
& \left. + \frac{(\alpha_i-\alpha_A)(\beta_j-\beta_B)(3\alpha_{ij}\beta_{ji}+\delta_{\alpha\beta}r_{ij}^2)+(\alpha_j-\alpha_B)(\beta_j-\beta_B)(3\alpha_{ji}\beta_{ji}-\delta_{\alpha\beta}r_{ij}^2)}{r_{ij}^5}\right]_{\mathbf{r}_{Ai},\mathbf{r}_{Bj}=\mathbf{0}} \qquad (3) \\
& + \cdots \\
= & \frac{1}{R_{AB}} + \sum_{\alpha=x,y,z}\left[\frac{(\alpha_i-\alpha_A)\,\alpha_{AB}}{R_{AB}^3}+\frac{(\alpha_j-\alpha_B)\,\alpha_{BA}}{R_{AB}^3}\right] \\
& + \frac{1}{2}\sum_{\alpha,\beta=x,y,z}\left[\frac{(\alpha_i-\alpha_A)(\beta_i-\beta_A)(3\alpha_{AB}\beta_{AB}-\delta_{\alpha\beta}R_{AB}^2)}{R_{AB}^5}\right. \\
& + \frac{(\alpha_j-\alpha_B)(\beta_i-\beta_A)(3\alpha_{BA}\beta_{AB}+\delta_{\alpha\beta}R_{AB}^2)}{R_{AB}^5} \\
& + \frac{(\alpha_i-\alpha_A)(\beta_j-\beta_B)(3\alpha_{AB}\beta_{BA}+\delta_{\alpha\beta}R_{AB}^2)}{R_{AB}^5} \\
& \left. + \frac{(\alpha_j-\alpha_B)(\beta_j-\beta_B)(3\alpha_{BA}\beta_{BA}-\delta_{\alpha\beta}R_{AB}^2)}{R_{AB}^5}\right] + \cdots .
\end{aligned}
$$

Here, $r_{ij} = |\mathbf{r}_j - \mathbf{r}_i|$, $R_{AB} = |\mathbf{R}_B - \mathbf{R}_A|$, and $\alpha_{AB} = -\alpha_{BA} = \alpha_B - \alpha_A$. α and β represent Cartesian components. Energy terms of higher order in $\mathbf{r}_{Ai}$ and $\mathbf{r}_{Bj}$ are not shown in Eq. 3 and will be neglected in our approach. This way, all energy terms up to second order are considered, which ensures that all electrostatic interactions decaying with R_{AB}^{-3} or slower are taken into account. Insertion of Eq. 3 into Eq. 2 and integrating over all positive (clamped

nuclei) and negative (electrons) particles then yields

$$\begin{aligned}\lim_{R_{AB}\to\infty} E_{\mathrm{ES}}^{(2)} =& \frac{1}{2}\sum_{A\neq B}\Bigg[\frac{q_A q_B}{R_{AB}} + \frac{q_B\left(\boldsymbol{\mu}_A^T\mathbf{R}_{AB}\right) + q_A\left(\boldsymbol{\mu}_B^T\mathbf{R}_{BA}\right)}{R_{AB}^3} \\ &- \frac{3\left(\boldsymbol{\mu}_A^T\mathbf{R}_{AB}\right)\left(\boldsymbol{\mu}_B^T\mathbf{R}_{AB}\right) - \left(\boldsymbol{\mu}_A^T\boldsymbol{\mu}_B\right)R_{AB}^2}{R_{AB}^5} \\ &+ \frac{q_B\mathbf{R}_{AB}^T\boldsymbol{\Theta}_A\mathbf{R}_{AB}}{R_{AB}^5} + \frac{q_A\mathbf{R}_{AB}^T\boldsymbol{\Theta}_B\mathbf{R}_{AB}}{R_{AB}^5}\Bigg]\end{aligned} \tag{4}$$

with

$$\begin{aligned}q_A &= Z_A - \sum_{\kappa\in A}\sum_{\lambda} P_{\kappa\lambda}\underbrace{\langle\phi_\lambda \mid \phi_\kappa\rangle}_{S_{\lambda\kappa}} \\ \mu_A^\alpha &= \sum_{\kappa\in A}\sum_{\lambda} P_{\kappa\lambda}\left(\alpha_A S_{\lambda\kappa} - \underbrace{\langle\phi_\lambda\,|\alpha_i|\,\phi_\kappa\rangle}_{D^\alpha_{\lambda\kappa}}\right) = -\sum_{\kappa\in A}\sum_{\lambda} P_{\kappa\lambda}D^\alpha_{\lambda\kappa A} \\ \theta_A^{\alpha\beta} &= \sum_{\kappa\in A}\sum_{\lambda} P_{\kappa\lambda}\left(\alpha_A D^\beta_{\lambda\kappa} + \beta_A D^\alpha_{\lambda\kappa} - \alpha_A\beta_A S_{\lambda\kappa} - \underbrace{\langle\phi_\lambda\,|\alpha_i\beta_i|\,\phi_\kappa\rangle}_{Q^{\alpha\beta}_{\lambda\kappa}}\right) = -\sum_{\kappa\in A}\sum_{\lambda} P_{\kappa\lambda}Q^{\alpha\beta}_{\lambda\kappa A}\end{aligned} \tag{5}$$

and

$$\Theta_A^{\alpha\beta} = \frac{3}{2}\theta_A^{\alpha\beta} - \frac{\delta_{\alpha\beta}}{2}\left(\theta_A^{xx} + \theta_A^{yy} + \theta_A^{zz}\right) . \tag{6}$$

Here, q_A, $\boldsymbol{\mu}_A$, and $\boldsymbol{\Theta}_A$ are the cumulative atomic monopole (i.e., charge), dipole, and quadrupole moment, respectively. $S_{\lambda\kappa}$, $D^\alpha_{\lambda\kappa}$, and $Q^{\alpha\beta}_{\lambda\kappa}$ are the overlap, electric dipole, and electric quadrupole moments, respectively. The extra index A in the multipole integrals $D^\alpha_{\lambda\kappa A}$ and $Q^{\alpha\beta}_{\lambda\kappa A}$ indicates that these are evaluated with the origin at the corresponding atomic center, while $D^\alpha_{\lambda\kappa}$ and $Q^{\alpha\beta}_{\lambda\kappa}$ are given with origin $\mathbf{O} = (0\ 0\ 0)^T$.

Since the isotropic charge-charge term (first term in Eq. 4) is already captured in a shell-wise manner, the anisotropic terms represent the newly introduced terms in the anisotropic electrostatic (AES) energy. Adding a damping function to damp the AES contributions at short range then leads to the expression given in the manuscript.

1.2 Anisotropic XC kernel contribution

The second order change in the XC energy takes the form of a static XC kernel $f_{\mathrm{XC}}(\mathbf{r}_i, \mathbf{r}_j)$. In accordance with the (semi-)local density functional origin, a local density approximation is applied, which restricts this contribution to same-site terms only:

$$\begin{aligned} E_{\mathrm{XC}}^{(2)} =& \frac{1}{2} \iint \left. \frac{\partial^2 E_{\mathrm{XC}}}{\partial\rho(\mathbf{r}_i)\partial\rho(\mathbf{r}_j)} \right|_{\rho=\rho_0} \delta\rho(\mathbf{r}_i)\delta\rho(\mathbf{r}_j)\mathrm{d}\mathbf{r}_i\mathrm{d}\mathbf{r}_j \\ \approx& \frac{1}{2} \iint \delta(\mathbf{r}_i - \mathbf{r}_j)\, f_{\mathrm{XC}}(\mathbf{r}_i, \mathbf{r}_j)\, \delta\rho(\mathbf{r}_i)\delta\rho(\mathbf{r}_j)\mathrm{d}\mathbf{r}_i\mathrm{d}\mathbf{r}_j \\ =& \frac{1}{2} \int f_{\mathrm{XC}}(\mathbf{r}_i)\, \delta\rho^2(\mathbf{r}_i)\mathrm{d}\mathbf{r}_i \end{aligned} \tag{7}$$

In the last line, we have dropped the second redundant index of f_{XC}. For distant atoms with non-overlapping atomic reference densities ρ_{A_0}, the integral can be approximately partitioned in terms of atomic contributions.

$$E_{\mathrm{XC}}^{(2)} \approx \frac{1}{2} \sum_A \int f_{\mathrm{XC}}(\mathbf{r}_i)\, \delta\rho_A^2(\mathbf{r}_i)\mathrm{d}\mathbf{r}_i = \sum_A E_{\mathrm{XC}}^{(2),A} \tag{8}$$

The XC kernel is evaluated on the reference density ρ_0 (see Eq. 7). Hence, in terms of spherically symmetric atomic reference densities, there is no angular dependence (cf. spherical coordinates) of this kernel. So, in terms of spherical coordinates, this integral around the nuclear center A can be written as ($x_{Ai} = r\sin(\vartheta)\cos(\varphi)$, $y_{Ai} = r\sin(\vartheta)\sin(\varphi)$, and $z_{Ai} = r\cos(\vartheta)$):

$$E_{\mathrm{XC}}^{(2),A} = \frac{1}{2} \int_0^{2\pi} \int_0^{\pi} \int_0^{\infty} \left(\sqrt{f_{\mathrm{XC}}(r)}\delta\rho(r,\vartheta,\varphi) \right)^2 r^2 \sin(\vartheta)\mathrm{d}r\mathrm{d}\vartheta\mathrm{d}\varphi \tag{9}$$

We will not consider details regarding specific prefactors or the explicit decay of ρ_{A_0}, as the final energy expression relies on fitted parameters anyhow. Nevertheless, for the monotonically decaying reference densities, an LDA-based exchange kernel takes the form $f_{\mathrm{XC}}(r) \sim \rho_{A_0}^{-2/3}(r) \sim e^{cr}$ (c is an atom-specific constant) , i.e., a monotonically increasing function with

distance from the nucleus. Due to the faster exponential decay of the density fluctuation, the term in parentheses in Eq. 9 will decay to zero for $r \to \infty$. Next, we decompose this term into a product of a radial part and an angular part. The angular part will be expressed by a series of spherical harmonic functions.

$$\sqrt{f_{\mathrm{XC}}\,(r)}\delta\rho(r,\vartheta,\varphi) = R(r)\sum_{l}^{\infty}\sum_{m=-l}^{l} a_{l,m}Y_l^m(\vartheta,\varphi) \tag{10}$$

The integral in Eq. 9 then becomes:

$$E_{\mathrm{XC}}^{(2),A} = \frac{1}{2}\int_0^{2\pi}\int_0^{\pi}\left[\sum_{l}^{\infty}\sum_{m=-l}^{l} a_{l,m}Y_l^m(\vartheta,\varphi)\right]^2 sin(\vartheta)\mathrm{d}\vartheta\mathrm{d}\varphi\int_0^{\infty}\left[R(r)r\right]^2\ \mathrm{d}r \tag{11}$$

Next, we exploit the orthogonality properties of spherical harmonics, to expand the squared sum, i.e., we use:

$$\int_0^{2\pi}\int_0^{\pi} Y_l^m(\vartheta,\varphi)Y_{l'}^{m'}(\vartheta,\varphi)sin(\vartheta)\mathrm{d}\vartheta\mathrm{d}\varphi = \delta_{mm'}\delta_{ll'} \tag{12}$$

This way, the integral in Eq. 13 is equivalent to:

$$E_{\mathrm{XC}}^{(2),A} = \sum_{l}^{\infty}\sum_{m=-l}^{l} a_{l,m}^2\frac{1}{2}\int_0^{2\pi}\int_0^{\pi}\left[Y_l^m(\vartheta,\varphi)\right]^2 sin(\vartheta)\mathrm{d}\vartheta\mathrm{d}\varphi\int_0^{\infty}\left[R(r)r\right]^2\ \mathrm{d}r \tag{13}$$

In the derivation of the anisotropic XC contribution in GFN2-xTB, we now make the following approximations: First, we approximate the angular part as:
$\int_0^{2\pi}\int_0^{\pi}\left[Y_l^m(\vartheta,\varphi)\right]^2 sin(\vartheta)\mathrm{d}\vartheta\mathrm{d}\varphi \approx \left[\int_0^{2\pi}\int_0^{\pi} Y_l^m(\vartheta,\varphi)sin(\vartheta)\mathrm{d}\vartheta\mathrm{d}\varphi\right]^2$
This is a lower bound approximation to the integral, but this way, it becomes possible to use the same CAMM expressions as in the anisotropic electrostatic case. Then we truncate the series at $l = 2$ and express the product of the angular and radial integral as the squared atomic multipole moment (Mulliken approximation) times an l-dependent constant. The XC energy to second order in the density fluctuations is then given as a sum over local CAMM

contributions:

$$E_{\mathrm{XC}}^{(2)} \approx \sum_{A} \left(f_{\mathrm{XC}}^{q_A} q_A^2 + \underbrace{f_{\mathrm{XC}}^{\mu_A} \left|\boldsymbol{\mu}_A\right|^2 + f_{\mathrm{XC}}^{\Theta_A} \left|\left|\boldsymbol{\Theta}_A\right|\right|^2}_{E_{\mathrm{AXC}}^{A}} \right) \tag{14}$$

The zeroth order monopole term is already incorporated within the short-ranged damping/on-site contribution in the second order isotropic Coulomb interaction as in DFTB[S1,S2] and GFN-xTB.[S3] Here the proportionality constant relates to the chemical hardness of the atom. The same is true (in a shell-wise manner) in GFN2-xTB. The other terms in Eq. 14 describe polarization effects of the atomic density up to second order in the multipole expansion and are included in E_{AXC} in GFN2-xTB.

1.3 Derivation of the potential

A derivation for the tight-binding Fock matrix elements is given below. The final expressions are given in Eqs. 22, 23, 24, and 25, as well as in the manuscript. To derive the potential for the Hamiltonian (or Fock) matrix construction, the total energy including orthonormality constraint needs to be differentiated w.r.t. the molecular orbital coefficients.

$$\frac{\partial}{\partial c_{\nu i}} \left[E_{\mathrm{GFN2\text{-}xTB}} - \sum_{j} n_j \epsilon_j \left(\sum_{A,B} \sum_{\kappa \in A} \sum_{\lambda \in B} c_{\kappa j} c_{\lambda j} S_{\kappa \lambda} - 1 \right) \right] = 0 \tag{15}$$

The derivation for an isotropic second and third order density fluctuation tight-binding model is described extensively in Ref. S2.

1.3.1 Anisotropic electrostatic terms

The potential of the newly included terms can be derived analogously to Ref. S2.

$$\begin{aligned}\frac{\partial E_{q\mu}}{\partial c_{\nu i}} =& \sum_{A,B} f_3(R_{AB}) \left[\frac{\partial q_B}{\partial c_{\nu i}} \left(\boldsymbol{\mu}_A^T \mathbf{R}_{AB} \right) - q_A \left(\frac{\partial \boldsymbol{\mu}_B^T}{\partial c_{\nu i}} \mathbf{R}_{AB} \right) \right] \\ =& \sum_{A,C} \sum_{\lambda \in C} \Big[f_3(R_{AD}) \left(q_A \mathbf{D}_{\nu\lambda D}^T \mathbf{R}_{AD} - S_{\nu\lambda} \boldsymbol{\mu}_A^T \mathbf{R}_{AD} \right) \\ &+ f_3(R_{AC}) \left(q_A \mathbf{D}_{\nu\lambda C}^T \mathbf{R}_{AC} - S_{\nu\lambda} \boldsymbol{\mu}_A^T \mathbf{R}_{AC} \right) \Big] n_i c_{\lambda i}\end{aligned} \tag{16}$$

$$\begin{aligned}\frac{\partial E_{\mu\mu}}{\partial c_{\nu i}} =& -\sum_{A,B} f_5(R_{AB}) R_{AB}^2 \left(\boldsymbol{\mu}_A^T \frac{\partial \boldsymbol{\mu}_B}{\partial c_{\nu i}} \right) - 3 f_5(R_{AB}) \left(\boldsymbol{\mu}_A^T \mathbf{R}_{AB} \right) \left(\frac{\partial \boldsymbol{\mu}_B^T}{\partial c_{\nu i}} \mathbf{R}_{AB} \right) \\ =& \sum_{A,C} \sum_{\lambda \in C} \Big[3 f_5(R_{AC}) \left(\boldsymbol{\mu}_A^T \mathbf{R}_{AC} \right) \left(\mathbf{D}_{\nu\lambda C}^T \mathbf{R}_{AC} \right) - f_5(R_{AC}) R_{AC}^2 \left(\boldsymbol{\mu}_A^T \mathbf{D}_{\nu\lambda C} \right) \\ &+ 3 f_5(R_{AD}) \left(\boldsymbol{\mu}_A^T \mathbf{R}_{AD} \right) \left(\mathbf{D}_{\nu\lambda D}^T \mathbf{R}_{AD} \right) - f_5(R_{AD}) R_{AD}^2 \left(\boldsymbol{\mu}_A^T \mathbf{D}_{\nu\lambda D} \right) \Big] n_i c_{\lambda i}\end{aligned} \tag{17}$$

$$\begin{aligned}\frac{\partial E_{q\Theta}}{\partial c_{\nu i}} =& \sum_{A,B} f_5(R_{AB}) \left[\frac{\partial q_B}{\partial c_{\nu i}} \mathbf{R}_{AB}^T \boldsymbol{\Theta}_A \mathbf{R}_{AB} + q_A \mathbf{R}_{AB}^T \frac{\partial \boldsymbol{\Theta}_B}{\partial c_{\nu i}} \mathbf{R}_{AB} \right] \\ =& \sum_{A,C} \sum_{\lambda \in C} \Big[q_A \Big(f_5(R_{AD}) \operatorname{Tr}(\mathbf{Q}_{\nu\lambda D}) R_{AD}^2 + f_5(R_{AC}) \operatorname{Tr}(\mathbf{Q}_{\nu\lambda C}) R_{AC}^2 \\ &- f_5(R_{AD}) \mathbf{R}_{AD}^T \mathbf{Q}_{\nu\lambda D} \mathbf{R}_{AD} - f_5(R_{AC}) \mathbf{R}_{AC}^T \mathbf{Q}_{\nu\lambda C} \mathbf{R}_{AC} \Big) \\ &- S_{\nu\lambda} \left(f_5(R_{AD}) \mathbf{R}_{AD}^T \boldsymbol{\Theta}_A \mathbf{R}_{AD} + f_5(R_{AC}) \mathbf{R}_{AC}^T \boldsymbol{\Theta}_A \mathbf{R}_{AC} \right) \Big] n_i c_{\lambda i}\end{aligned} \tag{18}$$

Here, the terms involving the traces of the Cartesian quadrupole moment integral tensors $\mathbf{Q}_{\nu\lambda D}$ and $\mathbf{Q}_{\nu\lambda D}$ originate from the trace removal term (c.f. Eq. 6).

1.3.2 Anisotropic XC terms

The anisotropic XC (AXC) kernel contributions can be derived analogously:

$$\begin{aligned}\frac{\partial E_{\mu XC}}{\partial c_{\nu i}} &= \frac{\partial}{\partial c_{\nu i}} \sum_A f_{XC}^{\mu_A} \left(\boldsymbol{\mu}_A^T \boldsymbol{\mu}_A\right) \\ &= 2 \sum_A f_{XC}^{\mu_A} \left(\boldsymbol{\mu}_A^T \frac{\partial \boldsymbol{\mu}_A}{\partial c_{\nu i}}\right) \\ &= -2 n_i \sum_C \sum_{\lambda \in C} c_{\lambda i} \left(f_{XC}^{\mu_D} \boldsymbol{\mu}_D^T \mathbf{D}_{\nu\lambda D} + f_{XC}^{\mu_C} \boldsymbol{\mu}_C^T \mathbf{D}_{\nu\kappa C}\right)\end{aligned} \tag{19}$$

$$\begin{aligned}\frac{\partial E_{\Theta XC}}{\partial c_{\nu i}} &= \frac{\partial}{\partial c_{\nu i}} \sum_A f_{XC}^{\Theta_A} \left|\left|\boldsymbol{\Theta}_A\right|\right|^2 \\ &= 2 \sum_A f_{XC}^{\Theta_A} \sum_{\alpha,\beta} \Theta_A^{\alpha\beta} \frac{\partial \Theta_A^{\alpha\beta}}{\partial c_{\nu i}} \\ &= \sum_A f_{XC}^{\Theta_A} \sum_{\alpha,\beta} \Theta_A^{\alpha\beta} \left(3 \frac{\partial \theta_A^{\alpha\beta}}{\partial c_{\nu i}} - \delta_{\alpha\beta} \frac{\partial \left(\theta_A^{xx} + \theta_A^{yy} + \theta_A^{zz}\right)}{\partial c_{\nu i}}\right) \\ &= -n_i \sum_C \sum_{\lambda \in C} c_{\lambda i} \sum_{\alpha,\beta} \left(3 f_{XC}^{\Theta_D} \Theta_D^{\alpha\beta} Q_{\nu\lambda D}^{\alpha\beta} + 3 f_{XC}^{\Theta_C} \Theta_C^{\alpha\beta} Q_{\nu\kappa C}^{\alpha\beta}\right) \\ &\quad + n_i \sum_C \sum_{\lambda \in C} c_{\lambda i} \sum_{\alpha} \left[f_{XC}^{\Theta_D} \Theta_D^{\alpha\alpha} \mathrm{Tr}(\mathbf{Q}_{\nu\lambda D}) + f_{XC}^{\Theta_C} \Theta_C^{\alpha\alpha} \mathrm{Tr}(\mathbf{Q}_{\nu\kappa C})\right]\end{aligned} \tag{20}$$

1.3.3 Fock matrix elements

Dividing all Eqs. 16–20 by $2n_i$, we can now obtain the AES and AXC expressions entering the Fock matrix.

$$\sum_C \sum_{\lambda \in C} c_{\lambda i} \left[\left(H_{\kappa\lambda}^0 + F_{\kappa\lambda}^{\mathrm{IES+IXC}} + \underbrace{F_{\kappa\lambda}^{\mathrm{AES}} + F_{\kappa\lambda}^{\mathrm{AXC}}}_{F_{\kappa\lambda}^{\mathrm{aniso}}}\right) - \varepsilon_i S_{\kappa\lambda}\right] = 0 \tag{21}$$

It is more convenient during the SCF calculations to work with the dipole and quadrupole integrals with the same overall origin (c.f. Eq. 5). By rearrangement of all terms, the respective Fock matrix elements for the anisotropic contributions $F_{\kappa\lambda}^{\mathrm{aniso}} = F_{\kappa\lambda}^{\mathrm{AES}} + F_{\kappa\lambda}^{\mathrm{AXC}}$ to terms, which are proportional to the respective overlap, dipole and quadrupole integrals, yields to

the following expression:

$$F_{\kappa\lambda}^{\text{aniso}} = \frac{1}{2} S_{\kappa\lambda} \left[V_S(\mathbf{R}_B) + V_S(\mathbf{R}_C) \right] \tag{22a}$$

$$+ \frac{1}{2} \mathbf{D}_{\kappa\lambda}^T \left[\mathbf{V}_D(\mathbf{R}_B) + \mathbf{V}_D(\mathbf{R}_C) \right] \tag{22b}$$

$$+ \frac{1}{2} \sum_{\alpha,\beta} Q_{\kappa\lambda}^{\alpha\beta} \left[V_Q^{\alpha\beta}(\mathbf{R}_B) + V_Q^{\alpha\beta}(\mathbf{R}_C) \right] , \ \forall \ \kappa \in B, \lambda \in C \tag{22c}$$

With the respective integral (overlap, dipole, and quadrupole) proportional potential terms given as:

$$\begin{aligned} V_S(\mathbf{R}_C) = & \sum_A \Big\{ \mathbf{R}_C^T \left[f_5(R_{AC}) \boldsymbol{\mu}_A R_{AC}^2 - \mathbf{R}_{AC} 3 f_5(R_{AC}) \left(\boldsymbol{\mu}_A^T \mathbf{R}_{AC} \right) - f_3(R_{AC}) q_A \mathbf{R}_{AC} \right] \\ & - f_5(R_{AC}) \mathbf{R}_{AC}^T \boldsymbol{\Theta}_A \mathbf{R}_{AC} - f_3(R_{AC}) \boldsymbol{\mu}_A^T \mathbf{R}_{AC} + q_A f_5(R_{AC}) \frac{1}{2} \mathbf{R}_C^2 \mathbf{R}_{AC}^2 \\ & - \frac{3}{2} q_A f_5(R_{AC}) \sum_{\alpha,\beta} \alpha_{AB} \beta_{AB} \alpha_C \beta_C \Big\} \\ & + 2 f_{XC}^{\mu_C} \mathbf{R}_C^T \boldsymbol{\mu}_C - f_{XC}^{\Theta_C} \mathbf{R}_C^T \left[3 \boldsymbol{\Theta}_C - \text{Tr}(\boldsymbol{\Theta}_C) \, \mathbf{I} \right] \mathbf{R}_C \end{aligned} \tag{23}$$

$$\begin{aligned} \mathbf{V}_D(\mathbf{R}_C) = & \sum_A \Big[\mathbf{R}_{AC} 3 f_5(R_{AC}) \left(\boldsymbol{\mu}_A^T \mathbf{R}_{AC} \right) - f_5(R_{AC}) \boldsymbol{\mu}_A R_{AC}^2 + f_3(R_{AC}) q_A \mathbf{R}_{AC} \\ & - q_A f_5(R_{AC}) \mathbf{R}_C R_{AC}^2 + 3 q_A f_5(R_{AC}) \mathbf{R}_{AC} \sum_\alpha \alpha_C \alpha_{AC} \Big] \\ & - 2 f_{XC}^{\mu_C} \boldsymbol{\mu}_C + 2 f_{XC}^{\Theta_C} \left[3 \boldsymbol{\Theta}_C - \text{Tr}(\boldsymbol{\Theta}_C) \, \mathbf{I} \right] \mathbf{R}_C \end{aligned} \tag{24}$$

$$\begin{aligned} V_Q^{\alpha\beta}(\mathbf{R}_C) = & - \sum_A q_A f_5(R_{AC}) \left[\frac{3}{2} \alpha_{AC} \beta_{AC} - \frac{1}{2} R_{AB}^2 \right] \\ & - f_{XC}^{\Theta_C} \left[3 \Theta_C^{\alpha\beta} - \delta_{\alpha\beta} \sum_\alpha \Theta_C^{\alpha\alpha} \right] \end{aligned} \tag{25}$$

1.4 Cartesian gradients of $E^{\mathbf{AES}}$

A derivation for the nuclear gradients is given below. For the final expressions, see Section 1.4.5.

1.4.1 General

$$\frac{\partial R_{AB}}{\partial \alpha_A} = -\frac{\alpha_{AB}}{R_{AB}} = -\frac{\partial R_{BA}}{\partial \alpha_A} = -\frac{\partial R_{AB}}{\partial \alpha_B} \tag{26}$$

The $f_n(R_{AB})$ terms given in the manuscript define the damping and distance dependence of the AES as:

$$f_n(R_{AB}) = R_{AB}^{-n} f_{\text{damp}}(a_n, R_{AB}) \tag{27}$$

with

$$f_{\text{damp}}(a_n, R_{AB}) = \frac{1}{1 + 6\left(\frac{R_0^{AB}}{R_{AB}}\right)^{a_n}} \ . \tag{28}$$

$$\begin{aligned}
\left.\frac{\partial f_n(R_{AB})}{\partial R_{AB}}\right|_{R_0^{AB}=\text{const.}} =& \frac{\partial R_{AB}^{-n}}{\partial R_{AB}} f_{\text{damp}}(a_n, R_{AB}) + \frac{\partial f_{\text{damp}}(a_n, R_{AB})}{\partial R_{AB}} R_{AB}^{-n} \\
=& - n f_{\text{damp}}(a_n, R_{AB}) R_{AB}^{-n-1} + 6 a_n f_{\text{damp}}^2(a_n, R_{AB}) \left(\frac{R_0^{AB}}{R_{AB}}\right)^{a_n} R_{AB}^{-n-1} \\
=& - \frac{n}{R_{AB}} f_n(R_{AB}) - \frac{a_n}{R_{AB}} \left[f_{\text{damp}}(a_n, R_{AB}) f_n(R_{AB}) - f_n(R_{AB})\right]
\end{aligned} \tag{29}$$

The above equation would be sufficient if the R_0^{AB} were constants. In GFN2-xTB, however, they are dependent on the coordination numbers of the atoms via:

$$R_0^{AB} = \frac{1}{2}\left(R_0^{A\prime} + R_0^{B\prime}\right) \tag{30}$$

$$R_0^{A\prime} = R_0^A + \frac{R_{\text{max}} - R_0^A}{1 + \exp[-\beta(CN_A - N_{\text{val}} - \Delta_{\text{val}})]} \tag{31}$$

Consequently a term, which takes into account the CN-dependence needs to be added, i.e.:

$$\frac{\partial R_0^{A\prime}}{\partial CN_A} = \beta \frac{R_{\max} - R_0^A}{\left[\{1 + \exp[-\beta(CN_A - N_{\text{val}} - \Delta_{\text{val}})]\}\right]^2} \exp[-\beta(CN_A - N_{\text{val}} - \Delta_{\text{val}})] \tag{32}$$

$$\frac{\partial R_0^{A\prime}}{\partial \alpha_C} = \frac{\partial R_0^{A\prime}}{\partial CN_A} \frac{\partial CN_A}{\partial \alpha_C} \tag{33}$$

$$\begin{aligned} \frac{\partial f_n(R_{AB})}{\partial CN_A} &= -\frac{6a_n f_{\text{damp}}^2(a_n, R_{AB})}{R_{AB}^{n+a_n}} \left(R_0^{AB}\right)^{a_n-1} \left(\frac{\partial R_0^{AB}}{\partial CN_A}\right) \\ &= -\frac{3a_n f_{\text{damp}}^2(a_n, R_{AB})}{R_{AB}^{n+a_n}} \left(R_0^{AB}\right)^{a_n-1} \left(\frac{\partial R_0^{A\prime}}{\partial CN_A}\right) \\ &= -\frac{3a_n f_n(R_{AB}) f_{\text{damp}}(a_n, R_{AB})}{R_{AB}^{a_n}} \left(R_0^{AB}\right)^{a_n-1} \left(\frac{\partial R_0^{A\prime}}{\partial CN_A}\right) \end{aligned} \tag{34}$$

Hence for nuclear gradients of $f_n(R_{AB})$, this leads to:

$$f_n^{[\alpha_A]}(R_{AB}) = \frac{\partial f_n(R_{AB})}{\partial \alpha_A} = \left.\frac{\partial f_n(R_{AB})}{\partial R_{AB}}\right|_{R_0^{AB}=\text{const.}} \frac{\partial R_{AB}}{\partial \alpha_A} + \frac{\partial f_n(R_{AB})}{\partial CN_A} \frac{\partial CN_A}{\partial \alpha_A} + \frac{\partial f_n(R_{AB})}{\partial CN_B} \frac{\partial CN_B}{\partial \alpha_A} \tag{35}$$

It should be noted that due to the CN-dependence, the last terms also survive if an atom different from A or B is moved:

$$f_n^{[\alpha_C]}(R_{AB}) = \frac{\partial f_n(R_{AB})}{\partial CN_A} \frac{\partial CN_A}{\partial \alpha_C} + \frac{\partial f_n(R_{AB})}{\partial CN_B} \frac{\partial CN_B}{\partial \alpha_C} \tag{36}$$

The nuclear derivatives of the CAMM expressions yields:

$$\frac{\partial q_A}{\partial \alpha_C} = (\delta_{AC} - 1) \sum_{\kappa \in A} \sum_{\lambda \in C} P_{\kappa\lambda} \frac{\partial S_{\lambda\kappa}}{\partial \alpha_C} - \delta_{AC} \sum_{\kappa \in C} \sum_{B \neq C} \sum_{\lambda \in B} P_{\kappa\lambda} \frac{\partial S_{\lambda\kappa}}{\partial \alpha_C} \tag{37}$$

$$
\begin{aligned}
\frac{\partial \mu_A^\beta}{\partial \alpha_C} &= (1-\delta_{AC}) \sum_{\kappa\in A}\sum_{\lambda\in C} P_{\kappa\lambda}\left[\beta_A \frac{\partial S_{\lambda\kappa}}{\partial \alpha_C} - \frac{\partial D^\beta_{\lambda\kappa}}{\partial \alpha_C}\right] \\
&\quad + \delta_{AC} \sum_{\kappa\in C}\sum_{B}\sum_{\lambda\in B} P_{\kappa\lambda}\left[\delta_{\alpha\beta} S_{\lambda\kappa} + \beta_C \frac{\partial S_{\lambda\kappa}}{\partial \alpha_C} - \frac{\partial D^\beta_{\lambda\kappa}}{\partial \alpha_C}\right] \\
&= (\delta_{AC}-1) \sum_{\kappa\in A}\sum_{\lambda\in C} P_{\kappa\lambda} \frac{\partial D^\beta_{\lambda\kappa A}}{\partial \alpha_C} - \delta_{AC} \sum_{\kappa\in C}\sum_{B\neq C}\sum_{\lambda\in B} P_{\kappa\lambda} \frac{\partial D^\beta_{\lambda\kappa C}}{\partial \alpha_C}
\end{aligned}
\tag{38}
$$

Here, we have made use of the notation of dipole integrals with origin on the respective atoms, which is more convenient for the evaluation of nuclear gradients.

$$
\begin{aligned}
\frac{\partial \theta_A^{\beta\gamma}}{\partial \alpha_C} &= (1-\delta_{AC}) \sum_{\kappa\in A}\sum_{\lambda\in C} P_{\kappa\lambda}\left[\beta_A \frac{\partial D^\gamma_{\lambda\kappa}}{\partial \alpha_C} + \gamma_A \frac{\partial D^\beta_{\lambda\kappa}}{\partial \alpha_C} - \beta_A\gamma_A \frac{\partial S_{\lambda\kappa}}{\partial \alpha_C} - \frac{\partial Q^{\beta\gamma}_{\lambda\kappa}}{\partial \alpha_C}\right] \\
&\quad + \delta_{AC} \sum_{\kappa\in C}\sum_{B}\sum_{\lambda\in B} P_{\kappa\lambda}\left[\beta_C \frac{\partial D^\gamma_{\lambda\kappa}}{\partial \alpha_C} + \gamma_C \frac{\partial D^\beta_{\lambda\kappa}}{\partial \alpha_C} - \beta_C\gamma_C \frac{\partial S_{\lambda\kappa}}{\partial \alpha_C} - \frac{\partial Q^{\beta\gamma}_{\lambda\kappa}}{\partial \alpha_C}\right. \\
&\quad \left. + \delta_{\alpha\beta}\left(D^\gamma_{\lambda\kappa} - \gamma_C S_{\lambda\kappa}\right) + \delta_{\alpha\gamma}\left(D^\beta_{\lambda\kappa} - \beta_C S_{\lambda\kappa}\right)\right] \\
&= (\delta_{AC}-1) \sum_{\kappa\in A}\sum_{\lambda\in C} P_{\kappa\lambda} \frac{\partial Q^{\beta\gamma}_{\lambda\kappa A}}{\partial \alpha_C} - \delta_{AC} \sum_{\kappa\in C}\sum_{B\neq C}\sum_{\lambda\in B} P_{\kappa\lambda} \frac{\partial Q^{\beta\gamma}_{\lambda\kappa C}}{\partial \alpha_C}
\end{aligned}
\tag{39}
$$

Again, we made use of the more convenient notation of quadrupole integrals with origin on the respective atoms.

The nuclear gradient for the AXC terms are obtained as $\mathrm{Tr}(\mathbf{P}\frac{\partial}{\partial \alpha_C}\mathbf{F}^{\mathrm{AXC}})$and are trivial due to their on-site nature. The derivative of the AES energy is given by:

$$
\frac{\partial E_{\mathrm{AES}}}{\partial \alpha_C} = \frac{\partial E_{q\mu}}{\partial \alpha_C} + \frac{\partial E_{q\Theta}}{\partial \alpha_C} + \frac{\partial E_{\mu\mu}}{\partial \alpha_C} \tag{40}
$$

1.4.2 Nuclear gradients for $E_{q\mu}$

$$
\begin{aligned}
\left.\frac{\partial E_{q\mu}}{\partial \alpha_C}\right|_{R_0^{AB}=\text{const.}} &= \frac{\partial}{\partial \alpha_C} \sum_{A,B} f_3(R_{AB}) q_B \boldsymbol{\mu}_A^T \mathbf{R}_{AB} \\
&= \sum_{A,B} \left[\frac{\partial f_3(R_{AB})}{\partial \alpha_C} q_B \boldsymbol{\mu}_A^T \mathbf{R}_{AB} + f_3(R_{AB}) q_B \boldsymbol{\mu}_A^T \frac{\partial \mathbf{R}_{AB}}{\partial \alpha_C} \right. \\
&\quad \left. + f_3(R_{AB}) \left(\frac{\partial q_B}{\partial \alpha_C} \boldsymbol{\mu}_A + q_B \frac{\partial \boldsymbol{\mu}_A}{\partial \alpha_C} \right)^T \mathbf{R}_{AB} \right] \\
&= \sum_B \left[f_3^{[\alpha_C]}(R_{CB}) \left(q_B \boldsymbol{\mu}_C - q_C \boldsymbol{\mu}_B \right)^T \mathbf{R}_{CB} \Big|_{R_0^{AB}=\text{const.}} - f_3(R_{CB}) \left(q_B \mu_C^\alpha - q_C \mu_B^\alpha \right) \right] \\
&\quad + \sum_{A \neq C} \sum_{\kappa \in A} \sum_{\lambda \in C} P_{\kappa\lambda} \frac{\partial S_{\lambda\kappa}}{\partial \alpha_C} \sum_B \boldsymbol{\mu}_B \left[f_3(R_{AB}) \mathbf{R}_{AB} + f_3(R_{CB}) \mathbf{R}_{CB} \right] \\
&\quad - \sum_{A \neq C} \sum_{\kappa \in A} \sum_{\lambda \in C} P_{\kappa\lambda} \sum_B q_B \left[f_3(R_{AB}) \left(\frac{\partial \mathbf{D}_{\lambda\kappa A}}{\partial \alpha_C} \right)^T \mathbf{R}_{AB} + f_3(R_{CB}) \left(\frac{\partial \mathbf{D}_{\lambda\kappa C}}{\partial \alpha_C} \right)^T \mathbf{R}_{CB} \right]
\end{aligned}
\tag{41}
$$

Here, we exploited the symmetry of $\partial S_{\lambda\kappa}/\partial \alpha_C = \partial S_{\kappa\lambda}/\partial \alpha_C$. Note that $\partial D_{\lambda\kappa A}/\partial \alpha_C \neq \partial D_{\kappa\lambda C}/\partial \alpha_C$, but that $\partial D_{\lambda\kappa A}/\partial \alpha_C = -\partial D_{\kappa\lambda A}/\partial \alpha_A$, which can save a factor of two in the gradient calculation.

1.4.3 Nuclear gradients for $E_{q\Theta}$

Let us first consider the derivative

$$
\begin{aligned}
\frac{\partial \left(\mathbf{R}_{AB}^T \boldsymbol{\Theta}_A \mathbf{R}_{AB} \right)}{\partial \alpha_C} =& \delta_{BC} \left(\frac{\partial \mathbf{R}_{AC}^T}{\partial \alpha_C} \boldsymbol{\Theta}_A \mathbf{R}_{AC} + \mathbf{R}_{AC}^T \boldsymbol{\Theta}_A \frac{\partial \mathbf{R}_{AC}}{\partial \alpha_C} \right) \\
&+ \delta_{AC} \left(\frac{\partial \mathbf{R}_{CB}^T}{\partial \alpha_C} \boldsymbol{\Theta}_C \mathbf{R}_{CB} + \mathbf{R}_{CB}^T \boldsymbol{\Theta}_C \frac{\partial \mathbf{R}_{CB}}{\partial \alpha_C} \right) + \mathbf{R}_{AB}^T \frac{\partial \boldsymbol{\Theta}_A}{\partial \alpha_C} \mathbf{R}_{AB} \\
=& 2 \sum_{\beta=x,y,z} \left(\delta_{BC} \Theta_A^{\alpha\beta} \beta_{AC} - \delta_{AC} \Theta_C^{\alpha\beta} \beta_{CB} \right) + \mathbf{R}_{AB}^T \frac{\partial \boldsymbol{\Theta}_A}{\partial \alpha_C} \mathbf{R}_{AB} \\
=& 2 \left(\delta_{BC} \mathbf{R}_{AC}^T \boldsymbol{\Theta}_A^\alpha - \delta_{AC} \mathbf{R}_{CB}^T \boldsymbol{\Theta}_C^\beta \right) + \mathbf{R}_{AB}^T \frac{\partial \boldsymbol{\Theta}_A}{\partial \alpha_C} \mathbf{R}_{AB}
\end{aligned}
\tag{42}
$$

,

with

$$\frac{\partial \mathbf{\Theta}_A}{\partial \alpha_C} = \frac{3}{2}\frac{\partial \theta_A^{\alpha\beta}}{\partial \alpha_C} - \frac{\delta_{\alpha\beta}}{2}\frac{\partial \left(\theta_A^{xx} + \theta_A^{yy} + \theta_A^{zz}\right)}{\partial \alpha_C}. \tag{43}$$

Then the derivative of the charge-quadrupole interaction is given as:

$$\begin{aligned}
\left.\frac{\partial E_{q\Theta}}{\partial \alpha_C}\right|_{R_0^{AB}=\text{const.}} &= \frac{\partial}{\partial \alpha_C}\sum_{AB} f_5(R_{AB})\left(q_B \mathbf{R}_{AB}^T \mathbf{\Theta}_A \mathbf{R}_{AB}\right) \\
&= \sum_B \Big[f_5^{[\alpha_C]}(R_{CB})\left(q_C \mathbf{R}_{CB}^T \mathbf{\Theta}_B \mathbf{R}_{CB} + q_B \mathbf{R}_{CB}^T \mathbf{\Theta}_C \mathbf{R}_{CB}\right)\Big|_{R_0^{AB}=\text{const.}} \\
&\quad - 2 f_5(R_{CB})\left(q_C \mathbf{R}_{CB}^T \mathbf{\Theta}_B^{\alpha} + q_B \mathbf{R}_{CB}^T \mathbf{\Theta}_C^{\beta}\right)\Big] \\
&\quad + \sum_{AB} f_5(R_{AB})\left[\frac{\partial q_A}{\partial \alpha_C}\mathbf{R}_{AB}^T \mathbf{\Theta}_B \mathbf{R}_{AB} + q_B \mathbf{R}_{AB}^T \frac{\partial \mathbf{\Theta}_A}{\partial \alpha_C}\mathbf{R}_{AB}\right] \\
&= \sum_B \Big[f_5^{[\alpha_C]}(R_{CB})\left(q_C \mathbf{R}_{CB}^T \mathbf{\Theta}_B \mathbf{R}_{CB} + q_B \mathbf{R}_{CB}^T \mathbf{\Theta}_C \mathbf{R}_{CB}\right)\Big|_{R_0^{AB}=\text{const.}} \\
&\quad - 2 f_5(R_{CB})\left(q_C \mathbf{R}_{CB}^T \mathbf{\Theta}_B^{\alpha} + q_B \mathbf{R}_{CB}^T \mathbf{\Theta}_C^{\beta}\right)\Big] \\
&\quad - \sum_{A\neq C}\sum_{\kappa\in A}\sum_{\lambda\in C} P_{\kappa\lambda}\frac{\partial S_{\lambda\kappa}}{\partial \alpha_C}\sum_B \left[f_5(R_{AB})\mathbf{R}_{AB}^T \mathbf{\Theta}_B \mathbf{R}_{AB} + f_5(R_{CB})\mathbf{R}_{CB}^T \mathbf{\Theta}_B \mathbf{R}_{CB}\right] \\
&\quad - \sum_{A\neq C}\sum_{\kappa\in A}\sum_{\lambda\in C} P_{\kappa\lambda}\sum_B \frac{q_B}{2}\Big[3 f_5(R_{AB})\mathbf{R}_{AB}^T \frac{\partial \mathbf{Q}_{\lambda\kappa A}}{\partial \alpha_C}\mathbf{R}_{AB} - f_5(R_{AB})\,\mathrm{Tr}(\mathbf{Q}_{\lambda\kappa A})\,R_{AB}^2 \\
&\quad + 3 f_5(R_{CB})\mathbf{R}_{CB}^T \frac{\partial \mathbf{Q}_{\lambda\kappa C}}{\partial \alpha_C}\mathbf{R}_{CB} - f_5(R_{CB})\,\mathrm{Tr}(\mathbf{Q}_{\lambda\kappa C})\,R_{CB}^2\Big]
\end{aligned} \tag{44}$$

1.4.4 Nuclear gradients for $E_{\mu\mu}$

Let us first consider the derivative of the following terms:

$$\begin{aligned}
\frac{\partial\left[\left(\boldsymbol{\mu}_A^T \mathbf{R}_{AB}\right)\left(\boldsymbol{\mu}_B^T \mathbf{R}_{AB}\right)\right]}{\partial \alpha_C} &= \left(\left(\frac{\partial \boldsymbol{\mu}_A}{\partial \alpha_C}\right)^T \mathbf{R}_{AB}\right)\left(\boldsymbol{\mu}_B^T \mathbf{R}_{AB}\right) + \left(\boldsymbol{\mu}_A^T \mathbf{R}_{AB}\right)\left(\left(\frac{\partial \boldsymbol{\mu}_B}{\partial \alpha_C}\right)^T \mathbf{R}_{AB}\right) \\
&\quad + (\delta_{BC} - \delta_{AC})\,\mu_A^{\alpha}\left(\boldsymbol{\mu}_B^T \mathbf{R}_{AB}\right) + (\delta_{BC} - \delta_{AC})\,\mu_B^{\alpha}\left(\boldsymbol{\mu}_A^T \mathbf{R}_{AB}\right)
\end{aligned} \tag{45}$$

$$\frac{\partial\left(\boldsymbol{\mu}_A^T\boldsymbol{\mu}_B R_{AB}^2\right)}{\partial\alpha_C}=\left(\frac{\partial\boldsymbol{\mu}_A}{\partial\alpha_C}\right)^T\boldsymbol{\mu}_B R_{AB}^2+\boldsymbol{\mu}_A^T\left(\frac{\partial\boldsymbol{\mu}_B}{\partial\alpha_C}\right)R_{AB}^2+2\left(\delta_{BC}\boldsymbol{\mu}_A^T\boldsymbol{\mu}_C\alpha_{AC}-\delta_{AC}\boldsymbol{\mu}_C^T\boldsymbol{\mu}_B\alpha_{CB}\right) \tag{46}$$

$$\begin{aligned}\left.\frac{\partial E_{\mu\mu}}{\partial\alpha_C}\right|_{R_0^{AB}=\text{const.}}=&\frac{\partial}{2\partial\alpha_C}\sum_{A,B}f_5(R_{AB})\left[\left(\boldsymbol{\mu}_A^T\boldsymbol{\mu}_B\right)R_{AB}^2-3\left(\boldsymbol{\mu}_A^T\mathbf{R}_{AB}\right)\left(\boldsymbol{\mu}_B^T\mathbf{R}_{AB}\right)\right]\\=&\sum_B f_5^{[\alpha_C]}(R_{CB})\left[\left(\boldsymbol{\mu}_C^T\boldsymbol{\mu}_B\right)R_{CB}^2\Big|_{R_0^{AB}=\text{const.}}-3\left(\boldsymbol{\mu}_C^T\mathbf{R}_{CB}\right)\left(\boldsymbol{\mu}_B^T\mathbf{R}_{CB}\right)\right]\\&-\sum_B f_5(R_{CB})\left[2\boldsymbol{\mu}_C^T\boldsymbol{\mu}_B\alpha_{CB}+3\left(\mu_C^\alpha\boldsymbol{\mu}_B^T+\mu_B^\alpha\boldsymbol{\mu}_C^T\right)\mathbf{R}_{CB}\right]\\&-\sum_{A\neq C}\sum_{\kappa\in A}\sum_{\lambda\in C}P_{\kappa\lambda}\left\{\frac{\partial\mathbf{D}_{\lambda\kappa A}}{\partial\alpha_C}\sum_B\left[f_5(R_{AB})\left(\boldsymbol{\mu}_B R_{AB}^2-3\mathbf{R}_{AB}\left(\boldsymbol{\mu}_B^T\mathbf{R}_{AB}\right)\right)\right]\right.\\&\left.+\frac{\partial\mathbf{D}_{\lambda\kappa C}}{\partial\alpha_C}\sum_B\left[f_5(R_{CB})\left(\boldsymbol{\mu}_B R_{CB}^2-3\mathbf{R}_{CB}\left(\boldsymbol{\mu}_B^T\mathbf{R}_{CB}\right)\right)\right]\right\}\end{aligned} \tag{47}$$

1.4.5 Additional terms in AES derivatives due to CN-dependence of R_0^{AB}

The aforementioned formulas for $\partial E_{\text{AES}}/\partial\alpha_C$ have been given for constant values of R_0^{AB}. Due to their CN-dependence, these terms must, however, be included. Therefore, the aforementioned expression have to be augmented with the respective derivatives:

$$\frac{\partial E_{q\mu}}{\partial\alpha_C}=\left.\frac{\partial E_{q\mu}}{\partial\alpha_C}\right|_{R_0^{AB}=\text{const.}}+\sum_{A,B}\left(\frac{\partial f_3(R_{AB})}{\partial CN_A}\frac{\partial CN_A}{\partial\alpha_C}+\frac{\partial f_3(R_{AB})}{\partial CN_B}\frac{\partial CN_B}{\partial\alpha_C}\right)\left(q_B\boldsymbol{\mu}_A^T\mathbf{R}_{AB}\right) \tag{48}$$

$$\frac{\partial E_{q\Theta}}{\partial\alpha_C}=\left.\frac{\partial E_{q\Theta}}{\partial\alpha_C}\right|_{R_0^{AB}=\text{const.}}+\sum_{A,B}\left(\frac{\partial f_5(R_{AB})}{\partial CN_A}\frac{\partial CN_A}{\partial\alpha_C}+\frac{\partial f_5(R_{AB})}{\partial CN_B}\frac{\partial CN_B}{\partial\alpha_C}\right)\left(q_B\mathbf{R}_{AB}^T\boldsymbol{\Theta}_A\mathbf{R}_{AB}\right) \tag{49}$$

$$\frac{\partial E_{\mu\mu}}{\partial \alpha_C} = \left.\frac{\partial E_{\mu\mu}}{\partial \alpha_C}\right|_{R_0^{AB}=\text{const.}} + \frac{1}{2}\sum_{A,B}\left(\frac{\partial f_5(R_{AB})}{\partial CN_A}\frac{\partial CN_A}{\partial \alpha_C} + \frac{\partial f_5(R_{AB})}{\partial CN_B}\frac{\partial CN_B}{\partial \alpha_C}\right) \times \left[\left(\boldsymbol{\mu}_A^T\boldsymbol{\mu}_B\right)R_{AB}^2 - 3\left(\boldsymbol{\mu}_A^T\mathbf{R}_{AB}\right)\left(\boldsymbol{\mu}_B^T\mathbf{R}_{AB}\right)\right] \tag{50}$$

The derived formulas and their implementation was tested by comparison of computed analytic and numerical numerical derivatives.

1.5 Dispersion

The two-body London dispersion energy for two distant atoms is given as:

$$E_{\text{disp}} = -\sum_{A<B}\frac{3}{\pi R_{AB}^6}\int \overline{\alpha}_A(i\omega)\overline{\alpha}_B(i\omega)\mathrm{d}\omega = -\sum_{A<B}\frac{C_6(\rho_A,\rho_B)}{R_{AB}^6} \tag{51}$$

Here, $\overline{\alpha}_A(i\omega)$ refers to the isotropic dynamic dipole polarizability of atom A. It is a function of the density around the atom, and consequently, the dispersion coefficient is a function of the density on both atoms. Hence, expanding the two-body dispersion energy in terms of density fluctuations leads to:

$$E_{\text{disp}} = -\sum_{A<B}\frac{1}{R_{AB}}\left[C_6(\rho_{A_0},\rho_{B_0}) + C_6(\delta\rho_A,\rho_{B_0}) + C_6(\rho_{A_0},\delta\rho_B) + C_6(\delta\rho_A,\delta\rho_B) + C_6((\delta\rho_A)^2,\rho_{B_0}) + C_6(\rho_{A_0},(\delta\rho_B)^2) + \cdots\right] \tag{52}$$

Here, all terms up to second order in the density fluctuations are shown. In the D4 dispersion model (see below), all terms in the first line of Eq. 52 will explicitly be taken into account, i.e., all terms of zeroth and first order, as well as the two-center second order terms. This is a result from considering first order effects in the polarizabilities and formation of their products to obtain the dispersion cofficient. Formally, the pairwise dipole-quadrupole dispersion coefficient is handled in the same way.

1.5.1 The D4 dispersion energy

The total dispersion energy in the context of GFN2-xTB is given by

$$
\begin{aligned}
E_{\mathrm{D4}} = &-\frac{1}{2}\sum_{A}^{} \sum_{a}^{N_{A,\mathrm{ref}}} \sum_{B}^{} \sum_{b}^{N_{B,\mathrm{ref}}} \xi_A^a(q_A, q_{A,a})\xi_B^b(q_B, q_{B,b}) \\
&\times W_A^a(q_A, q_{A,a}) W_B^b(q_B, q_{B,b}) \sum_{n=6,8}^{} s_n \frac{C_n^{ab}}{R_{AB}^n} f_n^{\mathrm{damp,BJ}}(R_{AB}) \\
&- s_9 \sum_{A>B>C}^{} \frac{(3\cos(\theta_{ABC})\cos(\theta_{BCA})\cos(\theta_{CAB})+1) C_9^{ABC}(CN_{\mathrm{cov}}^A, CN_{\mathrm{cov}}^B, CN_{\mathrm{cov}}^C)}{(R_{AB}R_{AC}R_{BC})^3} \\
&\times f_9^{\mathrm{damp,zero}}(R_{AB}, R_{AC}, R_{BC})
\end{aligned}
\tag{53}
$$

Here, the two-body contribution is augmented with a charge-independent three-body Axilrod-Teller-Muto (ATM) term. The rational damping Becke-Johnson-type damping as in DFT-D3 is used for the two-body contribution

$$
f_n^{\mathrm{damp,BJ}}(R_{AB}) = \frac{R_{AB}^n}{R_{AB}^n + (a_1 \cdot R_{AB}^{\mathrm{crit.}} + a_2)^6} \qquad \text{with} \quad R_{AB}^{\mathrm{crit.}} = \sqrt{\frac{C_8^{AB}}{C_6^{AB}}} \tag{54}
$$

and the zero damping function used for the ATM dispersion is defined slightly different compared to previous implementations of DFT-D3. Namely, we adjust the cutoff radii in the damping function by dropping the factor 4/3 and using the same cutoff radii as in the two-body damping function for a more consistent description of the dispersion energy.

$$
f_9^{\mathrm{damp,zero}}(R_{AB}, R_{AC}, R_{BC}) = \left(1 + 6\left(\sqrt[3]{\frac{R_{AB}^{\mathrm{crit.}} R_{BC}^{\mathrm{crit.}} R_{CA}^{\mathrm{crit.}}}{R_{AB}R_{BC}R_{CA}}}\right)^{16}\right)^{-1} \tag{55}
$$

1.5.2 Derivation of the potential

The potential for the dispersion energy is derived by taking the derivative of the dispersion energy expression with respect to the orbital coefficients. Since the ATM term is not

charge dependent it appears not in the potential. The dependencies on the charge and the coordination number are dropped for brevity.

$$\begin{aligned}\frac{\partial E_{\mathrm{D4}}}{\partial c_{\nu i}} = &-\frac{1}{2}\sum_{A}^{N_{A,\mathrm{ref}}}\sum_{a}^{}\sum_{B}^{N_{B,\mathrm{ref}}}\sum_{b}^{}\frac{\partial \xi_A^a}{\partial c_{\nu i}}\xi_B^b W_A^a W_B^b \sum_{n=6,8} s_n \frac{C_n^{ab}}{R_{AB}^n} f_n \\ &-\frac{1}{2}\sum_{A}^{N_{A,\mathrm{ref}}}\sum_{a}^{}\sum_{B}^{N_{B,\mathrm{ref}}}\sum_{b}^{}\frac{\partial \xi_B^b}{\partial c_{\nu i}}\xi_A^a W_A^a W_B^b \sum_{n=6,8} s_n \frac{C_n^{ab}}{R_{AB}^n} f_n \end{aligned} \tag{56}$$

By renaming the indices we can easily simplify above expression

$$\frac{\partial E_{\mathrm{D4}}}{\partial c_{\nu i}} = -\sum_{A}^{N_{A,\mathrm{ref}}}\sum_{a}^{}\sum_{B}^{N_{B,\mathrm{ref}}}\sum_{b}^{}\frac{\partial \xi_A^a}{\partial q_A}\frac{\partial q_A}{\partial c_{\nu i}}\xi_B^b W_A^a W_B^b \sum_{n=6,8} s_n \frac{C_n^{ab}}{R_{AB}^n} f_n \tag{57}$$

$$= -\sum_{A}^{N_{A,\mathrm{ref}}}\sum_{a}^{}\sum_{B}^{N_{B,\mathrm{ref}}}\sum_{b}^{}\frac{\partial \xi_A^a}{\partial q_A}\left(\delta_{AD}\sum_{C}\sum_{\kappa\in C} n_i c_{\kappa i} S_{\nu\kappa} + \sum_{\mu\in A} n_i c_{\mu i} S_{\nu\mu}\right)\xi_B^b W_A^a W_B^b \sum_{n=6,8} s_n \frac{C_n^{ab}}{R_{AB}^n} f_n \tag{58}$$

$$= -\sum_{d}^{N_{D,\mathrm{ref}}}\frac{\partial \xi_D^d}{\partial q_D}\sum_{B}^{}\sum_{b}^{N_{B,\mathrm{ref}}}\xi_B^b W_D^d W_B^b \sum_{n=6,8} s_n \frac{C_n^{DB}}{R_{DB}^n} f_n \sum_{C}\sum_{\kappa\in C} n_i c_{\kappa i} S_{\nu\kappa} \tag{59}$$

We rename and reorder the terms

$$\frac{\partial E_{\mathrm{D4}}}{\partial c_{\nu i}} = \underbrace{-\sum_{a}^{N_{A,\mathrm{ref}}}\frac{\partial \xi_A^a}{\partial q_A}\sum_{B}^{}\sum_{b}^{N_{B,\mathrm{ref}}}\xi_B^b W_A^a W_B^b \sum_{n=6,8} s_n \frac{C_n^{ab}}{R_{AB}^n} f_n}_{d_A} \sum_{C}\sum_{\kappa\in C} n_i c_{\kappa i} S_{\nu\kappa} \tag{60}$$

Which leads to the compact expression

$$F_{\kappa\lambda}^{\mathrm{D4}} = \frac{1}{2} S_{\kappa\lambda}(d_A + d_B), \quad \forall \kappa \in A, \lambda \in B \tag{61}$$

2 Detailed results

2.1 Structures

Table S1: Comparison of computed rotational constants of twelve medium sized molecules to experimentally derived ones (ROT34)[a] for different semiempirical methods. The individual values are given in MHz.

		GFN2-xTB	ref.
1	A	4299.3	4293.9
	B	1411.8	1395.9
	C	1143.5	1130.2
2	A	2630.8[b]	3322.5
	B	912.7[b]	719.8
	C	868.4[b]	698.0
3	A	3072.3	3071.1
	B	1302.1	1285.0
	C	1246.2	1248.7
4	A	2789.4	2755.9
	B	2699.8	2675.6
	C	2682.4	2653.3
5	A	2336.8	2336.9
6	A	1459.6	1464.2
	B	772.8	768.2
	C	587.5	580.6
7	A	1175.1	1165.7
	B	658.6	661.2
	C	456.3	454.0
8	A	1236.0	1166.3
	B	759.8	767.6
	C	525.9	513.0
9	A	876.0	862.5
	B	748.5	754.2
	C	513.6	513.7
10	A	3100.8	3086.2
	B	730.8	723.7
	C	687.2	685.0
11	A	1451.9	1432.1
	B	819.0	820.5
	C	687.1	679.4
12	A	1523.5	1523.2
	B	1086.4	1070.5
	C	728.2	719.9

[a] Rotational constants B_e (excluding vibrational effects) from Ref. S4 with an estimated reference error of 0.2%.
[b] A conformer other than the experimental one is obtained. This value is neglected in the statistical analysis of the data set presented in the manuscript.
1: ethynyl-cyclohexane, **2**: isoamyl-acetate,
3: diisopropyl-ketone, **4**: bicyclo[2.2.2]octadiene,
5: triethylamine, **6**: vitamin C, **7**: serotonine, **8**: aspirin,
9: cassyrane, **10**: proline, **11**: lupinene, and **12**: limonene.

Table S2: Untypically long intramolecular bonds (LB12)[a] obtained by geometry optimizations with the GFN-xTB and GFN2-xTB methods in comparison to experimental values. The values are given in pm.

system	bond	GFN2-xTB	GFN-xTB	ref.
DIAD	C–C	167.5	167.3	171.0
FLP	P–B	205.4	210.2	212.0
DTFS	Si–N	272.8	207.3	227.0
MESITRAN	Si–N	239.7	226.7	245.0
S_8^{2+}	S–S	385.6	230.2	286.0
HAPPOD	Rh–Cr	306.2	298.6[b]	308.0
KAMDOR	Os–Cr	306.6	297.6[b]	310.0
PP	C–C	313.8	312.6	312.0
BRCLNA	Br–Cl	302.8	305.8	313.0
C_2Br_6	Br–Br	337.0	340.3	342.0
RESVAN	S–S	374.9	382.6	419.0
BHS	Si–Si	454.1	452.7	443.0
	MD:	6.5 (-1.9)[c]	-13.0	–
	MAD:	19.9 (12.6)[c]	14.7	–
	SD:	35.4 (20.9)[c]	17.9	–
	MAX:	99.6 (45.8)[c]	55.8	–

[a] Reference bond lengths of long bonds as used in Ref. S5.
[b] Bonds are different w.r.t. Ref. S3 due to the modified GFN-xTB Hamiltonian with CN-dependence for Cr.
[c] Statistics without system S_8^{2+} is given in parentheses.

Table S3: Covalent bonds of heavy main group elements (HMGB11)[a] from experiment and computed with GFN2-xTB. The values are given in pm.

system	bond	GFN2-xTB	ref.
Cl_2	Cl–Cl	201.7	198.8
S_2H_2	S–S	204.5	205.5
P_2Me_4	P–P	220.1	221.2
Br_2	Br–Br	228.2	228.1
Se_2H_2	Se–Se	227.7	234.6
Ge_2H_6	Ge–Ge	236.3	241.0
As_2Me_4	As–As	244.3	242.9
Te_2Me_2	Te–Te	269.5	268.6
Sn_2Me_6	Sn–Sn	281.2	277.6
Sb_2Me_4	Sb–Sb	281.9	281.8
Pb_2Me_6	Pb–Pb	293.9	288.0
	MD:	0.1	–
	MAD:	2.6	–
	RMSD:	3.5	–
	SD:	3.6	–
	MAX:	7.0	–

[a] Reference bond lengths are the same as used in Ref. S5.

Table S4: Covalent bonds in transition metal complexes (TMC32)[a] from experiment and computed with different semiempirical methods. The values are given in pm.

	GFN2-xTB	GFN-xTB[b]	ref.
1	201.4	201.4	207.6
2	211.1	213.1	216.9
3	204.7	205.5	204.7
4	212.6	213.1	218.5
5	205.5	205.1	205.8
6	214.8	213.7	219.6
7	228.6	229.3	217.5
8	185.7	178.1	198.4
9	151.7	154.5	157.0
10	162.0	163.7	172.9
11	164.3	169.0	173.4
12	165.1	167.9	170.8
13	151.9	153.3	157.3
14	200.4	210.8	213.8
15	178.5	180.9	187.9
16	180.7	182.3	196.3
17	154.8	152.5	157.4
18	167.1	159.8	171.9
19	155.1	153.1	157.7
20	203.4	203.5	212.2
21	154.6	153.0	158.4
22	191.0	180.7	195.4
23	220.3	214.3	215.0
24	226.8	220.3	220.8
25	182.3	182.9	186.3
26	169.9	168.3	175.0
27	147.0	155.0	158.6
28	159.6	163.9	172.4
29	216.8	217.0	214.7
30	178.5	180.6	180.6
31	179.6	178.4	182.9
32	177.1	178.1	181.0
33	200.9	193.4	193.8
34	217.6	210.7	212.3
35	180.2	182.4	187.2
36	166.9	160.8	167.4
37	215.8	206.5	206.4
38	217.2	214.0	211.7
39	176.6	177.5	181.5
40	177.9	178.5	180.6
41	232.1	239.6	237.7
42	179.8	179.3	181.8
43	160.4	160.0	165.8
44	181.1	181.3	183.0
45	181.6	179.7	182.5
46	185.4	185.8	187.6
47	214.8	211.1	209.9
48	194.6	198.9	188.4
49	186.9	190.5	183.2
50	195.9	199.0	191.4
MD:	-2.8	-3.2	–
MAD:	5.7	5.0	–
SD:	6.1	5.7	–
MAX:	15.6	20.3	–

[a] Reference bond lengths are from Ref. S6.
[b] GFN-xTB values differ from Ref. S3, due to the use of a revised GFN-xTB Hamiltonian, in which 3d-level shifting is activated for the early 3d-transition metals (Sc–Cr).

Table S5: Center of mass (CMA) distances of 22 non-covalently interacting systems (S22)[S7] computed with GFN2-xTB. The values are given in pm. See Ref. S3 and its supporting information for the results obtained with other semiempirical methods.

	GFN2-xTB	ref.
1	336.4	324.6
2	282.6	291.1
3	300.9	301.1
4	316.7	325.8
5	599.7	606.0
6	501.8	517.7
7	586.0	602.6
8	372.8	371.7
9	390.5	371.8
10	378.6	371.6
11	349.3	376.4
12	328.8	347.9
13	323.0	317.6
14	331.5	349.8
15	306.4	316.7
16	423.5	442.2
17	311.7	333.7
18	321.7	353.9
19	401.5	389.2
20	501.6	490.8
21	480.8	488.7
22	524.0	494.0

1: ammonia dimer, 2: water dimer, 3: formic acid dimer,
4: formamide dimer, 5: uracil dimer,
6: 2-pyridoxine · 2-aminopyridine, 7: adenine · thymine,
8: methane dimer, 9: ethene dimer, 10: benzene · methane,
11: benzene dimer, 12: pyracine dimer, 13: uracil dimer,
14: indole · benzene, 15: adenine · thymine (stack),
16: ethene · ethine, 17: benzene · water,
18: benzene · ammonia, 19: benzene · cyanide, 20: benzene
dimer, 21: indole · benzene (T-shape), 22: phenol dimer.

Table S6: Equilibrium center-of-mass distances between non-covalently bound systems of the S66x8 set[a]. The values are given in pm and obtained by a cubic spline interpolation. For the interpolation, interaction energies and center-of-mass distances computed on the 8 structures along the potential energy curve of each complex are used. See Ref. S3 and its supporting information for the results obtained with other semiempirical methods.

#	system	GFN2-xTB	ref.
1	H_2O · H_2O	286.6	293.9
2	H_2O · MeOH	309.0	309.6
3	H_2O · $MeNH_2$	325.5	334.4
4	H_2O · peptide	383.4	384.9
5	MeOH · MeOH	351.3	350.0
6	MeOH · $MeNH_2$	330.2	335.8
7	MeOH · peptide	419.5	420.7
8	MeOH · H_2O	324.5	328.6
9	$MeNH_2$ · MeOH	349.8	354.4
10	$MeNH_2$ · $MeNH_2$	349.8	348.0
11	$MeNH_2$ · peptide	363.1	366.9
12	$MeNH_2$ · H_2O	301.2	303.4
13	peptide · MeOH	391.0	388.3
14	peptide · $MeNH_2$	376.6	388.7
15	peptide · peptide	463.8	468.1
16	peptide · H_2O	378.9	382.3
17	uracil · uracil (BP)	569.7	574.6
18	H_2O · pyridine	420.2	426.9
19	MeOH · pyridine	445.1	449.9
20	AcOH · AcOH	407.5	407.9
21	$AcNH_2$ · $AcNH_2$	423.9	432.5
22	AcOH · uracil	503.9	506.0
23	$AcNH_2$ · uracil	505.7	512.0
24	benzene · benzene (π-π)	360.8	387.6
25	pyridine · pyridine (π-π)	344.7	369.9
26	uracil · uracil (π-π)	306.3	314.8
27	benzene · pyridine (π-π)	353.5	379.2
28	benzene · uracil (π-π)	326.4	338.9
29	pyridine · uracil (π-π)	320.3	333.7
30	benzene · ethene	329.7	353.2
31	uracil · ethene	325.1	331.2
32	uracil · ethyne	317.2	326.2
33	pyridine · ethene	326.3	345.5
34	pentane · pentane	403.3	382.5
35	neopentane · pentane	467.2	452.8
36	neopentane · neopentane	522.8	525.6
37	cyclopentane · neopentane	477.5	465.7
38	cyclopentane · cyclopentane	440.2	421.7
39	benzene · cyclopentane	396.7	394.3
40	benzene · neopentane	445.3	448.5
41	uracil · pentane	355.6	353.3
42	uracil · cyclopentane	378.7	375.8
43	uracil · neopentane	432.5	434.5
44	ethene · pentane	394.9	375.6
45	ethyne · pentane	361.3	362.9
46	peptide · pentane	376.0	362.2
47	benzene · benzene (TS)	493.9	490.4
48	pyridine · pyridine (TS)	487.7	481.9
49	benzene · pyridine (TS)	491.0	487.1
50	benzene · ethyne (CH-π)	419.7	410.1
51	ethyne · ethyne (TS)	421.2	435.6
52	benzene · AcOH (OH-π)	404.7	417.3
53	benzene · $AcNH_2$ (NH-π)	469.7	476.1
54	benzene · H_2O (OH-π)	319.4	329.2
55	benzene · MeOH (OH-π)	337.8	342.1
56	benzene · $MeNH_2$ (N H-π)	358.8	358.2
57	benzene · peptide (N H-π)	401.5	404.0
58	pyridine · pyridine (N H-π)	581.2	585.7
59	ethyne · H_2O (CH-O)	404.2	399.2
60	ethyne · AcOH (OH-π)	385.7	396.4
61	pentane · AcOH	377.8	373.1
62	pentane · $AcNH_2$	362.1	358.6
63	benzene · AcOH	375.6	375.3
64	peptide · ethene	364.6	360.0
65	pyridine · ethyne	519.3	533.0
66	$MeNH_2$ · pyridine	374.7	372.5

[a] Reference structures taken from Ref. S8.

Table S7: Equilibrium center-of-mass distances between non-covalently bound systems of the S22x5 set[a]. The values are given in pm and obtained by a cubic spline interpolation. For the interpolation, interaction energies and center-of-mass distances computed on the 5 structures along the potential energy curve of each complex are used.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB-D3(BJ)	ref.
1	586.9	595.6	610.4	608.4	602.6
2	312.6	309.9	331.7	310.5	324.2
3	334.1	326.5	340.6	337.2	329.9
4	287.1	293.8	294.0	289.1	293.8
5	390.6	386.7	359.7	334.6	374.7
6	395.5	407.3	393.8	372.3	381.4
7	419.9	431.7	431.1	414.6	446.3
8	303.9	301.3	310.8	302.9	304.0
9	318.4	329.7	329.0	326.9	328.5
10	319.7	309.3	324.2	309.3	331.3
11	351.4	349.2	338.9	332.9	354.6
12	381.5	380.6	367.7	355.3	378.7
13	503.7	504.2	497.5	491.8	497.5
14	371.5	379.0	384.8	362.5	395.7
15	494.3	493.7	496.7	489.4	493.6
16	350.2	357.2	361.7	345.9	364.2
17	337.1	347.6	358.6	336.1	361.9
18	502.7	514.2	527.7	523.8	519.3
19	498.4	497.4	502.0	493.7	494.7
20	313.5	304.6	329.8	310.8	323.2
21	602.5	609.1	611.4	610.6	607.0
22	402.5	414.3	403.0	390.2	391.3

1: ammonia dimer, 2: water dimer, 3: formic acid dimer, 4: formamide dimer, 5: uracil dimer, 6: 2-pyridoxine · 2-aminopyridine, 7: adenine · thymine, 8: methane dimer, 9: ethene dimer, 10: benzene · methane, 11: benzene dimer, 12: pyracine dimer, 13: uracil dimer, 14: indole · benzene, 15: adenine · thymine (stack), 16: ethene · ethine, 17: benzene · water, 18: benzene · ammonia, 19: benzene · cyanide, 20: benzene dimer, 21: indole · benzene (T-shape), 22: phenol dimer.

Table S8: Equilibrium center-of-mass distances between non-covalently bound systems of the X40 set[S9]. The values are given in pm and obtained by a cubic spline interpolation of the X40x10 set. For the interpolation, interaction energies and center-of-mass distances computed on the 10 structures along the potential energy curve of each complex are used.

#	system	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB-D3(BJ)	ref.
1	$CH_4 \cdot F_2$	422.1	363.3	427.8	386.4	391.5
2	$CH_4 \cdot Cl_2$	437.2	425.7	388.6	452.0	431.5
3	$CH_4 \cdot Br_2$	462.4	473.1	415.2	482.7	452.6
4	$CH_4 \cdot I_2$	500.7	491.7	425.2	510.5	491.0
5	$CH_3F \cdot CH_4$	367.7	367.9	351.1	338.6	348.2
6	$CH_3Cl \cdot CH_4$	382.5	381.0	363.1	356.9	366.4
7	$CHF_3 \cdot CH_4$	337.5	319.5	338.2	318.2	345.8
8	$CHCl_3 \cdot CH_4$	345.1	328.1	319.5	342.9	360.0
9	CH_3F dimer	452.1	415.3	466.0	430.1	434.5
10	CH_3Cl dimer	520.6	503.7	477.1	525.0	516.2
11	$C_6H_3F_3 \cdot C_6H_6$	353.5	363.9	366.6	347.8	367.9
12	$C_6F_6 \cdot C_6H_6$	334.2	350.8	350.4	331.3	350.4
13	$CH_3Cl \cdot CH_2O$	424.2	393.0	367.9	349.5	395.0
14	$CH_3Br \cdot CH_2O$	372.0	351.9	366.1	343.7	371.4
15	$CH_3I \cdot CH_2O$	367.3	358.8	372.6	337.2	373.5
16	$CF_3Cl \cdot CH_2O$	447.7	435.3	450.9	426.2	452.8
17	$CF_3Br \cdot CH_2O$	407.9	417.5	434.3	420.1	430.7
18	$CF_3I \cdot CH_2O$	413.0	420.7	417.2	409.0	426.7
19	$C_6H_5Cl \cdot C(CH_3)_2O$	611.9	596.9	598.5	572.9	612.8
20	$C_6H_5Br \cdot C(CH_3)_2O$	553.3	550.6	565.5	544.4	567.4
21	$C_6H_5I \cdot C(CH_3)_2O$	544.2	544.0	551.0	524.0	554.8
22	$C_6H_5Cl \cdot N(CH_3)_3$	566.3	545.9	560.4	539.1	570.4
23	$C_6H_5Br \cdot N(CH_3)_3$	506.2	506.1	514.1	504.2	515.8
24	$C_6H_5I \cdot N(CH_3)_3$	488.9	489.2	481.8	476.4	491.0
25	$C_6H_5Br \cdot CH_3SH$	539.2	511.6	568.7	516.5	520.0
26	$C_6H_5I \cdot CH_3SH$	507.2	497.0	500.1	491.6	505.6
27	$CH_3Br \cdot C_6H_6$	403.0	386.0	341.6	388.1	395.8
28	$CH_3I \cdot C_6H_6$	402.3	378.2	344.6	381.4	398.8
29	$CF_3Br \cdot C_6H_6$	441.6	429.3	395.2	449.7	449.2
30	$CF_3I \cdot C_6H_6$	434.5	415.7	385.1	440.9	446.7
31	$CF_3OH \cdot H_2O$	333.6	338.0	345.5	340.3	339.6
32	$CCl_3OH \cdot H_2O$	349.2	355.3	357.4	358.0	353.7
33	$HF \cdot CH_3OH$	299.4	303.2	324.0	306.8	303.1
34	$HCl \cdot CH_3OH$	346.0	364.9	346.3	345.3	352.1
35	$HBr \cdot CH_3OH$	360.5	400.6	363.6	370.1	370.5
36	$HI \cdot CH_3OH$	384.0	430.6	386.8	381.0	397.6
37	$HF \cdot N(CH_3)H_2$	280.0	270.9	326.1	302.7	288.2
38	$HCl \cdot N(CH_3)H_2$	317.6	304.8	329.0	337.8	324.0
39	$CH_3OH \cdot CH_3F$	377.6	358.2	431.7	362.3	359.5
40	$CH_3OH \cdot CH_3Cl$	378.6	360.8	394.3	375.5	375.1

2.2 Non-covalent interactions

Table S9: Association energies computed with semiempirical methods for six alkane dimers (ADIM6). Structures and reference energies are taken from Ref. S10. The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	0.92	0.69	1.42	1.50	1.34
2	1.29	1.27	1.77	2.32	1.99
3	1.97	2.09	2.79	3.44	2.89
4	2.57	2.75	3.59	4.47	3.78
5	3.08	3.53	4.35	5.61	4.60
6	3.45	3.78	4.77	6.52	5.55
MD:	-1.15	-1.01	-0.24	0.62	–
MAD:	1.15	1.01	0.27	0.62	–
SD:	0.60	0.41	0.29	0.34	–
MAX:	2.10	1.77	0.78	1.01	–

Table S10: Association energies of the HAL59 set[S9–S11] computed with different semiempirical methods. Numbering and reference values taken from Ref. S10. The values are given in kcal/mol.

#	system	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	PCH–C_6H_5Br	0.53	2.32	-2.05	2.66	0.85
2	NCH–C_6H_5Br	-0.90	1.46	0.67	3.58	1.15
3	NH_3–C_6H_5Br	-0.37	2.38	2.23	5.73	2.02
4	MeI–PCH	0.98	2.34	4.19	3.17	0.85
5	MeI–NCH	0.18	1.88	0.81	5.02	1.42
6	MeI–NH_3	1.85	3.45	2.86	7.64	2.73
7	PCH–C_6H_5I	0.96	2.36	4.28	3.14	0.92
8	NCH–C_6H_5I	0.32	2.04	0.86	4.99	1.87
9	NH_3–C_6H_5I	2.07	3.78	3.06	7.60	3.33
10	PCH–F_3CI	2.65	2.33	5.36	3.59	0.89
11	NCH–F_3CI	3.63	3.27	3.37	6.76	3.61
12	NH_3–F_3CI	8.22	6.88	7.97	10.03	5.88
13	Br_2–PCH	1.58	2.36	-3.07	3.53	1.18
14	Br_2–NCH	1.66	2.25	2.52	6.00	3.61
15	Br_2–NH_3	6.89	5.97	7.65	9.78	7.29
16	PCH–$C_4H_4BrNO_2$	1.87	2.71	-0.25	4.63	1.19
17	NCH–$C_4H_4BrNO_2$	1.79	3.84	14.21	8.66	4.32
18	NH_3–$C_4H_4BrNO_2$	7.24	10.23	36.83	13.40	8.02
19	PCH–FCl	2.43	3.25	-6.54	4.84	1.16
20	NCH–FCl	3.07	3.03	-3.94	7.99	4.81
21	NH_3–FCl	9.65	11.85	-69.17	11.15	10.54
22	PCH–$C_4H_4INO_2$	3.82	3.38	29.22	5.81	1.53
23	NCH–$C_4H_4INO_2$	5.59	5.56	83.24	12.29	5.91
24	NH_3–$C_4H_4INO_2$	12.13	13.61	182.55	18.43	10.99
25	PCH–FBr	4.22	3.55	-6.58	5.49	2.07
26	NCH–FBr	5.91	5.05	-2.51	9.55	7.53
27	NH_3–FBr	14.89	13.63	-2.23	10.41	15.30
28	FI–PCH	5.29	5.58	16.83	7.28	2.74
29	FI–NCH	8.35	7.67	4.34	13.42	9.33
30	FI–NH_3	15.09	17.10	4.92	17.11	17.11
31	MeI–FCCH	0.46	0.34	0.18	1.53	0.50
32	Br_2–FCCH	0.74	0.33	-0.23	1.33	0.74
33	FI–FCCH	-0.21	-0.40	-0.01	-0.08	0.29
34	MeI–FMe	0.30	1.09	1.38	3.06	1.70
35	Br_2–FMe	1.18	0.87	-0.19	3.16	2.87
36	FI–FMe	3.86	7.14	4.20	7.35	5.97
37	MeI–OCH_2	1.46	2.84	2.13	5.51	2.39
38	Br_2–OCH_2	3.70	3.27	3.25	7.28	4.41
39	FI–OCH_2	9.71	9.48	4.39	14.30	9.94
40	MeI–OPH_3	1.32	2.70	8.33	4.65	3.34
41	Br_2–OPH_3	1.92	1.15	3.90	5.64	5.95
42	FI–OPH_3	6.92	1.93	17.55	11.79	13.36
43	MeI–C_5H_5N	2.76	3.92	2.92	10.84	3.61
44	Br_2–C_5H_5N	9.31	5.12	4.49	14.94	9.07
45	FI–C_5H_5N	20.07	18.52	-9.50	25.48	20.34
46	$C_6H_3F_3$-C_6H_6	4.99	3.50	3.98	5.09	4.40
47	$C_6H_6F_6$-C_6H_6	6.70	3.96	4.97	6.91	6.12
48	C_6H_5Cl–CH_2O	0.65	0.47	1.27	4.65	1.49
49	C_6H_5Br–CH_2O	0.60	2.40	2.54	5.47	2.43
50	C_6H_5I–CH_2O	2.92	3.89	3.38	7.35	3.46
51	C_6H_5Cl–$N(CH_3)_3$	0.94	0.92	1.62	7.71	2.11
52	C_6H_5Br–$N(CH_3)_3$	1.21	3.76	3.92	9.27	3.78
53	C_6H_5I–$N(CH_3)_3$	4.07	5.28	5.74	12.24	5.81
54	C_6H_5Br–CH_3SH	1.04	2.50	0.69	3.33	2.32
55	C_6H_5I–CH_3SH	1.84	3.24	3.10	4.81	3.08
56	CH_3Br–C_6H_6	0.89	1.21	2.63	3.27	1.81
57	CH_3I–C_6H_6	1.69	1.60	3.50	3.83	2.48
58	CF_3Br–C_6H_6	1.82	1.57	3.65	4.28	3.11
59	CF_3I–C_6H_6	3.54	2.13	5.03	4.85	3.91
	MD:	-0.73	-0.36	2.36	2.76	–
	MAD:	1.28	1.34	9.71	3.00	–
	SD:	1.51	2.09	27.92	2.22	–
	MAX:	6.44	11.43	171.56	7.44	–

Table S11: Association energies of the PNICO23 set[S10,S12] computed with different semiempirical methods. Numering and refence values taken from Ref. S10. The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	2.19	1.86	0.39	1.67	1.43
2	5.96	18.56	-10.70	-1.86	8.02
3	3.92	1.87	-0.90	-0.61	0.64
4	4.37	9.41	-7.50	4.63	4.26
5	4.96	3.10	-0.55	3.66	2.86
6	1.18	1.29	1.52	1.38	1.32
7	3.37	7.62	-7.87	4.36	4.29
8	2.62	4.12	-2.37	3.08	2.63
9	6.24	7.34	-6.22	4.05	4.91
10	2.89	2.99	-0.55	2.23	2.21
11	2.44	2.56	-0.17	1.57	1.40
12	4.22	6.03	-4.01	2.88	2.62
13	1.78	1.27	2.25	0.44	3.98
14	4.18	6.80	-4.46	4.02	4.10
15	2.91	2.99	3.54	2.25	4.34
16	2.16	1.97	1.49	1.41	1.78
17	7.22	10.94	5.51	5.43	7.10
18	2.24	1.50	1.24	1.27	2.35
19	5.12	7.70	5.28	3.92	5.95
20	11.47	14.70	9.64	6.16	8.18
21	5.53	6.02	3.83	4.02	4.92
22	8.88	9.02	5.87	6.04	8.03
23	12.41	12.37	8.34	7.88	10.97
MD:	0.43	1.90	-4.12	-1.23	–
MAD:	1.10	2.33	4.26	1.45	–
SD:	1.42	2.79	5.05	2.23	–
MAX:	3.29	10.54	18.72	9.88	–

Table S12: Association energies of 22 non-covalently interacting systems (S22)[a] computed with GFN2-xTB. The values are given in kcal/mol. Reference energies taken from Ref. S13. Structures taken from Ref. S7. Running number as in Ref. S14.

system #	GFN2-xTB	ref.
1	2.05	3.13
2	4.90	4.99
3	17.20	18.75
4	16.64	16.06
5	19.63	20.64
6	16.72	16.93
7	15.89	16.66
8	0.38	0.53
9	1.07	1.47
10	1.29	1.45
11	3.83	2.65
12	5.43	4.25
13	9.81	9.80
14	5.40	4.52
15	12.20	11.73
16	1.37	1.50
17	2.26	3.27
18	1.88	2.31
19	2.60	4.54
20	2.31	2.72
21	4.00	5.63
22	5.79	7.10
MD:	-0.36	–
MAD:	0.75	–
SD:	0.88	–
MAX:	1.94	–

1: ammonia dimer, 2: water dimer, 3: formic acid dimer, 4: formamide dimer, 5: uracil dimer, 6: 2-pyridoxine · 2-aminopyridine, 7: adenine · thymine, 8: methane dimer, 9: ethene dimer, 10: benzene · methane, 11: benzene dimer, 12: pyracine dimer, 13: uracil dimer, 14: indole · benzene, 15: adenine · thymine (stack), 16: ethene · ethine, 17: benzene · water, 18: benzene · ammonia, 19: benzene · cyanide, 20: benzene dimer, 21: indole · benzene (T-shape), 22: phenol dimer.

Table S13: Association energies computed with semiempirical methods for 66 non-covalent complexes consisting of main group elements (S66).[S8] The values are given in kcal/mol.

#	system	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	$H_2O \cdot H_2O$	4.91	4.76	5.03	4.86	4.92
2	$H_2O \cdot$ MeOH	4.78	4.46	5.58	5.09	5.59
3	$H_2O \cdot MeNH_2$	5.38	6.15	7.44	4.42	6.91
4	$H_2O \cdot$ peptide	7.40	7.08	7.99	8.41	8.10
5	MeOH · MeOH	4.77	4.62	6.32	4.85	5.76
6	MeOH · $MeNH_2$	5.74	6.65	6.98	4.48	7.55
7	MeOH · peptide	7.31	7.28	8.19	7.82	8.23
8	MeOH · H_2O	4.61	4.76	5.75	4.43	5.01
9	$MeNH_2$ · MeOH	2.71	2.51	4.24	2.54	3.06
10	$MeNH_2 \cdot MeNH_2$	2.70	2.56	4.54	2.36	4.16
11	$MeNH_2 \cdot$ peptide	4.58	4.22	6.14	4.32	5.42
12	$MeNH_2 \cdot H_2O$	4.98	5.71	7.30	4.09	7.27
13	peptide · MeOH	5.32	4.85	6.49	5.07	6.19
14	peptide · $MeNH_2$	6.39	5.76	7.45	4.72	7.45
15	peptide · peptide	8.05	7.54	8.81	7.92	8.63
16	peptide · H_2O	4.76	4.78	5.51	4.47	5.12
17	uracil · uracil (BP)	16.54	14.77	16.44	15.11	17.18
18	$H_2O \cdot$ pyridine	5.11	5.44	6.78	4.24	6.86
19	MeOH · pyridine	5.57	5.96	6.14	4.31	7.41
20	AcOH · AcOH	17.56	18.68	18.43	19.15	19.09
21	$AcNH_2 \cdot AcNH_2$	16.43	14.45	16.98	15.49	16.26
22	AcOH · uracil	18.24	17.66	18.42	18.40	19.49
23	$AcNH_2 \cdot$ uracil	19.07	16.71	18.97	17.48	19.19
24	benzene · benzene (π-π)	3.60	3.12	2.89	3.38	2.82
25	pyridine · pyridine (π-π)	4.60	3.35	4.03	3.96	3.90
26	uracil · uracil (π-π)	9.87	8.62	9.08	9.25	9.83
27	benzene · pyridine (π-π)	4.06	3.15	3.55	3.71	3.44
28	benzene · uracil (π-π)	5.81	3.97	5.37	6.02	5.71
29	pyridine · uracil (π-π)	7.39	5.22	6.88	6.92	6.82
30	benzene · ethene	1.88	1.78	1.59	1.89	1.43
31	uracil · ethene	3.14	1.96	3.00	3.60	3.38
32	uracil · ethyne	3.34	1.84	2.70	3.84	3.74
33	pyridine · ethene	2.18	1.81	2.03	2.08	1.87
34	pentane · pentane	2.38	2.56	3.48	4.73	3.78
35	neopentane · pentane	1.93	2.10	2.73	3.45	2.61
36	neopentane · neopentane	1.62	2.08	2.17	2.65	1.78
37	cyclopentane · neopentane	1.95	2.09	2.52	3.39	2.40
38	cyclopentane · cyclopentane	1.96	2.10	2.33	3.67	3.00

39	benzene · cyclopentane	3.27	3.03	3.40	3.73	3.58
40	benzene · neopentane	2.88	2.73	3.46	3.19	2.90
41	uracil · pentane	4.93	4.09	5.55	5.61	4.85
42	uracil · cyclopentane	4.01	3.59	4.53	4.71	4.14
43	uracil · neopentane	3.81	2.81	3.78	3.79	3.71
44	ethene · pentane	1.34	1.15	1.85	2.46	2.01
45	ethyne · pentane	1.60	1.43	1.73	1.91	1.75
46	peptide · pentane	3.21	3.22	4.11	4.72	4.26
47	benzene · benzene (TS)	2.55	2.05	2.52	2.70	2.88
48	pyridine · pyridine (TS)	2.75	2.44	3.03	3.07	3.54
49	benzene · pyridine (TS)	2.65	1.99	2.92	2.86	3.33
50	benzene · ethyne (CH-π)	2.04	0.93	1.96	2.52	2.87
51	ethyne · ethyne (TS)	1.48	0.63	0.82	1.28	1.52
52	benzene · AcOH (OH-π)	3.65	3.08	4.17	4.60	4.71
53	benzene · $AcNH_2$ (NH-π)	3.26	3.43	3.92	3.79	4.36
54	benzene · H_2O (OH-π)	2.29	1.71	3.36	3.18	3.28
55	benzene · MeOH (OH-π)	3.17	2.49	3.63	3.80	4.19
56	benzene · $MeNH_2$ (N H-π)	2.66	2.00	3.23	2.96	3.23
57	benzene · peptide (N H-π)	4.08	3.14	4.81	4.81	5.28
58	pyridine · pyridine (N H-π)	3.09	2.59	3.25	2.05	4.15
59	ethyne · H_2O (CH-O)	1.84	1.11	1.95	2.68	2.85
60	ethyne · AcOH (OH-π)	3.95	3.51	2.38	4.22	4.87
61	pentane · AcOH	2.80	2.46	3.59	3.52	2.91
62	pentane · $AcNH_2$	3.08	3.28	3.94	4.15	3.53
63	benzene · AcOH	2.98	2.35	3.73	3.79	3.80
64	peptide · ethene	2.07	2.15	2.74	2.91	3.00
65	pyridine · ethyne	3.35	2.18	1.78	2.85	3.99
66	$MeNH_2$ · pyridine	2.94	2.68	4.43	2.77	3.97
	MD:	-0.61	-1.05	-0.15	-0.45	–
	MAD:	0.73	1.08	0.47	0.79	–
	SD:	0.65	0.64	0.64	1.04	–
	MAX:	2.29	2.48	2.49	3.18	–

Table S14: Association energies of water clusters (WATER27 set) computed with different semiempirical methods. The values are given in kcal/mol.

system	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
H_2O_2	4.93	4.79	5.19	4.87	5.01
H_2O_3	14.26	14.63	15.00	14.75	15.80
H_2O_4	26.46	24.29	26.25	25.93	27.40
H_2O_5	34.20	30.99	34.01	33.77	35.90
H_2O_6	44.57	45.33	45.68	44.90	46.00
H_2O_6c	44.16	43.31	44.98	43.94	45.80
H_2O_6b	43.44	40.71	43.61	42.92	45.30
H_2O_6c2	42.09	37.80	41.33	42.10	44.30
H_2O_8d2d	71.48	68.18	72.06	69.92	72.60
H_2O_8s4	71.35	68.21	71.85	69.76	72.60
H_2O_{20}	193.19	176.71	195.88	186.12	198.60
H_2O_{20}fc	208.95	201.24	207.22	204.39	208.00
H_2O_{20}fs	207.71	197.38	206.93	202.69	208.00
H_2O_{20}es	208.62	196.93	206.49	203.25	209.70
$H_3O^+H_2O$	31.16	33.19	28.46	30.55	33.50
$H_3O^+H_2O_2$	53.46	54.34	49.88	52.42	56.90
$H_3O^+H_2O_3$	73.30	72.59	68.15	71.31	76.50
$H_3O^+H_2O_6$3d	112.67	112.76	110.37	108.52	117.80
$H_3O^+H_2O_6$2d	110.36	109.80	108.86	108.74	114.90
OH^-H_2O	35.35	38.90	35.08	30.33	26.60
$OH^-H_2O_2$	55.32	60.02	54.36	53.04	48.40
$OH^-H_2O_3$	74.70	79.30	73.91	72.51	67.60
$OH^-H_2O_4$c4	89.98	96.28	88.20	91.12	84.80
$OH^-H_2O_4$cs	89.72	97.25	90.37	90.85	84.80
$OH^-H_2O_5$	104.43	114.86	103.27	108.40	100.70
$OH^-H_2O_6$	120.02	129.04	119.01	122.43	115.70
$H_3O^+H_2O_6OH^-$	32.41	42.88	21.09	30.96	28.50
MD:	0.24	0.00	-0.90	-1.16	–
MAD:	3.15	7.51	3.55	4.31	–
SD:	3.92	9.43	4.45	5.07	–
MAX:	8.75	21.89	8.47	12.47	–

Table S15: Association energies of anion-neutral dimer systems (AHB21).[S17] The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	-13.45	-16.77	-9.44	-15.41	-17.79
2	-26.50	-33.06	-24.13	-30.73	-32.50
3	-58.43	-67.14	-45.41	-61.41	-65.68
4	-7.16	-5.15	-8.59	-8.62	-8.98
5	-13.08	-11.39	-13.77	-17.17	-15.61
6	-24.78	-20.75	-17.38	-28.73	-25.52
7	-18.15	-5.60	-14.37	-19.45	-14.35
8	-37.75	-35.11	-46.30	-42.37	-41.79
9	-20.48	-21.40	-20.08	-17.63	-17.03
10	-43.90	-48.37	-39.26	-40.72	-37.31
11	-7.60	-8.13	-9.18	-7.13	-7.97
12	-14.09	-16.62	-14.84	-13.67	-14.13
13	-29.96	-34.22	-10.45	-24.79	-26.01
14	-17.61	-10.74	-11.14	-12.90	-11.07
15	-7.62	-14.19	-8.85	-7.83	-8.62
16	-13.98	-26.50	-13.87	-15.26	-15.73
17	-26.51	-45.21	-15.95	-25.70	-26.24
18	-10.46	-11.54	-14.08	-9.95	-12.80
19	-18.19	-20.55	-20.50	-19.37	-20.65
20	-18.93	-22.50	-22.86	-23.65	-21.03
21	-32.41	-33.61	-21.66	-33.11	-31.40
MD:	0.53	-1.73	3.34	-0.16	–
MAD:	2.97	4.68	4.75	1.80	–
SD:	3.72	6.51	6.61	2.28	–
MAX:	7.25	18.97	20.27	5.10	–

Table S16: Association energies of the systems in the CARBHB12.[S10] The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	ref.
1	3.34	4.65	4.74	5.37
2	4.07	5.27	3.56	6.05
3	2.82	3.43	2.77	2.42
4	7.25	10.21	10.86	9.97
5	1.83	2.08	2.63	2.36
6	2.39	2.48	2.39	3.02
7	1.47	1.75	1.46	1.21
8	3.32	4.89	4.43	4.18
9	7.53	7.48	8.30	7.84
10	9.54	10.54	6.81	10.48
11	5.57	5.36	3.65	3.24
12	16.33	16.98	13.51	16.30
MD:	-0.58	0.23	-0.61	–
MAD:	1.08	0.67	1.09	–
SD:	1.33	0.85	1.52	–
MAX:	2.71	2.12	3.67	–

Table S17: Association energies of cation-neutral dimer systems (CHB6).[S17] The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	ref.
1	-44.31	-42.80	-20.29	-34.43
2	-30.07	-23.15	-11.59	-23.83
3	-23.84	-19.33	-7.45	-17.83
4	-36.74	-40.77	-12.23	-39.09
5	-28.28	-14.76	-29.04	-25.63
6	-24.61	-20.47	-20.35	-19.90
MD:	-4.52	-0.10	9.96	–
MAD:	5.31	3.95	11.25	–
SD:	4.12	6.23	10.92	–
MAX:	9.88	10.87	26.86	–

Table S18: Association energies of the HEAVY28[S10,S18] set computed with different semiempirical methods. The values are given in kcal/mol.

#	system	GFN2-xTB	GFN-xTB	PM6-D3H4X	ref.
1	$(BiH_3)_2$	1.99	1.36	0.60	1.16
2	$BiH_3{-}H_2O$	1.85	1.49	-1.73	2.49
3	$BiH_3{-}H_2S$	1.41	1.18	0.60	1.36
4	$BiH_3{-}HCl$	1.06	0.78	-0.08	0.77
5	$BiH_3{-}HBr$	1.13	0.94	0.69	0.98
6	$BiH_3{-}HI$	1.04	1.69	4.50	1.30
7	$BiH_3{-}NH_3$	1.69	1.13	-2.14	0.60
8	$(PbH_4)_2$	0.94	0.03	2.62	1.25
9	$PbH_4{-}BiH_3$	0.57	0.44	0.32	0.55
10	$PbH_4{-}H_2O$	0.82	0.35	0.87	0.36
11	$PbH_4{-}HCl$	0.64	0.41	0.68	0.75
12	$PbH_4{-}HBr$	0.59	0.33	-0.16	0.93
13	$PbH_4{-}HI$	0.65	0.53	1.34	1.18
14	$PbH_4{-}TeH_2$	0.61	0.46	0.60	0.65
15	$(SbH_3)_2$	2.19	4.04	-1.61	1.28
16	$SbH_3{-}H_2O$	0.80	1.97	-2.25	1.57
17	$SbH_3{-}H_2S$	0.83	1.41	-0.55	1.06
18	$SbH_3{-}HCl$	1.01	2.02	-2.92	2.02
19	$SbH_3{-}HBr$	0.66	1.61	-1.96	1.89
20	$SbH_3{-}HI$	0.92	2.00	-0.98	1.49
21	$SbH_3{-}NH_3$	5.41	8.45	-29.18	2.84
22	$(TeH_2)_2$	0.64	0.80	1.85	0.52
23	$TeH_2{-}H_2O$	2.46	0.40	-1.45	0.68
24	$TeH_2{-}H_2S$	0.94	0.41	0.03	0.48
25	$TeH_2{-}HCl$	1.92	0.89	0.64	1.23
26	$TeH_2{-}HBr$	1.26	0.76	0.11	1.22
27	$TeH_2{-}HI$	0.40	1.25	1.81	0.80
28	$TeH_2{-}NH_3$	4.46	2.33	2.18	3.35
	MD:	0.15	0.17	-2.15	–
	MAD:	0.61	0.65	2.70	–
	SD:	0.83	1.29	6.14	–
	MAX:	2.57	5.61	32.02	–

Table S19: Association energies of the anion-cation dimer set (IL16)[S17] computed with different semiempirical methods. The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	-99.31	-94.08	-95.10	-103.22	-100.41
2	-113.75	-120.68	-115.46	-120.99	-120.80
3	-108.50	-115.22	-109.75	-115.42	-116.91
4	-99.47	-94.89	-97.81	-103.72	-105.01
5	-96.87	-100.54	-99.07	-96.85	-104.44
6	-83.17	-94.93	-80.78	-85.80	-87.42
7	-109.04	-104.27	-102.58	-113.63	-114.00
8	-109.28	-104.49	-103.83	-114.36	-113.51
9	-111.65	-106.41	-106.47	-116.92	-114.91
10	-107.51	-112.63	-112.70	-109.00	-112.75
11	-101.14	-116.68	-95.87	-100.03	-104.47
12	-114.78	-110.39	-109.14	-121.29	-118.19
13	-106.28	-112.91	-110.75	-110.38	-112.02
14	-102.33	-118.59	-99.00	-104.47	-106.53
15	-111.55	-110.93	-113.31	-108.70	-110.98
16	-102.18	-103.41	-95.97	-98.54	-102.37
MD:	4.24	1.48	6.07	1.34	–
MAD:	4.32	5.69	6.36	2.46	–
SD:	2.50	7.25	3.64	2.81	–
MAX:	8.41	12.21	11.43	7.59	–

Table S20: Association energies of rare gas complexes (RG18)[S10] computed with different semiempirical methods. The values are given in kcal/mol.

#	system	GFN2-xTB	GFN-xTB	PM6-D3H4X	ref.
1	Ne_2	0.05	0.03	0.06	0.08
2	Ar_2	0.23	0.17	0.01	0.27
3	Kr_2	0.40	0.74	-0.52	0.40
4	Ne_3	0.15	0.09	0.18	0.27
5	Ar_3	0.66	0.46	0.34	0.77
6	Kr_3	1.11	1.90	-0.48	1.18
7	Ne_4	0.30	0.18	0.36	0.54
8	Ar_4	1.30	0.91	0.71	1.51
9	Ne_6	0.64	0.40	0.76	1.13
10	HFNe	0.27	0.32	-0.29	0.23
11	HFAr	0.95	1.04	-0.10	0.59
12	HFKr	0.76	1.86	1.70	0.72
13	C_2H_2−Ne	0.14	0.10	0.16	0.12
14	C_2H_2−Ar	0.38	0.25	-0.11	0.33
15	C_2H_6−Ne	0.16	0.17	0.26	0.24
16	C_2H_6−Ar	0.47	0.47	0.12	0.54
17	Benzene–Ne	0.38	0.27	0.40	0.40
18	Benzene–Ar	1.13	0.73	-1.24	1.12
	MD:	-0.05	-0.02	-0.45	–
	MAD:	0.11	0.32	0.57	–
	SD:	0.17	0.45	0.71	–
	MAX:	0.49	1.14	2.36	–

Table S21: Association energies computed with semiempirical methods for 30 large non-covalent complexes containing only main group elements (S30L)[a]. The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	-25.93	-21.97	-28.83	-28.62	-29.04
2	-18.02	-14.51	-18.63	-19.88	-20.78
3	-22.50	-18.75	-22.41	-23.93	-23.54
4	-21.72	-16.52	-19.18	-18.80	-20.27
5	-33.88	-27.16	-33.96	-34.07	-28.99
6	-25.57	-18.70	-20.83	-24.61	-25.50
7	-42.20	-34.75	-30.95	-38.78	-35.06
8	-48.70	-39.87	-35.57	-44.03	-36.79
9	-34.83	-33.02	-27.66	-36.37	-28.38
10	-35.86	-34.88	-29.12	-38.03	-29.78
11	-41.74	-41.52	-38.69	-44.49	-32.95
12	-42.21	-41.72	-38.45	-44.33	-33.92
13	-22.30	-25.25	-29.32	-26.75	-30.83
14	-25.64	-26.89	-29.45	-30.64	-31.33
15	-24.10	-22.78	-21.86	-36.63	-17.39
16	-25.85	-29.95	-30.06	-42.40	-25.12
17	-26.78	-25.18	-39.73	-31.14	-33.38
18	-20.38	-20.49	-29.11	-23.09	-23.31
19	-13.05	-17.85	-19.32	-17.15	-17.47
20	-15.23	-22.00	-23.48	-22.62	-19.25
21	-22.14	-28.43	-31.35	-28.23	-24.21
22	-36.57	-33.98	-44.06	-33.76	-42.63
23	-60.72	-47.03	-61.72	-41.70	-61.32
24	-136.59	-167.06	-162.49	-162.06	-135.51
25	-28.08	-24.30	-25.95	-29.39	-25.96
26	-28.21	-24.60	-25.89	-29.49	-25.77
27	-83.16	-94.76	-104.03	-95.17	-82.18
28	-79.49	-90.49	-101.22	-89.88	-80.11
29	-50.95	-54.68	-59.62	-58.39	-53.54
30	-50.52	-51.57	-56.39	-56.78	-49.28
MD:	-0.64	-0.90	-3.86	-4.25	–
MAD:	4.05	6.08	5.15	6.90	–
SD:	5.09	8.50	7.47	8.68	–
MAX:	11.91	31.55	26.98	26.55	–

[a] Reference structures and numbering are taken from Ref. S19.

2.3 Conformers

Table S22: Relative conformer energies for different alkane conformers (ACONF)[a] computed with GFN2-xTB. The results for the other semiempirical methods mentioned in the manuscript can be found in Ref. S3. The values are given in kcal/mol.

	GFN2-xTB	ref.
1	0.604	0.598
2	0.566	0.614
3	1.214	0.961
4	2.483	2.813
5	0.573	0.595
6	0.545	0.604
7	1.176	0.934
8	1.082	1.178
9	1.233	1.302
10	1.741	1.250
11	2.384	2.632
12	2.432	2.740
13	2.862	3.283
14	3.049	3.083
15	4.647	4.925
MD:	-0.061	–
MAD:	0.194	–
SD:	0.246	–
MAX:	0.491	–

[a] Reference data taken from Ref. S20.

Table S23: Conformational energies for different amino acid conformers (Amino20x4)[a] computed with different semiempirical methods. The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	-0.01	-0.49	-2.15	-0.58	1.17
2	1.47	1.91	0.95	1.60	3.05
3	1.19	1.42	0.56	1.25	3.32
4	5.55	3.90	5.08	4.35	5.04
5	5.18	3.67	6.32	1.81	1.58
6	5.64	3.67	4.66	0.97	2.53
7	7.21	6.02	5.32	3.03	2.80
8	6.28	6.04	5.08	3.79	6.46
9	0.90	1.24	0.99	1.38	0.39
10	4.02	4.42	1.99	4.77	4.11
11	3.79	4.35	3.27	4.57	4.76
12	7.83	8.17	7.72	8.70	6.53
13	-0.03	-0.83	-0.92	1.44	0.16
14	1.64	1.98	0.67	1.68	1.97
15	3.69	2.36	4.54	2.45	2.90
16	1.51	0.11	0.35	3.14	3.11
17	-0.83	-0.63	-3.41	-1.67	0.28
18	0.60	0.17	-0.71	0.31	0.98
19	-0.10	0.04	-1.93	-1.15	0.99
20	2.02	1.84	0.13	0.17	2.37
21	0.28	1.44	-0.72	-0.32	0.42
22	1.93	3.47	0.42	2.56	3.27
23	5.39	6.82	4.84	5.21	4.04
24	4.99	5.78	2.34	5.12	4.15
25	1.78	0.85	2.21	2.02	1.33
26	2.08	1.52	2.42	2.28	1.54
27	3.29	2.57	3.09	3.27	2.94
28	4.63	3.27	4.57	3.58	5.22
29	1.82	2.48	3.88	1.10	1.09
30	3.54	3.24	4.62	3.34	2.63
31	2.14	3.31	4.29	0.88	2.74
32	3.12	3.58	4.44	3.12	4.09
33	2.22	2.18	2.69	2.10	2.77
34	2.29	1.82	2.49	2.02	2.99
35	6.19	5.93	9.22	6.36	7.32
36	7.69	6.45	8.13	7.10	7.37
37	0.02	0.12	-0.16	0.49	0.19
38	0.10	0.06	0.16	0.32	0.59

39	1.07	0.88	2.34	0.96	0.96
40	1.33	1.54	1.62	1.01	0.99
41	-0.71	-0.24	-0.43	-0.42	0.34
42	0.29	0.02	-0.06	-0.16	1.52
43	-0.72	0.08	-0.45	-0.28	1.69
44	1.01	0.94	-0.57	0.55	1.95
45	-0.98	-0.98	-2.20	-0.14	0.06
46	0.29	0.13	0.37	0.12	0.20
47	0.32	0.09	0.17	0.01	0.46
48	-0.12	-0.69	-1.56	0.07	0.54
49	1.02	-0.17	-0.06	0.90	1.81
50	1.96	1.78	0.13	1.48	2.46
51	3.80	2.26	2.54	2.32	2.47
52	2.60	1.78	0.70	1.15	2.90
53	-0.33	-0.91	-0.47	-0.02	0.87
54	0.15	-0.52	-0.77	-0.01	1.72
55	1.90	0.39	2.47	1.81	1.84
56	1.20	0.51	2.29	1.19	1.89
57	1.18	1.59	0.81	1.36	1.40
58	4.05	4.44	3.19	4.69	3.21
59	4.42	5.02	3.24	4.90	4.19
60	6.51	7.67	6.16	7.05	6.01
61	3.40	3.45	3.18	2.68	3.03
62	2.68	2.99	2.99	2.09	3.10
63	3.85	3.77	3.76	3.71	3.51
64	2.17	3.23	3.55	1.98	4.18
65	2.14	0.81	3.44	2.53	1.34
66	0.95	0.27	1.05	1.37	3.08
67	2.18	2.11	3.57	2.89	3.51
68	2.38	1.29	2.12	1.43	4.22
69	-0.86	-0.94	0.46	-0.57	1.29
70	0.42	0.62	0.70	0.52	2.83
71	1.11	0.53	3.22	0.95	3.24
72	1.45	1.28	2.25	1.69	4.06
73	0.26	0.18	0.18	0.23	0.09
74	0.07	-0.64	-0.12	0.24	0.90
75	1.81	0.32	2.47	1.80	1.71
76	0.26	-0.46	-0.68	0.10	1.77
77	0.35	0.07	0.58	-0.29	0.85
78	1.27	0.85	2.70	0.72	0.86
79	1.54	0.65	0.99	0.38	1.35

80	1.26	1.11	-0.42	0.60	1.48
MD:	-0.31	-0.55	-0.53	-0.61	–
MAD:	0.95	1.11	1.37	1.00	–
SD:	1.24	1.28	1.62	1.09	–
MAX:	4.41	3.22	4.74	2.80	–

[a] Reference data taken from Ref. S21.

Table S24: Conformational energies of butane-1,4-diol conformers (BUT14DIOL)[S10,S22] computed with different semiempirical methods. The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	-1.07	-0.79	-1.94	-0.41	0.15
2	0.81	0.03	0.75	0.13	0.30
3	0.94	0.08	-0.07	1.02	1.31
4	1.73	0.76	1.19	0.45	1.44
5	0.29	0.16	-1.33	-0.24	1.72
6	-0.05	-0.31	0.53	1.04	2.07
7	3.29	2.65	5.42	2.65	1.77
8	0.34	0.78	1.01	1.85	2.25
9	1.89	1.76	3.29	2.26	2.03
10	2.86	1.84	4.10	2.29	2.10
11	4.72	2.92	6.57	2.98	2.00
12	1.51	1.75	3.20	2.07	2.15
13	0.06	0.64	0.74	1.72	2.22
14	1.49	1.02	1.89	1.91	2.42
15	1.81	2.18	3.30	2.05	2.23
16	2.27	2.20	3.66	2.48	2.25
17	0.55	1.32	1.23	1.77	2.58
18	1.06	1.34	1.60	2.18	2.59
19	1.43	1.70	1.99	2.35	2.63
20	2.05	2.58	3.49	2.08	2.48
21	0.60	1.46	1.44	1.82	2.74
22	2.24	2.53	3.84	2.29	2.55
23	4.01	3.56	6.64	3.25	2.53
24	0.68	1.70	1.17	1.66	2.72
25	1.17	1.70	1.58	2.02	2.69
26	3.27	3.61	5.25	2.13	2.49
27	2.45	2.62	4.28	2.66	2.72
28	1.21	1.75	1.93	2.23	2.83

29	1.19	1.49	1.85	2.21	2.85
30	1.19	1.49	1.85	2.21	2.85
31	3.60	3.55	5.88	2.65	2.52
32	2.05	2.69	3.22	1.91	2.63
33	1.87	1.74	2.21	2.24	3.10
34	1.10	1.74	1.62	1.98	2.72
35	3.48	2.73	4.62	2.61	2.83
36	0.81	1.78	1.28	1.71	2.79
37	2.18	2.62	3.71	2.26	2.79
38	1.08	1.54	1.50	1.81	3.06
39	3.74	2.81	4.97	2.56	3.10
40	2.26	2.05	2.94	2.32	3.30
41	4.39	3.16	6.05	3.60	3.15
42	2.63	1.96	3.44	2.18	3.29
43	1.04	1.20	1.20	2.22	3.59
44	2.66	2.87	3.88	2.36	3.18
45	2.66	2.87	3.88	2.36	3.18
46	1.37	2.11	1.71	1.86	3.37
47	2.26	2.43	2.49	2.31	3.45
48	2.71	3.14	3.67	2.12	3.33
49	2.71	3.14	3.67	2.12	3.33
50	1.60	2.06	2.23	2.20	3.37
51	1.31	1.36	1.62	2.11	3.61
52	2.83	2.11	3.46	2.52	3.42
53	4.60	4.12	6.63	3.08	3.15
54	2.37	3.03	3.25	2.04	3.31
55	1.22	2.11	1.40	1.81	3.45
56	3.83	4.14	5.31	2.50	3.32
57	1.27	2.19	1.42	1.90	3.57
58	3.34	2.56	4.55	3.38	3.52
59	2.57	3.20	3.35	2.23	3.50
60	1.23	2.21	1.34	1.98	3.65
61	1.98	2.39	1.91	2.25	3.78
62	3.30	3.15	3.77	3.01	4.15
63	2.13	2.22	2.01	2.69	4.31
64	2.41	2.57	2.40	2.96	4.70
MD:	-0.82	-0.74	-0.03	-0.69	–
MAD:	1.25	0.95	1.48	0.81	–
SD:	1.19	0.87	1.79	0.65	–
MAX:	2.72	2.39	4.57	1.96	–

Table S25: Conformational energies of small inorganic molecules (ICONF)[S10] computed with different semiempirical methods. The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	0.67	1.09	-1.04	0.85	0.90
2	5.94	5.32	9.41	3.49	5.29
3	1.65	-2.14	-1.77	1.45	0.13
4	4.26	0.61	3.37	3.52	2.33
5	10.41	10.25	4.34	7.78	12.16
6	-0.10	-0.55	0.13	0.00	0.10
7	0.15	0.37	0.72	0.00	1.03
8	0.56	-0.24	2.89	0.00	3.51
9	1.39	2.23	1.16	0.00	1.69
10	-2.88	-2.67	6.83	0.45	1.40
11	2.63	2.11	4.33	1.17	4.39
12	7.50	7.55	9.79	4.01	9.16
13	1.24	0.59	0.73	0.15	0.55
14	6.53	2.92	2.81	0.86	3.55
15	-1.49	-2.33	-3.12	1.52	1.33
16	1.72	-5.35	-7.94	-1.52	3.66
17	3.19	-7.29	-7.42	-2.35	4.35
MD:	-0.72	-2.53	-1.78	-2.01	–
MAD:	1.63	2.63	3.13	2.33	–
SD:	1.89	3.29	4.72	2.37	–
MAX:	4.28	11.64	11.78	6.70	–

Table S26: Conformational energies of melatonin conformers (MCONF)[S10,S23] computed with different semiempirical methods. The values are given in kcal/mol. Though the results for all semiempirical methods apart from GFN2-xTB are also found in Ref. S3 and its supporting information, they are listed here along with the statistical data, since the reference values have been revised in Ref. S10.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	1.06	0.65	1.37	1.46	0.39
2	1.30	1.53	2.11	1.56	1.74
3	1.60	1.27	2.52	2.01	1.16
4	-0.28	0.29	1.91	0.27	2.20

5	-0.29	-0.05	0.56	-0.63	2.20
6	-0.20	0.52	0.44	-0.17	2.68
7	-0.29	0.48	2.46	0.64	2.92
8	2.15	1.79	3.48	2.62	2.23
9	0.22	1.38	1.42	0.13	2.84
10	2.60	2.40	3.11	4.68	4.24
11	2.90	2.47	3.41	4.68	4.45
12	5.21	3.77	4.17	6.76	3.60
13	2.57	2.11	3.05	2.96	2.25
14	5.44	3.93	4.34	6.92	3.74
15	3.38	2.53	2.23	3.97	5.00
16	3.42	2.78	1.89	4.61	5.11
17	3.53	3.18	4.56	3.94	3.18
18	0.52	1.70	2.14	0.60	3.83
19	1.48	2.06	2.82	1.63	3.80
20	2.82	2.41	3.71	3.38	3.11
21	3.67	3.77	2.98	4.80	5.27
22	3.80	3.80	3.13	4.81	5.31
23	5.92	4.42	4.44	7.51	4.50
24	3.68	4.01	4.90	3.04	3.85
25	5.96	4.44	4.53	7.49	4.55
26	1.83	2.40	3.54	2.12	4.76
27	2.16	2.62	2.94	1.82	4.37
28	6.81	6.07	5.62	7.26	5.27
29	2.24	3.12	3.80	2.35	5.67
30	2.44	3.03	3.33	2.47	4.86
31	5.00	4.50	4.40	6.32	6.24
32	5.08	4.51	4.42	6.34	6.26
33	2.95	3.61	4.54	2.81	5.85
34	3.02	3.97	4.87	2.12	5.37
35	2.60	3.17	3.75	2.56	5.53
36	6.19	5.85	5.21	7.21	7.53
37	2.85	3.50	4.22	3.04	5.88
38	3.41	4.09	4.57	2.84	5.58
39	5.83	5.24	4.59	6.77	6.98
40	5.97	5.34	4.74	6.86	7.07
41	3.51	4.41	5.69	2.67	6.39
42	6.04	5.58	5.02	7.24	7.32
43	6.15	5.65	5.13	7.30	7.39
44	3.89	4.83	5.92	2.98	6.18
45	6.70	6.46	6.42	6.91	7.82

46	6.74	6.57	6.56	6.94	7.89
47	4.03	4.70	5.58	3.53	6.74
48	7.24	6.79	6.34	7.82	8.19
49	7.24	6.81	6.32	7.82	8.20
50	4.47	5.36	6.89	3.59	7.28
51	7.69	7.48	7.68	7.87	8.75
MD:	-1.36	-1.38	-0.98	-0.91	–
MAD:	1.73	1.44	1.34	1.66	–
SD:	1.42	0.89	1.19	1.89	–
MAX:	3.43	2.55	3.22	3.72	–

Table S27: Conformational energies of tri- and tetrapeptide conformers (PCONF21)[S10,S24,S25] computed with different semiempirical methods. The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	3.05	3.45	2.36	0.50	0.02
2	5.19	5.69	5.09	2.21	1.01
3	2.81	3.71	4.68	3.93	0.70
4	3.16	3.02	3.10	0.13	0.85
5	1.70	1.62	1.79	2.85	0.78
6	6.16	7.78	7.84	2.47	1.92
7	2.65	2.75	1.86	1.83	2.18
8	3.56	3.16	4.56	4.09	1.61
9	4.43	3.57	3.85	2.54	1.89
10	3.80	5.42	4.81	0.32	2.07
11	0.64	0.41	-2.03	-2.13	1.07
12	-0.58	0.25	2.49	-0.53	1.23
13	2.88	-1.06	1.20	-1.29	2.44
14	0.14	0.77	-0.19	0.64	2.14
15	0.82	1.59	-2.18	-1.45	1.47
16	1.23	2.67	4.16	1.91	2.80
17	2.32	-2.10	0.41	-1.86	2.27
18	1.58	1.97	0.78	2.14	2.74
MD:	0.91	0.86	0.85	-0.60	–
MAD:	1.76	2.17	2.46	1.79	–
SD:	1.97	2.67	2.73	2.12	–
MAX:	4.24	5.86	5.92	4.13	–

Table S28: Conformational energies of sugar conformers (SCONF)[S10,S26] computed with different semiempirical methods. The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	0.72	0.31	0.62	0.29	0.86
2	4.27	3.54	5.97	2.30	2.28
3	4.84	3.75	6.45	2.51	3.08
4	4.46	4.72	7.54	3.13	4.60
5	4.41	4.32	6.48	3.71	4.87
6	5.28	5.69	8.82	4.16	4.16
7	5.70	6.15	9.02	4.60	4.38
8	5.83	6.22	8.84	4.87	6.19
9	5.85	7.37	8.90	4.98	6.18
10	7.01	8.09	10.64	6.03	5.65
11	6.77	7.88	10.62	5.71	5.59
12	4.83	5.64	6.22	3.04	5.93
13	5.65	6.92	8.54	4.67	6.31
14	5.56	7.83	7.97	4.58	6.22
15	-1.36	-1.47	-2.19	-1.73	0.20
16	-0.40	-7.21	-15.21	-1.25	6.16
17	-1.71	-7.06	-14.26	-0.65	5.54
MD:	-0.62	-0.91	-0.19	-1.60	–
MAD:	1.64	2.50	4.96	1.69	–
SD:	2.59	4.68	7.93	2.16	–
MAX:	7.25	13.37	21.37	7.41	–

Table S29: Conformational energies of RNA-backbone conformers (UPU23)[S10,S27] computed with different semiempirical methods. The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	9.84	6.46	10.06	7.09	4.87
2	6.16	3.69	4.78	4.35	2.97
3	13.56	11.88	13.33	10.46	8.90
4	6.54	3.05	6.68	2.25	2.22
5	4.32	2.68	2.20	3.32	2.02
6	5.24	2.22	5.29	2.63	3.14
7	4.51	0.92	0.97	-1.86	0.57

8	3.63	2.36	5.16	2.67	3.32
9	12.20	6.96	10.64	6.50	7.26
10	8.30	5.00	5.31	4.24	3.96
11	14.56	11.98	13.45	11.61	11.13
12	7.77	5.60	9.07	5.17	4.82
13	17.15	13.85	17.73	13.23	14.41
14	8.56	3.38	7.82	4.15	5.15
15	4.86	4.15	4.98	2.78	5.48
16	10.23	7.83	10.24	5.82	6.84
17	4.46	2.21	4.01	1.29	3.90
18	8.57	4.72	6.82	4.78	6.43
19	9.81	4.89	8.83	4.11	5.42
20	5.43	3.53	5.32	3.60	6.70
21	8.05	7.50	10.77	6.97	5.60
22	13.48	12.31	14.87	11.42	10.42
23	7.50	5.10	7.79	4.17	6.09
MD:	2.74	0.03	2.37	-0.47	–
MAD:	2.91	1.24	2.53	1.34	–
SD:	1.72	1.48	1.89	1.53	–
MAX:	4.97	3.17	5.20	3.10	–

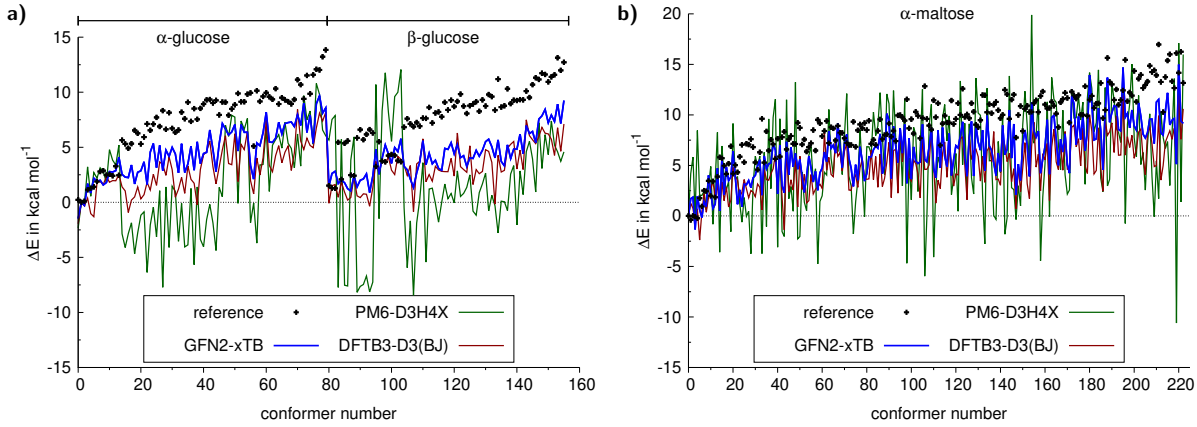


Figure S1: Conformational energies **a)** of 80 α-glucose and 76 β-glucose conformers and **b)** of 205 α-maltose conformers. The energies are given in kcal mol^{-1} and are computed with GFN2-xTB, DFTB3-D3(BJ), and PM6-D3H4X. The structures and reference conformational energies are taken from Ref. S28.

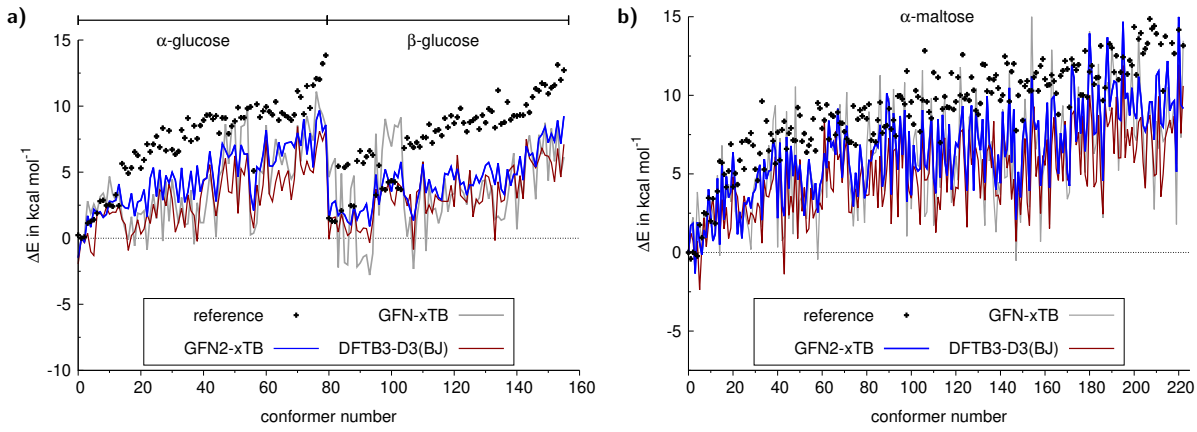


Figure S2: Conformational energies **a)** of 80 α-glucose and 76 β-glucose conformers and **b)** of 205 α-maltose conformers. The energies are given in kcal mol^{-1} and are computed with GFN-xTB, DFTB3-D3(BJ), and PM6-D3H4X. The structures and reference conformational energies are taken from Ref. S28.

Table S30: Conformational energies for different glucose conformers[S28] computed with different semiempirical methods. The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
			α-glucose		
1	-1.47	-1.14	-2.34	-1.95	0.23
2	0.00	0.00	0.00	0.00	0.00
3	-0.28	0.20	-0.16	-0.34	0.10
4	1.69	2.80	3.22	0.38	1.14
5	0.84	1.87	0.55	-1.02	1.28
6	2.19	2.58	2.70	-1.31	1.41
7	1.90	1.97	2.20	1.79	1.90
8	2.00	3.14	3.14	1.91	2.80
9	1.58	3.40	2.81	1.60	2.87
10	1.80	2.63	1.79	2.36	2.48
11	2.79	4.68	4.78	1.82	2.54
12	2.75	3.34	3.83	1.98	2.41
13	2.88	2.87	3.40	1.96	3.28
14	4.11	3.51	5.20	2.53	2.46
15	2.54	1.63	-2.86	1.41	5.65
16	2.29	0.64	-4.11	1.04	5.31
17	2.25	-0.12	-2.84	-0.92	4.91
18	3.39	1.63	-1.86	0.20	5.33
19	1.67	-0.58	-4.67	0.49	6.53
20	2.73	1.89	-2.77	1.40	6.27

21	2.87	1.80	-1.17	0.32	5.30
22	3.68	2.08	-0.86	0.70	5.93
23	1.62	-0.89	-6.37	1.82	7.12
24	4.57	1.85	-2.05	1.12	6.54
25	5.34	3.33	-0.44	2.56	7.11
26	3.15	1.49	-3.14	1.54	7.71
27	5.17	4.66	1.06	3.67	6.89
28	2.88	-0.22	-7.72	3.43	8.33
29	4.74	4.28	1.40	2.45	6.63
30	4.17	1.29	-4.10	4.12	8.14
31	4.99	4.45	1.47	2.97	6.71
32	4.40	2.69	0.99	0.46	6.35
33	2.68	1.13	-3.54	1.20	6.57
34	5.32	3.09	-0.09	2.27	8.19
35	2.72	1.01	-4.06	0.82	7.59
36	3.48	1.10	-3.86	3.29	9.09
37	5.09	3.63	1.26	2.09	7.35
38	2.92	-0.46	-5.67	1.73	8.51
39	5.49	2.68	1.40	-0.23	7.54
40	3.52	1.19	-3.37	3.00	8.53
41	4.43	1.18	-4.04	4.31	8.93
42	4.54	2.79	-1.57	4.62	9.31
43	6.57	4.83	-0.19	5.57	9.34
44	4.79	3.32	-1.88	3.22	9.02
45	2.79	0.08	-4.38	1.68	9.26
46	5.34	4.72	1.36	3.21	7.94
47	7.15	5.17	0.50	4.77	9.11
48	6.67	8.64	7.01	6.05	9.09
49	6.12	5.62	6.00	3.53	7.91
50	6.40	8.38	7.90	3.42	8.48
51	7.07	8.82	7.74	6.23	8.86
52	4.66	2.15	-0.09	1.88	9.89
53	7.06	8.44	7.39	5.26	9.13
54	6.36	8.56	6.64	4.79	9.13
55	6.36	9.64	7.62	6.43	9.82
56	3.35	0.26	-1.86	1.98	9.94
57	2.11	0.39	-0.07	0.76	5.10
58	5.39	1.14	-3.71	4.03	10.14
59	5.14	5.14	3.40	2.48	9.16
60	6.32	4.90	3.44	4.04	9.62
61	8.18	8.56	6.63	7.45	9.95

62	5.48	3.32	0.99	3.35	9.59
63	5.45	5.41	5.18	2.78	9.35
64	5.82	7.51	5.70	4.25	10.33
65	7.27	6.50	6.14	5.15	8.93
66	6.21	5.54	4.92	3.59	9.50
67	7.15	6.46	7.26	4.55	9.31
68	7.24	5.26	5.36	5.12	9.15
69	7.29	4.49	4.90	2.90	8.94
70	6.88	4.97	3.13	4.48	9.34
71	7.97	8.54	6.21	8.44	11.14
72	6.99	5.82	3.29	5.56	10.69
73	9.01	7.87	9.70	4.79	9.40
74	6.35	9.25	7.50	5.82	11.52
75	7.94	8.06	8.29	5.17	9.86
76	5.36	7.40	5.27	4.96	11.58
77	8.95	11.09	10.80	6.76	12.08
78	9.67	9.86	9.79	8.10	12.02
79	8.10	9.34	6.57	7.38	13.22
80	8.57	8.49	6.04	8.17	13.84
			β-glucose		
81	1.46	4.62	5.65	-0.08	1.52
82	2.62	5.52	7.53	1.52	1.29
83	2.82	6.04	7.88	1.61	1.27
84	1.71	-2.21	-7.73	0.46	5.44
85	2.02	4.90	5.72	1.82	2.09
86	1.77	-1.84	-7.53	0.93	5.36
87	1.24	-2.01	-7.59	0.20	5.58
88	1.03	3.74	3.84	0.44	2.47
89	2.42	4.74	5.94	2.30	2.42
90	1.40	-2.17	-8.17	0.59	5.85
91	2.00	-1.88	-7.64	0.43	6.60
92	2.13	-1.41	-7.46	0.77	5.69
93	2.58	-1.26	-6.69	0.73	6.17
94	0.90	-2.79	-8.13	-0.35	6.61
95	2.37	-1.16	-7.25	1.32	6.30
96	4.81	7.91	10.70	3.72	3.26
97	2.55	1.73	-1.48	2.70	5.51
98	4.00	6.63	8.30	3.87	3.75
99	5.21	9.03	11.77	3.52	3.68
100	4.39	7.74	8.15	4.10	4.24
101	3.54	6.95	5.54	2.92	4.35

102	4.97	8.55	9.60	2.52	4.26
103	4.61	8.41	9.99	4.17	3.79
104	5.78	9.14	12.06	3.90	3.69
105	3.26	2.34	-0.50	3.39	6.83
106	2.96	0.13	-5.94	2.44	7.60
107	2.11	1.62	-1.69	3.34	7.13
108	1.34	-2.33	-8.45	-0.84	6.87
109	2.67	1.42	-1.88	3.49	7.25
110	3.68	3.06	0.27	3.73	6.92
111	5.60	5.83	3.60	5.77	7.65
112	3.42	-0.97	-5.43	1.87	7.62
113	4.46	4.08	1.21	2.52	8.62
114	3.50	2.14	1.28	1.92	7.21
115	3.83	2.25	-0.10	3.27	8.11
116	3.08	1.13	-1.80	2.93	8.21
117	2.91	1.99	0.02	2.81	7.77
118	4.90	4.52	1.67	2.79	9.08
119	4.81	1.99	0.33	2.78	8.39
120	5.41	3.28	0.80	3.99	9.28
121	4.40	3.37	-0.06	2.38	8.70
122	5.26	5.51	2.85	6.30	9.16
123	4.06	3.75	0.54	3.24	9.83
124	4.23	3.41	0.81	2.98	8.73
125	3.26	2.54	1.36	2.73	8.07
126	4.05	4.89	2.49	5.14	8.78
127	3.99	2.18	-0.80	1.62	8.93
128	4.72	2.68	-1.17	2.97	9.14
129	4.19	3.05	0.88	2.60	8.26
130	4.73	4.34	1.68	2.81	9.31
131	5.51	4.39	2.65	4.78	8.75
132	5.73	2.98	-0.78	4.63	9.79
133	5.05	4.00	0.64	3.63	9.60
134	4.04	1.44	-0.32	-0.17	8.37
135	5.67	1.67	-2.78	4.27	11.19
136	3.71	1.36	0.10	2.75	8.65
137	3.32	1.48	0.05	2.90	8.77
138	5.22	2.92	-0.75	4.92	10.31
139	3.98	4.77	2.57	5.27	8.97
140	4.84	6.44	3.33	5.53	9.28
141	4.82	0.68	-2.06	3.02	9.25
142	6.19	4.60	2.85	4.74	9.33

143	4.40	3.16	0.70	2.98	9.31
144	6.97	6.63	4.61	7.49	10.10
145	6.24	7.78	5.54	6.05	9.52
146	6.43	6.00	3.94	5.68	10.69
147	5.41	2.95	0.94	5.33	11.32
148	7.56	7.83	7.23	5.99	11.16
149	7.85	6.36	3.08	6.80	11.73
150	8.44	8.74	6.14	6.28	11.61
151	7.23	4.36	4.33	6.21	10.94
152	7.92	8.32	7.38	5.73	11.39
153	7.35	5.82	5.19	4.26	11.58
154	8.92	8.49	4.58	6.78	13.13
155	7.16	4.84	3.72	4.63	11.98
156	9.24	6.14	4.58	7.11	12.72
MD:	-3.07	-3.79	-6.01	-4.40	–
MAD:	3.24	4.57	7.05	4.42	–
SD:	1.95	3.55	5.40	2.13	–
MAX:	6.59	9.68	16.05	8.54	–

Table S31: Conformational energies for different maltose conformers[S28] computed with different semiempirical methods. The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	0.00	0.00	0.00	0.00	0.00
2	1.71	1.47	4.18	1.24	-0.40
3	1.86	2.24	5.83	0.76	0.01
4	-1.33	-0.53	-0.01	-0.87	-0.07
5	1.93	3.85	8.45	0.71	-0.23
6	0.63	1.69	2.06	-2.38	1.76
7	-0.15	0.83	2.02	0.41	0.94
8	1.43	2.44	2.45	2.20	2.51
9	0.94	1.39	0.60	0.37	2.45
10	4.02	2.73	2.98	1.83	3.47
11	2.27	3.24	4.90	2.62	1.99
12	0.73	2.44	-0.82	3.46	3.42
13	1.17	1.31	3.57	1.65	1.85
14	4.71	4.91	8.27	3.17	3.89
15	0.50	-0.22	-3.59	0.56	5.80
16	5.75	6.81	6.92	3.36	5.71
17	2.07	3.39	3.74	1.50	4.97
18	4.83	3.49	2.57	2.86	4.16

19	3.50	4.34	3.42	0.89	5.18
20	3.85	3.23	-1.56	1.55	6.89
21	6.39	6.19	9.15	4.81	4.18
22	3.94	2.61	3.71	3.90	5.03
23	1.16	2.30	3.86	3.55	4.32
24	2.58	1.90	1.23	1.90	5.86
25	0.90	0.88	-0.61	2.34	7.04
26	3.15	2.91	1.81	3.60	5.31
27	3.08	2.09	-0.04	2.69	7.15
28	3.51	1.47	0.47	1.80	7.06
29	2.81	0.18	-3.75	2.52	7.60
30	3.77	3.80	3.02	2.36	6.81
31	4.01	4.68	5.91	1.90	5.29
32	4.98	4.48	4.89	3.16	4.58
33	5.57	5.63	5.92	2.83	6.30
34	2.05	1.61	-3.73	0.59	9.62
35	5.94	3.51	6.30	4.58	5.28
36	4.67	4.07	-1.05	3.59	8.37
37	5.17	6.82	9.60	3.53	6.47
38	5.23	3.99	2.70	4.16	7.55
39	7.14	8.42	9.74	4.67	7.10
40	6.52	9.36	11.55	6.09	5.65
41	2.18	0.33	-3.09	2.50	7.65
42	6.99	7.28	9.62	6.30	5.87
43	2.98	0.74	-1.54	1.84	7.80
44	2.31	1.02	-2.03	-1.38	8.34
45	7.35	8.97	11.50	6.61	6.59
46	4.07	2.14	1.70	2.37	7.58
47	6.78	6.31	2.91	3.58	7.91
48	4.86	1.17	0.47	2.56	7.14
49	6.20	10.55	13.23	5.53	7.12
50	4.62	1.20	-2.27	0.89	9.22
51	5.59	3.28	-2.04	5.92	8.72
52	3.91	3.60	4.34	2.62	7.07
53	2.33	2.65	1.35	4.14	7.57
54	3.93	2.14	0.51	1.28	6.37
55	5.93	5.28	5.99	1.51	6.81
56	4.77	4.88	6.87	4.10	7.20
57	3.35	4.29	3.40	2.18	8.50
58	4.41	3.33	4.66	2.64	6.47
59	2.54	-0.45	-4.75	2.89	9.51

60	4.27	3.55	1.43	3.14	9.17
61	4.74	4.09	1.99	2.73	7.89
62	6.32	7.19	4.33	8.31	8.41
63	5.36	3.49	4.05	4.53	9.40
64	8.39	7.97	10.53	5.80	7.21
65	6.69	6.78	5.68	4.11	8.78
66	7.02	6.47	4.11	3.88	8.66
67	6.22	6.78	7.70	7.71	7.48
68	9.14	8.52	6.37	5.90	8.92
69	6.94	9.43	8.22	6.27	8.24
70	4.76	4.91	4.10	4.44	8.22
71	7.18	6.12	7.96	6.42	7.69
72	8.07	10.20	12.11	7.17	7.23
73	5.71	4.70	3.94	2.57	7.01
74	4.74	3.67	2.97	2.52	7.02
75	2.93	1.67	-1.79	3.70	9.32
76	5.07	4.28	1.39	4.04	6.84
77	3.53	2.00	3.51	2.83	8.47
78	6.85	6.90	8.52	4.78	9.06
79	5.44	5.10	4.53	5.63	8.68
80	3.45	1.89	-0.08	3.76	9.20
81	4.20	6.72	6.85	3.62	8.51
82	5.85	8.96	6.84	3.67	9.43
83	5.83	6.01	6.11	6.49	7.14
84	6.55	6.26	10.28	3.59	7.53
85	6.93	6.37	7.96	5.89	7.03
86	6.71	6.97	8.91	2.84	8.42
87	8.12	11.28	12.20	6.13	9.55
88	4.19	5.11	7.88	2.91	7.60
89	4.96	3.89	4.07	3.74	8.00
90	10.12	8.77	12.08	5.30	8.08
91	9.49	9.81	11.10	7.28	9.15
92	3.69	3.75	4.62	1.81	7.85
93	8.87	10.40	12.40	6.46	8.42
94	6.92	6.25	7.58	6.60	7.41
95	5.28	5.73	2.85	2.25	9.95
96	8.69	8.24	8.05	8.62	9.98
97	6.21	6.09	2.97	5.36	9.76
98	9.06	10.82	11.48	8.69	9.44
99	3.23	1.03	-4.68	5.03	11.54
100	7.70	7.59	5.70	3.91	8.67

101	3.64	2.76	1.59	3.45	9.58
102	6.83	7.17	8.78	5.50	8.15
103	5.43	7.85	8.07	5.19	10.30
104	5.19	5.55	2.42	4.14	9.91
105	5.34	5.41	6.95	1.77	7.69
106	9.06	11.59	12.54	7.87	9.70
107	5.58	2.28	-5.96	4.60	12.84
108	5.36	2.57	-0.05	4.62	9.60
109	6.85	6.56	8.90	5.05	8.94
110	8.65	9.83	11.06	8.61	8.54
111	4.73	1.72	-4.07	4.97	11.45
112	3.77	2.85	2.94	1.26	9.77
113	8.48	6.34	5.84	3.53	11.14
114	4.75	4.12	4.94	2.71	6.99
115	8.75	8.35	9.69	6.18	9.13
116	5.90	9.75	11.03	7.29	9.92
117	4.88	5.75	7.44	3.83	9.25
118	8.16	7.96	9.90	6.56	9.22
119	4.00	3.18	2.12	5.29	11.42
120	5.55	3.93	3.71	2.95	9.10
121	6.96	6.15	6.18	4.15	9.60
122	6.72	4.60	0.47	4.06	11.15
123	7.94	11.80	13.36	7.65	10.00
124	9.32	8.04	5.06	7.81	12.59
125	5.99	6.54	9.18	5.09	9.79
126	3.92	2.22	2.86	1.61	10.05
127	7.79	8.64	9.11	7.81	9.67
128	9.35	7.16	3.77	5.60	11.15
129	3.95	5.05	2.17	6.41	10.32
130	8.97	10.45	13.65	6.61	9.03
131	7.32	7.75	10.18	5.87	8.78
132	6.51	6.27	9.96	3.45	9.48
133	8.48	4.93	2.09	5.92	11.97
134	4.46	1.71	-2.77	3.32	11.94
135	9.94	9.39	12.64	6.44	10.26
136	6.93	5.82	5.46	6.86	11.06
137	6.50	7.06	5.49	5.37	10.16
138	5.27	3.50	0.04	8.35	11.00
139	4.00	1.29	-0.33	3.59	11.28
140	6.22	4.47	7.37	3.64	10.46
141	4.41	3.68	4.29	3.04	9.67

142	8.96	9.10	10.02	7.82	10.01
143	6.07	4.75	-1.81	6.98	12.48
144	6.59	4.83	-0.15	5.37	12.03
145	10.48	11.41	13.86	6.21	9.61
146	5.72	3.75	2.80	3.38	11.51
147	9.10	6.95	2.61	4.67	12.01
148	2.10	-0.53	-0.36	0.70	7.76
149	5.13	3.76	2.99	4.95	9.99
150	3.51	1.91	0.99	2.41	10.87
151	7.14	5.12	8.96	4.71	8.40
152	8.51	9.77	12.34	7.25	10.59
153	7.00	7.04	8.93	3.18	10.91
154	3.92	2.89	4.86	1.71	10.29
155	11.21	15.33	19.89	6.92	10.29
156	8.43	8.03	7.20	9.28	11.15
157	7.80	6.57	7.99	4.83	11.44
158	8.40	8.17	11.66	6.12	10.28
159	4.23	0.69	-4.44	2.48	11.87
160	5.64	4.89	1.86	6.16	12.31
161	6.87	5.79	5.18	5.54	11.03
162	9.53	8.83	11.29	7.18	10.64
163	4.12	1.78	-0.06	1.53	10.68
164	9.76	12.90	13.71	6.82	11.29
165	9.94	9.09	11.54	8.60	10.90
166	7.61	8.49	11.24	5.35	10.12
167	5.01	5.61	7.07	4.58	8.78
168	7.59	7.49	4.56	7.34	11.24
169	6.32	9.35	10.47	4.53	11.56
170	8.12	5.96	4.07	8.35	12.19
171	5.07	3.39	2.16	4.63	11.67
172	4.24	3.83	2.57	4.64	9.24
173	11.90	12.46	12.44	8.76	12.30
174	11.10	10.22	9.69	8.99	11.92
175	7.05	7.91	9.72	6.83	11.61
176	9.84	10.55	11.85	7.64	11.95
177	7.57	6.25	9.49	4.95	10.02
178	9.99	10.06	11.26	8.69	10.01
179	6.43	5.09	3.13	4.94	12.83
180	8.33	7.88	11.70	4.21	9.89
181	13.94	14.07	14.19	10.02	11.50
182	11.21	10.28	12.64	8.41	11.88

183	7.52	6.28	8.35	4.16	10.70
184	7.63	6.94	9.24	4.51	11.12
185	10.67	12.00	14.45	8.72	11.14
186	8.51	6.03	3.73	4.80	13.15
187	4.57	2.19	2.40	2.44	9.67
188	10.39	10.14	12.71	9.45	10.48
189	13.69	10.84	2.97	7.98	15.64
190	9.28	9.00	3.62	8.33	13.27
191	10.22	12.94	13.16	9.22	13.02
192	10.90	11.30	11.45	6.28	11.48
193	10.81	11.50	8.64	9.90	12.52
194	8.55	5.21	8.27	6.01	11.00
195	8.41	8.27	7.28	6.02	11.47
196	14.68	13.78	14.67	11.90	12.43
197	9.77	7.28	-0.74	6.02	15.44
198	10.15	9.68	9.08	7.84	11.35
199	8.72	9.16	11.11	6.78	12.22
200	10.48	11.90	15.02	7.98	11.91
201	8.88	7.50	7.75	5.71	14.35
202	8.50	7.11	5.86	6.58	11.30
203	11.19	13.88	11.93	8.29	13.58
204	9.87	9.69	5.09	8.60	13.96
205	8.09	7.85	6.28	5.74	12.66
206	7.71	4.25	2.56	4.09	12.96
207	8.24	8.68	7.05	8.24	13.30
208	9.58	9.29	10.11	8.74	14.86
209	9.58	7.21	3.90	7.16	14.24
210	7.17	2.74	-0.07	4.88	14.41
211	11.27	8.34	5.45	4.98	13.92
212	11.52	6.62	1.71	9.92	16.97
213	8.22	8.06	10.23	3.55	13.26
214	7.99	7.07	7.76	4.99	10.28
215	10.38	7.33	6.68	7.31	13.75
216	12.18	10.49	7.35	7.88	15.38
217	9.01	7.92	3.01	8.43	16.00
218	9.48	8.31	8.14	7.58	13.32
219	9.65	9.47	13.59	8.68	12.66
220	5.13	1.79	-10.60	5.62	16.11
221	14.99	15.32	17.11	11.94	14.17
222	9.39	7.31	1.38	7.58	16.25
223	9.16	13.47	15.98	10.60	13.16

MD:	-2.87	-3.27	-3.64	-4.39	–
MAD:	3.11	3.81	5.07	4.44	–
SD:	2.44	3.47	5.63	2.55	–
MAX:	10.98	14.32	26.71	10.49	–

2.4 Rotational and vibrational free energy computations

Table S32: Rotational and vibrational reaction free energies ΔG_{RRHO} for the AL2X6[S10] set computed with different methods. The values are given in kcal/mol and B97-3c served as reference in the manuscript.

reaction #	B97-3c	PBEh-3c	GFN-xTB	GFN2-xTB
1	-1.04	-1.09	-1.35	-1.25
2	-2.16	-2.20	-2.21	-2.43
3	-2.67	-2.70	-2.66	-2.96
4	-4.15	-1.63	-1.80	-2.57
5	-4.54	-1.70	-2.27	-3.10
6	-4.92	-1.50	-2.45	-2.04
MD		1.44	1.12	0.85
MAD		1.49	1.25	1.11
SD		1.65	1.37	1.32
MAX		3.42	2.47	2.87

Table S33: Rotational and vibrational reaction free energies ΔG_{RRHO} for the DARC[S10] set computed with different methods. The values are given in kcal/mol and B97-3c served as reference in the manuscript.

reaction #	B97-3c	PBEh-3c	GFN-xTB	GFN2-xTB
1	0.39	0.41	0.40	0.35
2	-0.01	0.12	0.13	0.09
3	0.53	0.54	0.57	0.57
4	0.06	0.18	0.22	0.28
5	0.58	0.59	0.48	0.52
6	0.04	0.17	0.14	0.27
7	1.31	1.33	1.32	1.32
8	1.32	1.35	1.31	1.46
9	1.21	1.26	1.25	1.29
10	1.22	1.28	1.29	1.34
11	1.14	1.16	1.15	1.15
12	1.15	1.14	1.15	1.13
13	1.04	1.09	1.09	1.12
14	1.04	1.07	1.12	1.19
MD		0.05	0.04	0.07
MAD		0.05	0.06	0.09
SD		0.05	0.07	0.09

MAX		0.13	0.17	0.22

Table S34: Rotational and vibrational reaction free energies ΔG_{RRHO} for the HEAVYSB11[S10] set computed with different methods. The values are given in kcal/mol and B97-3c served as reference in the manuscript.

reaction #	B97-3c	PBEh-3c	GFN-xTB	GFN2-xTB
1	-1.37	-1.36	-1.73	-1.59
2	-2.56	-2.68	-2.57	-2.40
3	-5.98	-2.64	-2.71	-2.59
4	-0.39	-0.38	-0.45	-0.48
5	-0.61	-0.59	-0.63	-0.69
6	-2.53	-2.30	-2.67	-2.69
7	-1.80	-1.87	-1.96	-2.01
8	-2.16	-2.23	-2.35	-2.45
9	-2.47	-2.47	-2.62	-2.61
10	-0.14	-0.13	-0.12	-0.14
11	-0.26	-0.25	-0.23	-0.29
MD		0.31	0.20	0.21
MAD		0.35	0.40	0.43
SD		1.01	1.02	1.06
MAX		3.34	3.27	3.40

Table S35: Rotational and vibrational reaction free energies ΔG_{RRHO} for the ISOL24[S10] set computed with different methods. The values are given in kcal/mol and B97-3c served as reference in the manuscript.

reaction #	B97-3c	PBEh-3c	GFN-xTB	GFN2-xTB
1	0.50	0.18	0.35	0.68
2	2.00	2.12	2.40	2.40
3	0.37	0.56	0.57	0.47
4	4.34	4.50	4.74	4.76
5	1.82	1.77	1.98	2.05
6	0.39	0.36	0.53	0.51
7	0.66	0.72	0.73	0.73
8	-1.38	-1.45	-1.67	-1.58
9	0.20	0.29	0.07	0.11

10	0.19	0.04	0.17	0.29
11	-0.55	-0.47	-0.74	-0.18
12	0.07	0.00	-0.06	-0.04
13	-0.27	-0.36	-0.29	-0.13
14	0.35	0.33	0.24	0.35
15	0.67	0.79	1.18	1.14
16	0.35	0.35	0.12	0.26
17	0.70	0.76	0.83	0.80
18	0.06	0.07	0.07	0.09
19	-0.06	-0.07	-0.06	-0.08
20	-0.01	-0.03	0.04	0.01
21	1.30	1.29	1.36	1.38
22	-0.83	-0.95	-1.01	-0.94
23	1.16	1.21	1.12	1.18
24	0.47	0.42	0.37	0.25
MD		-0.00	0.02	0.08
MAD		0.08	0.15	0.15
SD		0.11	0.21	0.19
MAX		0.32	0.51	0.47

Table S36: Rotational and vibrational reaction free energies ΔG_{RRHO} for the TAUT15[S10] set computed with different methods. The values are given in kcal/mol and B97-3c served as reference in the manuscript.

reaction #	B97-3c	PBEh-3c	GFN-xTB	GFN2-xTB
1	0.78	0.18	0.17	0.12
2	0.64	0.53	0.49	0.60
3	0.05	0.52	0.02	0.38
4	0.27	0.36	-0.27	0.15
5	0.17	0.27	-0.33	0.12
6	0.03	0.02	0.01	0.01
7	0.13	0.13	0.09	0.13
8	0.00	-0.17	-0.02	-0.19
9	0.07	0.01	0.01	-0.01
10	0.08	-0.00	0.00	-0.03
11	0.09	0.04	0.02	0.02
12	0.04	0.03	0.02	0.03
13	-0.01	-0.02	-0.03	-0.06
14	-0.19	-0.18	-0.07	-0.14

15	-0.22	-0.25	-0.09	-0.16
MD		-0.03	-0.13	-0.07
MAD		0.12	0.16	0.12
SD		0.21	0.23	0.20
MAX		0.60	0.62	0.67

Table S37: Rotational and vibrational reaction free energies ΔG_{RRHO} for the ALK8[S10] set computed with different methods. The values are given in kcal/mol and B97-3c served as reference in the manuscript.

reaction #	B97-3c	PBEh-3c	GFN-xTB	GFN2-xTB
1	-5.08	-5.42	-4.66	-5.49
2	-6.87	-7.04	-6.55	-6.53
3	-4.41	-4.44	-3.70	-3.75
4	-1.40	-1.97	-1.33	-1.91
5	-0.78	-1.30	-0.77	-0.85
6	-0.93	-1.10	-2.43	-2.33
7	-2.46	-2.45	-2.82	-2.49
8	0.22	0.23	0.15	0.20
MD		-0.22	-0.05	-0.18
MAD		0.23	0.43	0.43
SD		0.23	0.67	0.62
MAX		0.57	1.50	1.41

Table S38: Rotational and vibrational reaction free energies ΔG_{RRHO} for the G2RC[S10] set computed with different methods. The values are given in kcal/mol and B97-3c served as reference in the manuscript.

reaction #	B97-3c	PBEh-3c	GFN-xTB	GFN2-xTB
1	-1.51	-1.42	-1.65	-1.61
2	0.05	0.05	0.03	0.02
3	-0.65	-0.65	-0.71	-0.67
4	-1.10	-1.13	-1.31	-1.17
5	0.16	0.16	0.18	0.20
6	-0.12	0.43	0.47	0.50
7	0.59	1.03	0.16	0.47
8	-0.06	0.06	0.05	-0.00

9	-0.04	-0.06	-0.02	-0.05
10	0.01	0.05	-0.18	-0.14
11	0.15	0.10	0.31	0.27
12	0.28	0.28	0.31	0.25
13	0.10	0.23	0.16	0.11
14	0.15	0.16	0.22	0.25
15	0.04	0.04	0.05	0.03
16	0.20	0.18	0.20	0.23
17	-0.26	-0.13	-0.17	-0.23
18	-0.16	-0.14	-0.14	-0.14
19	0.02	0.03	0.03	0.02
20	-0.21	-0.18	-0.17	-0.18
21	-0.20	-0.18	-0.25	-0.24
22	0.20	0.20	0.17	0.16
23	-0.02	-0.02	-0.02	-0.07
24	-0.17	0.22	0.14	-0.02
25	-0.21	-0.21	-0.11	-0.20
MD		0.08	0.02	0.02
MAD		0.09	0.11	0.07
SD		0.15	0.18	0.14
MAX		0.55	0.60	0.62

Table S39: Rotational and vibrational reaction free energies ΔG_{RRHO} for the ISO34[S10] set computed with different methods. The values are given in kcal/mol and B97-3c served as reference in the manuscript.

reaction #	B97-3c	PBEh-3c	GFN-xTB	GFN2-xTB
1	-0.17	-0.07	0.24	-0.15
2	-0.47	-0.38	-0.45	-0.47
3	-0.46	-0.47	-0.56	-0.53
4	0.05	0.00	0.01	0.05
5	0.10	0.09	0.06	0.06
6	-0.14	-0.13	-0.18	-0.17
7	-0.51	-0.55	-0.62	-0.61
8	0.45	0.46	0.40	0.45
9	-0.07	-0.07	-0.06	-0.06
10	0.16	0.18	0.01	0.00
11	0.46	0.47	0.34	0.41
12	-0.32	-0.89	-0.87	-0.81

13	-0.02	-0.06	0.04	0.02
14	0.14	0.13	0.06	0.06
15	0.00	0.01	0.02	0.02
16	-0.40	-0.41	-0.51	-0.46
17	0.09	0.09	0.11	0.17
18	0.24	0.18	0.18	0.20
19	-0.01	-0.00	-0.00	0.03
20	0.05	0.03	0.03	-0.00
21	-0.54	0.03	0.56	0.56
22	0.16	0.13	-0.31	-0.31
23	0.00	-0.00	0.02	0.03
24	0.03	0.04	0.08	0.06
25	-0.47	-0.47	-0.55	-0.50
26	0.48	-0.06	0.02	0.04
27	0.35	-0.27	-0.22	0.44
28	-0.25	-0.32	-0.50	-0.37
29	0.68	0.67	0.74	0.66
30	-0.31	-0.34	-0.22	-0.18
31	0.59	1.10	1.04	1.19
32	1.29	1.27	1.33	1.27
33	0.22	-0.42	0.37	0.40
34	-0.09	-0.14	-0.13	-0.04
MD		-0.04	-0.02	0.00
MAD		0.12	0.17	0.14
SD		0.24	0.29	0.27
MAX		0.64	1.10	1.10

Table S40: Rotational and vibrational reaction free energies ΔG_{RRHO} for the MOR41[S29] set computed with different methods. The values are given in kcal/mol and B97-3c served as reference in the manuscript.

reaction #	B97-3c	PBEh-3c	GFN-xTB	GFN2-xTB
1	1.37	1.27	1.48	1.50
2	1.26	1.10	0.98	1.23
3	1.43	1.45	1.48	1.43
4	0.24	0.35	-1.56	-0.37
5	1.62	1.69	1.67	1.59
6	0.56	0.44	0.41	0.34
7	0.02	-0.11	-0.01	0.17

8	0.35	0.35	0.31	0.30
9	0.86	0.78	0.20	0.41
10	1.94	1.89	1.94	1.93
11	2.56	2.42	2.66	3.09
12	1.73	1.61	1.47	1.85
13	1.48	1.51	2.18	1.68
14	2.10	2.13	2.11	2.16
15	1.81	1.85	1.96	1.99
16	1.86	2.74	1.86	1.55
17	1.62	1.66	1.78	1.99
18	2.40	2.36	2.49	2.55
19	2.41	2.47	2.07	2.82
20	2.46	2.38	2.49	2.72
21	2.63	1.61	1.66	1.79
22	2.46	2.51	2.76	3.04
23	2.23	1.59	2.40	2.37
24	2.53	2.61	2.61	2.59
25	2.00	2.11	2.32	2.17
26	2.88	2.85	3.10	3.27
27	0.52	0.54	0.53	0.88
28	0.17	0.09	0.23	0.39
29	-0.32	0.35	0.16	0.53
30	0.23	-0.25	0.24	0.35
31	0.07	0.53	0.47	0.66
32	-1.51	-1.50	-1.71	-1.20
33	1.26	1.31	0.83	1.41
34	0.93	0.93	0.78	1.04
35	0.83	1.48	0.92	1.02
36	0.71	0.13	0.81	-0.31
37	3.99	5.50	5.46	5.63
38	5.02	5.10	5.22	4.95
39	2.57	2.54	2.37	2.44
40	2.08	0.83	3.52	2.01
41	1.00	0.68	0.17	0.69
MD		-0.01	0.01	0.10
MAD		0.25	0.32	0.31
SD		0.45	0.53	0.44
MAX		1.51	1.80	1.65

2.5 Other properties

Table S41: Barrier heights of divers reactions (BHDIV10)[S10] computed with different semiempirical methods. The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	32.76	32.42	20.23	23.91	25.65
2	50.31	50.87	31.70	70.47	56.90
3	32.95	36.65	23.99	26.48	36.53
4	87.32	86.99	91.73	86.29	96.17
5	8.05	8.70	12.68	-19.76	15.94
6	7.99	5.65	-9.91	-1.94	13.64
7	28.05	26.27	15.65	–	27.49
8	40.82	39.35	59.35	36.73	50.24
9	59.52	45.74	48.70	–	65.84
10	90.17	79.45	94.91	71.48	64.93
MD:	-1.54	-4.12	-6.43	-8.29[a]	–
MAD:	8.12	8.40	14.25	13.32[a]	–
SD:	10.65	9.70	16.36	15.02[a]	–
MAX:	25.24	20.10	29.98	35.70[a]	–

[a] Missing values are neglected in statistical analysis.

Table S42: Barrier heights of pericyclic reactions (BHPERI)[S30] computed with different semiempirical methods. The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	44.67	43.43	40.18	41.25	35.30
2	23.45	21.99	35.45	25.01	30.80
3	27.95	28.16	38.95	33.97	28.10
4	29.17	29.73	42.56	34.15	39.70
5	27.86	27.74	43.03	25.03	28.30
6	21.96	22.04	42.66	27.46	35.80
7	8.14	8.85	26.06	10.56	22.30
8	7.21	9.22	29.25	12.65	18.00
9	7.12	9.55	31.21	14.51	14.50
10	35.88	39.46	38.10	31.44	26.40
11	12.60	14.01	26.06	31.70	27.60
12	6.46	8.97	20.52	23.07	20.00
13	5.51	5.04	26.97	11.24	13.80
14	4.14	5.79	24.99	9.94	11.80

15	-1.11	0.24	10.75	8.60	6.50
16	-2.34	-2.60	8.91	4.62	4.70
17	-0.75	-0.22	18.46	8.09	13.10
18	-3.26	-3.46	12.18	3.07	5.90
19	-3.42	-4.50	13.55	-2.62	0.50
20	6.81	8.88	29.99	12.96	18.10
21	5.85	5.96	27.71	–	16.60
22	10.67	11.79	29.19	16.67	22.90
23	12.53	12.35	36.02	23.60	27.80
24	6.27	10.77	36.12	18.25	21.30
25	6.87	10.33	34.54	15.74	21.60
26	14.52	19.36	36.84	23.96	31.30
MD:	-8.77	-7.69	8.37	-2.45[a]	–
MAD:	10.22	9.32	8.49	4.54[a]	–
SD:	6.90	6.61	4.84	4.66[a]	–
MAX:	16.78	15.45	16.71	11.74[a]	–

[a] Missing value is neglected in statistical analysis.

Table S43: Barrier heights of bond rotations around single bonds (BHROT27)[S10] computed with different semiempirical methods. The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	2.58	1.95	1.26	2.18	2.73
2	6.12	4.87	5.49	4.91	7.01
3	2.62	2.50	2.13	3.02	3.46
4	2.55	2.12	2.16	2.64	3.72
5	1.62	1.02	1.05	0.99	1.01
6	2.74	1.53	1.80	1.75	2.28
7	-1.01	-0.79	-0.57	-0.91	1.01
8	7.24	6.31	8.11	3.91	7.17
9	2.06	2.07	3.49	2.05	5.79
10	5.97	4.62	4.81	3.69	8.03
11	1.05	0.17	-1.57	1.84	1.62
12	6.07	3.69	9.64	4.46	8.41
13	6.07	5.02	8.49	3.02	6.91
14	1.49	0.76	0.39	2.73	2.68
15	17.88	15.29	18.30	13.19	17.24
16	14.12	11.75	14.43	11.71	14.52
17	2.64	1.84	3.47	0.66	2.10

18	5.12	6.67	0.52	6.89	3.89
19	4.36	4.50	-0.73	6.72	2.09
20	3.88	3.52	0.59	4.42	1.78
21	3.01	3.51	-0.60	4.68	1.39
22	5.47	4.85	0.63	5.47	6.30
23	3.16	3.16	-0.03	3.47	3.35
24	10.71	6.54	5.64	14.84	10.36
25	10.50	5.21	5.02	15.01	10.24
26	15.15	10.96	12.17	18.02	17.20
27	14.95	9.63	11.54	18.20	17.08
MD:	-0.42	-1.71	-1.92	-0.36	–
MAD:	1.17	2.38	2.38	2.22	–
SD:	1.43	2.49	2.19	2.78	–
MAX:	3.73	7.45	5.67	4.77	–

Table S44: Barrier heights for inversions and racemizations (INV24)[S31] computed with different semiempirical methods. The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	31.39	21.45	15.03	16.50	31.70
2	70.15	79.34	66.09	36.13	69.30
3	62.58	69.50	5.59	122.64	60.60
4	42.84	29.21	35.42	31.54	37.00
5	78.08	85.22	71.57	57.43	74.20
6	7.80	10.05	10.48	5.48	9.70
7	14.68	14.29	0.00	10.66	18.90
8	54.13	65.53	34.27	36.74	43.20
9	72.17	60.78	59.94	–[a]	79.70
10	35.88	44.35	18.40	22.34	31.20
11	27.20	31.69	18.09	18.85	29.30
12	9.88	8.95	10.84	8.54	10.30
13	6.30	4.98	7.45	4.57	4.50
14	24.22	22.25	24.94	24.30	24.70
15	32.88	31.00	31.80	35.17	37.60
16	5.44	4.06	8.29	3.86	4.10
17	10.60	10.66	14.55	10.19	13.10
18	11.70	11.37	11.57	13.80	11.20
19	4.07	5.23	8.49	6.52	6.20
20	24.94	24.55	29.02	26.34	21.30

21	49.07	45.63	46.86	46.20	42.30
22	22.36	26.80	34.19	25.32	27.20
23	10.21	10.43	11.00	12.43	8.40
24	76.28	74.64	83.64	80.81	68.60
MD:	0.86	1.15	-4.45	-1.23[a]	–
MAD:	3.45	5.80	8.59	9.07[a]	–
SD:	4.39	8.37	13.82	16.52[a]	–
MAX:	10.93	22.33	55.01	62.04[a]	–

[a] Abnormally high repulsion energy in transition state geometry of PCL_3. Hence, this value is neglected in statistical analysis.

Table S45: Barrier heights for proton exchange reactions (PX13)[S10,S32] computed with different semiempirical methods. The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	67.02	71.22	60.14	69.08	59.30
2	40.36	43.69	37.03	52.38	46.90
3	46.16	54.41	33.18	56.37	48.40
4	51.94	47.30	46.39	49.94	48.60
5	26.94	24.94	18.06	26.96	29.80
6	22.60	23.88	12.18	22.94	26.60
7	27.76	29.23	10.56	26.17	30.10
8	34.96	35.86	14.83	30.07	35.10
9	39.42	29.90	59.91	46.46	42.30
10	18.74	5.51	43.13	31.96	20.70
11	14.16	0.34	36.79	30.41	14.70
12	14.95	-1.20	38.41	33.70	14.60
13	17.26	-1.91	44.24	38.92	16.60
MD:	-0.88	-5.43	1.63	6.28	–
MAD:	2.74	8.30	15.98	8.66	–
SD:	3.55	9.23	18.42	9.10	–
MAX:	7.72	18.51	27.64	22.32	–

Table S46: Barrier heights for proton transfer reactions (WCPT18)[S10,S33] computed with different semiempirical methods. The values are given in kcal/mol.

system #	GFN2-xTB	GFN-xTB	PM6-D3H4X	DFTB3-D3(BJ)	ref.
1	35.39	36.90	42.20	33.97	36.76
2	30.09	31.33	35.27	37.22	36.21
3	62.98	64.45	61.14	61.42	60.95
4	36.65	46.46	47.61	53.16	47.52
5	61.07	73.25	65.97	72.62	65.68
6	79.41	77.77	87.58	84.77	81.24
7	30.39	40.97	36.53	37.22	32.00
8	21.44	35.81	31.36	35.76	28.97
9	51.32	65.56	61.34	62.72	58.80
10	2.89	-0.83	0.57	3.90	5.40
11	1.54	-4.66	-1.05	4.95	2.68
12	27.57	23.68	23.93	25.81	28.78
13	9.92	9.78	6.61	11.71	8.68
14	28.72	33.73	31.03	26.27	33.89
15	51.74	49.35	44.90	51.98	59.63
16	3.84	12.18	4.28	9.73	5.83
17	0.86	6.91	0.14	7.83	3.54
18	31.38	45.51	34.90	37.03	33.22
MD:	-3.48	1.02	-0.86	1.57	–
MAD:	3.84	5.30	3.47	4.08	–
SD:	3.41	6.37	4.83	4.45	–
MAX:	10.87	12.29	14.73	7.65	–

Table S47: Molecular dipole moments computed with GFN2-xTB, GFN-xTB, and PM6. The reference is a CCSD(T)/CBS estimate and taken from Ref. S34. All quantities are given in Debye.

system	$N_\alpha - N_\beta$	GFN2-xTB	GFN-xTB	PM6	ref.
AlF	0	1.7690	3.3490	1.9660	1.4729
AlH_2	1	1.3360	1.6550	0.2020	0.4011
BeH	1	1.3590	1.2270	0.5000	0.2319
BF	0	1.4760	0.5170	0.2550	0.8194
BH	0	1.5380	0.9280	0.5070	1.4103
BH_2	1	0.0570	0.3170	0.5040	0.5004
BH_2Cl	0	1.0480	0.8410	1.0030	0.6838

BH_2F	0	0.8240	0.9630	1.1990	0.8269
$BHCl_2$	0	1.0090	1.0440	1.0360	0.6684
BHF_2	0	1.0320	1.1160	1.3590	0.9578
BN	2	2.1650	2.0600	2.5140	2.0366
BO	1	2.1910	2.8970	2.2260	2.3171
BS	1	2.3410	1.1380	1.7720	0.7834
C_2H	1	1.0190	0.5960	0.3090	0.7601
C_2H_3	1	0.8190	0.8290	0.7930	0.6867
C_2H_5	1	0.5610	0.5240	0.4780	0.3140
CF	1	1.0930	0.6960	0.0080	0.6793
CF_2	0	0.3990	0.0910	0.3440	0.5402
CH	1	1.6430	1.6120	1.3570	1.4328
CH_2BH	0	0.2670	0.5200	1.4510	0.6238
CH_2BOH	0	2.5010	2.8340	2.5990	2.2558
CH_2F	1	1.5450	1.5540	1.3770	1.3796
CH_2NH	0	1.9890	2.3310	2.4070	2.0673
CH_2PH	0	1.1980	0.5680	1.2460	0.8748
CH_2-singlet	0	1.7970	1.9180	1.9270	1.4942
CH_2-triplet	2	0.6790	0.5910	0.7780	0.5862
CH_3BH_2	0	0.8140	0.8650	0.6100	0.5751
CH_3BO	0	3.2660	4.0640	3.3970	3.6779
CH_3Cl	0	1.9770	1.8720	1.9800	1.8981
CH_3F	0	2.2740	2.1760	1.6350	1.8083
CH_3Li	0	4.6150	4.2520	5.2400	5.8304
CH_3NH_2	0	1.5380	1.8970	2.0520	1.3876
CH_3O	1	3.3450	4.0510	2.3290	2.0368
CH_3OH	0	1.9690	2.4940	2.1250	1.7091
CH_3SH	0	2.3830	1.4790	1.7700	1.5906
ClCN	0	3.2960	3.0390	2.5610	2.8496
ClF	0	1.5800	1.5770	0.0830	0.8802
ClO_2	1	2.1230	3.2720	5.1490	1.8627
CN	1	0.6100	1.5010	1.3500	1.4318
CO	0	0.6140	0.5040	0.0960	0.1172

CS	0	1.5060	1.6040	0.9750	1.9692
CSO	0	0.2610	1.6170	0.4800	0.7327
FCN	0	1.7630	2.0650	1.9950	2.1756
FCO	1	0.8190	1.6440	0.5290	0.7678
FH-BH_2	1	3.0470	3.4460	2.0560	2.9730
FH-NH_2	1	4.5960	5.7790	4.4340	4.6265
FH-OH	1	2.7590	2.8860	2.6710	3.3808
FNO	0	1.6060	1.4040	1.8310	1.6971
H_2CN	1	2.9730	3.0610	2.8030	2.4939
H_2O	0	2.2710	2.8400	2.1170	1.8601
H_2O-Al	1	4.0710	1.2550	3.4880	4.3573
H_2O-Cl	1	2.7640	3.2920	2.3630	2.2383
H_2O-F	1	3.3960	3.6030	2.1940	2.1875
H_2O-H_2O	0	2.9470	3.6130	2.6880	2.7303
H_2O-Li	1	6.0590	4.2560	0.9300	3.6184
H_2O-NH_3	0	3.7330	4.5840	4.0230	3.5004
H_2S-H_2S	0	1.2410	1.1010	0.9520	0.9181
H_2S-HCl	0	2.3110	3.0090	2.3880	2.1328
HBH_2BH	0	1.5500	1.3870	2.0880	0.8429
HBO	0	2.1720	3.0380	2.6370	2.7322
HBS	0	2.4970	1.4420	1.7920	1.3753
HCCCl	0	0.2030	0.2530	0.3550	0.5009
HCCF	0	1.4140	0.9450	0.5750	0.7452
HCHO	0	2.3870	3.3280	2.8020	2.3927
HCHS	0	2.2010	1.5610	1.4730	1.7588
HCl	0	1.0580	1.8390	1.4580	1.1055
HCl-HCl	0	1.7160	2.5480	1.9740	1.7766
HCN	0	2.6600	2.9140	2.6740	3.0065
HCNO	0	3.0650	3.5010	1.9790	2.9560
HCO	1	1.6680	2.5050	1.8220	1.6912
HCOF	0	2.0880	2.8230	2.3230	2.1169
$HCONH_2$	0	4.3700	5.1600	4.1830	3.9152
HCOOH	0	1.7630	1.8870	1.5510	1.3835

HCP	0	0.9730	0.9600	0.5610	0.3542
HF	0	2.5140	2.3850	1.3600	1.8059
HF-HF	0	4.3790	4.1870	2.3840	3.3991
HN_3	0	2.0470	2.3340	2.1060	1.6603
HNC	0	2.7090	3.1300	2.4480	3.0818
HNCO	0	2.2530	2.9820	2.2570	2.0639
HNO	0	1.7270	2.8660	1.8610	1.6536
HNO_2	0	2.5880	3.6690	1.5300	1.9345
HNS	0	1.4520	1.9980	1.8370	1.4062
HO_2	1	2.9370	3.3380	2.0910	2.1659
HOCl	0	1.5840	2.4420	1.6260	1.5216
HOCN	0	4.0520	4.2440	3.4310	3.7998
HOF	0	2.3570	2.4980	1.6050	1.9168
HOOH	0	2.0350	2.4110	1.5760	1.5732
HPO	0	2.4570	4.0430	3.1330	2.6291
$LiBH_4$	0	5.7880	5.2390	7.4990	6.1281
LiCl	0	6.4350	6.6300	7.8990	7.0960
LiCN	0	6.9370	6.5560	7.4270	6.9851
LiF	0	6.2180	5.6960	6.4710	6.2879
LiH	0	6.2200	6.3000	3.8400	5.8286
LiN	2	6.2690	5.8470	7.0160	7.0558
LiOH	0	3.8580	3.0640	4.3200	4.5664
N_2H_2	0	2.8770	3.7340	3.4080	2.8771
N_2H_4	0	3.0760	3.7680	3.7040	2.7179
NaCl	0	7.3410	9.8670	9.8760	9.0066
NaCN	0	8.0880	9.1610	9.0750	8.8903
NaF	0	7.5970	8.0630	8.3250	8.1339
NaH	0	7.0670	7.9210	5.8930	6.3966
NaLi	0	1.9860	3.2990	6.2780	0.4837
NaOH	0	5.7750	5.6160	6.1820	6.7690
NCl	2	2.4670	2.5910	1.8780	1.1279
NCO	1	0.9960	0.2820	1.4860	0.7935
NF	2	0.5030	0.4910	0.6230	0.0671

NF_2	1	0.2140	0.1440	0.6670	0.1904
NH	2	1.7030	2.2410	1.8270	1.5433
NH_2	1	1.9950	2.5490	2.3290	1.7853
NH_2Cl	0	1.7410	2.4130	2.4130	1.9468
NH_2F	0	2.6380	2.9490	2.3520	2.2688
NH_2OH	0	0.8060	1.3610	0.9420	0.7044
NH_3	0	1.8390	2.1680	2.3150	1.5289
NH_3-BH_3	0	6.1630	6.2280	5.8760	5.2810
NH_3-NH_3	0	2.2460	2.6630	2.7650	2.1345
NH_3O	0	6.4900	7.2100	6.3200	5.3942
NO	1	0.1730	0.6150	0.7770	0.1271
NO_2	1	0.9040	1.4740	0.5770	0.3350
NOCl	0	0.2430	0.7030	3.1080	2.0773
NP	0	1.9520	4.0770	2.6680	2.8713
NS	1	2.0260	1.9780	1.8140	1.8237
O_3	0	1.1780	1.3300	2.0050	0.5666
OCl	1	2.3060	3.5640	2.0680	1.2790
OCl_2	0	0.8550	1.8340	0.8530	0.5625
OF	1	0.5810	1.1200	0.4650	0.0205
OF_2	0	0.4580	0.1690	0.5870	0.3252
OH	1	2.0750	2.5100	1.4920	1.6550
P_2H_4	0	1.6320	1.4440	2.7680	0.9979
PCl	2	0.6730	1.0530	0.3530	0.5657
PF	2	0.8810	1.1920	0.8570	0.8104
PH	2	1.1180	0.6230	1.4470	0.4375
PH_2	1	1.2320	0.8210	1.8020	0.5472
PH_2OH	0	0.5050	1.2450	0.7790	0.6836
PH_3	0	1.1930	1.0330	1.9480	0.6069
PH_3O	0	3.2420	4.4790	5.5290	3.7704
PO	1	1.7150	3.4710	1.9100	1.9617
PO_2	1	1.1350	2.1950	2.1090	1.4426
PPO	0	2.0310	3.2430	0.9870	1.8812
PS	1	0.8040	1.8950	0.5970	0.6825

S_2H_2	0	1.8460	1.5780	1.3300	1.1425
SCl	1	1.3440	0.4260	0.0500	0.0690
SCl_2	0	0.5170	0.6470	0.7220	0.3891
SF	1	0.8180	1.1010	0.9750	0.8139
SF_2	0	1.3920	1.5520	1.5240	1.0555
SH	1	1.4420	1.2010	0.7880	0.7727
SH_2	0	1.9480	1.5860	1.3710	0.9939
SiH	1	0.2730	0.6130	0.6440	0.1138
SiH_3Cl	0	1.7150	1.0290	1.4490	1.3645
SiH_3F	0	1.3580	0.9240	1.3790	1.3123
SiO	0	3.0560	5.1760	5.3940	3.1123
SO_2	0	2.9590	2.3450	3.3750	1.6286
SO-triplet	2	1.9100	2.4540	2.1480	1.5606

3 Element-specific parameters in GFN2-xTB

Table S48: Element-specific atomic parameters employed in GFN2-xTB: atomic Hubbard parameter (η_A), its charge derivative (Γ_A), the exponential scaling parameter α_A and Y_A^{eff} (both entering the repulsion potential), the anisotropic XC scaling parameters $f_{\text{XC}}^{\mu_A}$ and $f_{\text{XC}}^{\Theta_A}$, and the offset radius R_0^A for the damping in the AES energy. All quantities are given in atomic units.

element	η_A	Γ_A	α_A	Y_A^{eff}	$f_{\text{XC}}^{\mu_A}$	$f_{\text{XC}}^{\Theta_A}$	R_0^A
H	0.405771	0.08	2.213717	1.105388	0.0556389	0.00027431	1.4
He	0.642029	0.2	3.604670	1.094283	-0.01	-0.00337528	3.0
Li	0.245006	0.130382	0.475307	1.289367	-0.005	0.0002	5.0
Be	0.684789	0.0574239	0.939696	4.221216	-0.00613341	-0.00058586	5.0
B	0.513556	0.0946104	1.373856	7.192431	-0.00481186	-0.00058228	5.0
C	0.538015	0.15	1.247655	4.231078	-0.00411674	0.00213583	3.0
N	0.461493	-0.063978	1.682689	5.242592	0.0352127	0.0202679	1.9
O	0.451896	-0.0517134	2.165712	5.784415	-0.0493567	-0.00310828	1.8
F	0.531518	0.142621	2.421394	7.021486	-0.0833918	-0.00245955	2.4
Ne	0.850000	0.05	3.318479	11.041068	0.1	-0.005	5.0
Na	0.271056	0.179873	0.572728	5.244917	0	0.0002	5.0
Mg	0.344822	0.234916	0.917975	18.083164	-0.00082005	-0.00005516	5.0
Al	0.364801	0.14	0.876623	17.867328	0.0263334	-0.00021887	5.0
Si	0.720000	0.193629	1.187323	40.001111	-0.0002575	-0.0008	3.9
P	0.297739	0.0711291	1.143343	19.683502	0.0211022	0.00028679	2.1
S	0.339971	-0.0501722	1.214553	14.995090	-0.00151117	0.00442859	3.1
Cl	0.248514	0.149548	1.577144	17.353134	-0.0253696	0.00122783	2.5
Ar	0.502376	-0.0315455	0.896198	7.266606	-0.0207733	-0.010834	5.0
K	0.247602	0.203309	0.482206	10.439482	-0.00103383	0.00025	5.0
Ca	0.320378	0.20069	0.683051	14.786701	-0.00236675	0.0001	5.0
Sc	0.472633	0.05	0.574299	8.004267	-0.00515177	-0.00042004	5.0
Ti	0.513586	0.176727	0.723104	12.036336	-0.00434506	0.0005966	5.0
V	0.589187	0.09	0.928532	15.677873	-0.0035	0.00009764	5.0
Cr	0.396299	0.03	0.966993	19.517914	0.00149669	0.00137744	5.0
Mn	0.346651	0.06	1.071100	18.760605	-0.00759168	0.00229903	5.0

Fe	0.271594	-0.05	1.113422	20.360089	0.00412929	0.00267734	5.0
Co	0.477760	0.03	1.241717	27.127744	-0.00247938	0.00048237	5.0
Ni	0.344970	-0.02	1.077516	10.533269	-0.0126189	-0.0008	5.0
Cu	0.202969	0.05	0.998768	9.913846	-0.007	-0.00345631	5.0
Zn	0.564152	0.23129	1.160262	22.099503	-0.001	0.00007658	5.0
Ga	0.432236	0.233427	1.122923	31.146750	0.00267219	-0.00003616	5.0
Ge	0.802051	-0.0064775	1.222349	42.100144	0.0010846	-0.00003589	5.0
As	0.571748	0.110604	1.249372	39.147587	-0.00201294	0.00014149	5.0
Se	0.235052	0.0913725	1.230284	27.426779	-0.00288648	0.00085728	3.9
Br	0.261253	0.13	1.296174	32.845361	-0.0108859	0.00216935	4.0
Kr	0.424373	0.0239815	0.908074	17.363803	-0.00889357	-0.00415024	5.0
Rb	0.210481	0.29162	0.574054	44.338211	-0.00093328	0.00015	5.0
Sr	0.340000	0.18	0.697345	34.365525	-0.00459925	0.00015	5.0
Y	0.711958	0.01	0.706172	17.326237	-0.00637291	0.0001046	5.0
Zr	0.461440	0.07	0.681106	24.263093	-0.00599615	-0.00012944	5.0
Nb	0.952957	0.05	0.865552	30.562732	-0.00288729	0.00041491	5.0
Mo	0.586134	0.0919928	1.034519	48.312796	0.00346327	0.00312549	5.0
Tc	0.368054	0.06	1.019565	44.779882	-0.00458416	0.00155242	5.0
Ru	0.711205	-0.05	1.031669	28.070247	-0.00081922	0.00359228	5.0
Rh	0.509183	0.03	1.094599	38.035941	0.00007016	0.0000857	5.0
Pd	0.273310	0.08	1.092745	28.674700	-0.00310361	-0.00040485	5.0
Ag	0.263740	0.02	0.678344	6.493286	-0.00800314	-0.0002081	5.0
Cd	0.392012	0.207322	0.936236	26.226628	-0.00105364	0.0001225	5.0
In	0.461812	0.19	1.024007	63.854240	0.00951079	-0.00002031	5.0
Sn	0.900000	-0.0178396	1.139959	80.053438	0.00085029	-0.00008243	5.0
Sb	0.942294	0.11	1.122937	77.057560	-0.00015519	-0.0002063	5.0
Te	0.750000	0.0953683	1.000712	48.614745	-0.00263414	-0.00026864	5.0
I	0.383124	0.12	1.017946	63.319176	-0.00603648	0.0006966	5.0
Xe	0.424164	-0.0118925	1.012036	51.188398	-0.00214447	-0.001562	5.0
Cs	0.236569	0.240419	0.585257	67.249039	-0.0008	0.00008	5.0
Ba	0.245937	0.20691	0.716259	46.984607	-0.0026	0.00015	5.0
La	0.597716	0.0012793	0.737643	50.927529	-0.00395198	-0.0003	5.0
Ce	0.662889	-0.01	0.729950	48.676714	-0.00723806	-0.00025	5.0

Pr	0.660710	-0.0100002	0.734624	47.669448	-0.00704819	-0.00024615	5.0
Nd	0.658531	-0.0100004	0.739299	46.662183	-0.00685832	-0.00024231	5.0
Pm	0.656352	-0.0100006	0.743973	45.654917	-0.00666845	-0.00023846	5.0
Sm	0.654173	-0.0100008	0.748648	44.647651	-0.00647858	-0.00023462	5.0
Eu	0.651994	-0.010001	0.753322	43.640385	-0.00628871	-0.00023077	5.0
Gd	0.649815	-0.0100012	0.757996	42.633120	-0.00609884	-0.00022692	5.0
Tb	0.647635	-0.0100013	0.762671	41.625854	-0.00590897	-0.00022308	5.0
Dy	0.645456	-0.0100015	0.767345	40.618588	-0.0057191	-0.00021923	5.0
Ho	0.643277	-0.0100017	0.772020	39.611322	-0.00552923	-0.00021538	5.0
Er	0.641098	-0.0100019	0.776694	38.604057	-0.00533936	-0.00021154	5.0
Tm	0.638919	-0.0100021	0.781368	37.596791	-0.00514949	-0.00020769	5.0
Yb	0.636740	-0.0100023	0.786043	36.589525	-0.00495961	-0.00020385	5.0
Lu	0.634561	-0.0100025	0.790717	35.582259	-0.00476974	-0.0002	5.0
Hf	0.662597	-0.01	0.852852	40.186772	-0.00537685	-0.00016478	5.0
Ta	0.449812	0.02	0.990234	54.666156	-0.00200343	0.00039599	5.0
W	0.685426	-0.02	1.018805	55.899801	0.00065886	0.0106331	5.0
Re	0.224623	0.08	1.170412	80.410086	-0.00587636	0.0030687	5.0
Os	0.364388	0.08	1.221937	62.809871	-0.0051009	0.00759049	5.0
Ir	0.548507	-0.01	1.197148	56.045639	-0.00673822	0.00322935	5.0
Pt	0.353574	0.06	1.204081	53.881425	-0.00423684	0.00098019	5.0
Au	0.438997	0.085	0.919210	14.711475	0.00393418	-0.0002032	5.0
Hg	0.457611	-0.0116312	1.137360	51.577544	-0.0025	-0.00032901	5.0
Tl	0.418841	-0.0533933	1.399312	58.801614	0.00374018	-0.00008506	5.0
Pb	0.168152	0.02	1.179922	102.368258	0.0100702	-0.0000167	5.0
Bi	0.900000	-0.0337508	1.130860	132.896832	-0.00737252	0.00162529	5.0
Po	1.023267	0.187798	0.957939	52.301232	-0.0134485	0.00013818	5.0
At	0.288848	0.184648	0.963878	81.771063	-0.00348123	0.00021624	5.0
Rn	0.303400	0.0097834	0.965577	128.133580	-0.00167597	-0.00111556	5.0

[a] It is noted that R_0^A is a fitted parameter only for 12 elements and set to a value of 5.0 for the rest of the periodic table.

Table S49: Element-specific shell parameters employed in GFN2-xTB: the polynomial scaling parameters $k_{A,l}^{poly}$, the shell-specific scaling parameters of the Hubbard parameter κ_A^l, the CN'_A dependent enhancement factors for the energy levels, the constant part of the energy levels (H_A^l), and the corresponding Slater exponents ζ_l. The energy levels and their CN'_A dependent enhancement factors are given in eV, ζ_l is given in atomic units, whereas $k_{A,l}^{poly}$ and κ_A^l are dimensionless.

element	level	$k_{A,l}^{poly}$	${\kappa_A^l}^{(a)}$	$H_{CN'_A}^l$ / eV	H^l / eV	ζ_l
H	1s	-0.00953618	0.0	-0.05	-10.707211	1.230000
He	1s	-0.0438682	0.0	0.207428	-23.716445	1.669667
	2p	0.00710647	0.0	0.0	-1.822307	1.500000
Li	2s	-0.047504	0.0	0.162084	-4.900000	0.750060
	2p	0.204249	0.197261	-0.0623876	-2.217789	0.557848
Be	2s	-0.0791039	0.0	0.118776	-7.743081	1.034720
	2p	-0.00476438	0.965847	0.0550528	-3.133433	0.949332
B	2s	-0.0518315	0.0	0.0120462	-9.224376	1.479444
	2p	-0.0245332	0.399408	-0.0141086	-7.419002	1.479805
C	2s	-0.0229432	0.0	-0.0102144	-13.970922	2.096432
	2p	-0.00271102	0.105636	0.0161657	-10.063292	1.800000
N	2s	-0.08506	0.0	-0.195534	-16.686243	2.339881
	2p	-0.025042	0.116489	0.0561076	-12.523956	2.014332
O	2s	-0.149553	0.0	0.0117826	-20.229985	2.439742
	2p	-0.0335082	0.149702	-0.0145102	-15.503117	2.137023
F	2s	-0.130119	0.0	0.0394362	-23.458179	2.416361
	2p	-0.123008	0.167738	-0.0538373	-15.746583	2.308399
Ne	2s	-0.163778	0.0	-0.0014933	-24.500000	3.084104
	2p	-0.0486055	0.119058	0.0232093	-18.737298	2.312051
	3d	-0.169223	-0.32	0.109671	-5.517827	2.815609
Na	3s	-0.040335	0.0	-0.0042211	-4.546934	0.763787
	3p	0.208739	0.101889	-0.0144323	-1.332719	0.573553
Mg	3s	-0.111674	0.0	0.116444	-6.339908	1.184203
	3p	0.39077	1.4	-0.0079924	-0.697688	0.717769
	3d	0.126911	-0.05	0.119241	-1.458197	1.300000
Al	3s	-0.106781	0.0	0.0715422	-9.329017	1.352531

	3p	-0.124428	-0.0603699	-0.0244485	-5.927846	1.391201
	3d	0.163111	0.2	0.0406173	-3.042325	1.000000
Si	3s	0.0235852	0.0	0.185848	-14.360932	1.773917
	3p	-0.0790041	-0.558004	-0.138307	-6.915131	1.718996
	3d	0.113662	-0.23	-0.193549	-1.825036	1.250000
P	3s	-0.198318	0.0	0.054761	-17.518756	1.816945
	3p	-0.0551558	-0.155806	-0.048993	-9.842286	1.903247
	3d	0.263975	-0.35	0.242951	-0.444893	1.167533
S	3s	-0.258555	0.0	-0.0256951	-20.029654	1.981333
	3p	-0.0804806	-0.108587	-0.0098465	-11.377694	2.025643
	3d	0.259939	-0.25	0.200769	-0.420282	1.702555
Cl	3s	-0.16562	0.0	0.0617972	-29.278781	2.485265
	3p	-0.0698643	0.49894	-0.0181618	-12.673758	2.199650
	3d	0.380456	0.5	0.167277	-0.240338	2.476089
Ar	3s	-0.238939	0.0	0.0000554	-16.487730	2.329679
	3p	-0.0372732	-0.0461133	0.0065921	-13.910539	2.149419
	3d	0.268129	-0.01	-0.273217	-1.167213	1.950531
K	4s	-0.0607606	0.0	-0.0339245	-4.510348	0.875961
	4p	0.211873	0.348366	0.0174542	-0.934377	0.631694
Ca	4s	-0.0971872	0.0	0.057093	-5.056506	1.267130
	4p	0.319734	1.5	-0.0074926	-1.150304	0.786247
	3d	0.0952865	-0.25	0.101375	-0.776883	1.380000
Sc	3d	-0.345023	-0.08	0.202678	-5.196187	2.440000
	4s	0.00686569	0.0	0.0991293	-8.877940	1.358701
	4p	0.380449	-0.204672	-0.0281241	-2.008206	1.019252
Ti	3d	-0.277244	-0.38	0.102819	-7.234331	1.849994
	4s	0.0456123	0.0	0.100702	-10.900000	1.469983
	4p	0.518016	-0.492111	-0.0237074	-1.928783	0.957410
V	3d	-0.298276	-0.45	0.0164476	-9.015342	1.673577
	4s	0.0970248	0.0	0.0235696	-9.573347	1.383176
	4p	0.511783	-0.0379088	-0.0108232	-0.706647	0.938025
Cr	3d	-0.279716	-0.47	0.0289291	-7.209794	1.568211
	4s	0.133762	0.0	-0.0232087	-9.201304	1.395427

	4p	0.480922	0.740587	-0.0188919	-0.696957	1.080270
Mn	3d	-0.312559	-0.6	-0.0195827	-10.120933	1.839250
	4s	0.285197	0.0	-0.0275	-5.617346	1.222190
	4p	0.263466	0.0545811	-0.0015839	-4.198724	1.240215
Fe	3d	-0.28615	-0.65	-0.0274654	-10.035473	1.911049
	4s	0.115278	0.0	-0.404988	-5.402911	1.022393
	4p	0.394599	0.404662	-0.075648	-3.308988	1.294467
Co	3d	-0.223556	-0.65	0.012198	-10.580430	2.326507
	4s	0.0916846	0.0	-0.0227872	-8.596723	1.464221
	4p	0.254247	-0.241849	0.0076513	-2.585753	1.298678
Ni	3d	-0.253856	-0.6	-0.0066417	-12.712236	2.430756
	4s	0.208395	0.0	0.0310301	-8.524281	1.469945
	4p	0.308864	-0.0611188	0.0226796	-2.878873	1.317046
Cu	3d	-0.265089	0.07	-0.0173684	-9.506548	2.375425
	4s	0.177983	0.0	0.334905	-6.922958	1.550837
	4p	0.149778	1.33331	-0.261945	-2.267723	1.984703
Zn	4s	-0.0924032	0.0	0.201191	-7.177294	1.664847
	4p	0.222718	0.0684343	-0.0055135	-0.991895	1.176434
Ga	4s	-0.190182	0.0	-0.0234627	-12.449656	1.720919
	4p	-0.0113779	-0.541655	0.130583	-4.469873	1.591570
	4d	0.354019	-0.3	0.0165604	-0.582255	1.050000
Ge	4s	-0.213337	0.0	0.0361068	-16.369792	1.990429
	4p	-0.0974904	-0.380909	-0.0014474	-8.207673	1.830340
	4d	0.286347	-0.15	-0.104256	-0.994226	1.100000
As	4s	-0.238207	0.0	-0.012964	-16.421504	2.026128
	4p	-0.106442	-0.410474	-0.023647	-9.311147	1.949257
	4d	0.307111	-0.5	0.233014	-0.276830	1.040181
Se	4s	-0.245064	0.0	-0.0061654	-20.584732	2.230969
	4p	-0.137658	0.119211	-0.0435018	-10.910799	2.150656
	4d	0.296111	-0.25	0.276856	-0.110636	1.317549
Br	4s	-0.250051	0.0	0.000615	-23.583718	2.077587
	4p	-0.145201	0.5203	-0.0058347	-12.588824	2.263120
	4d	0.36614	0.4	0.225018	0.047980	1.845038

Kr	4s	-0.326587	0.0	-0.0070305	-17.221422	2.445680
	4p	-0.136001	-0.250322	0.0076023	-13.633377	2.210494
	4d	0.232047	-0.07	0.0349523	-0.940657	1.884991
Rb	5s	0.043254	0.0	-0.151693	-4.353793	1.017267
	5p	0.232551	0.938649	0.0203437	-1.392938	0.870130
Sr	5s	-0.145068	0.0	0.040902	-6.291692	1.419028
	5p	0.20214	1.5	-0.0418725	-1.872475	0.928932
	4d	0.108162	-0.25	0.0401255	-0.890492	1.500000
Y	4d	-0.395295	-0.45	-0.127034	-8.015206	2.670141
	5s	-0.0212587	0.0	0.193752	-12.194181	1.633876
	5p	0.521619	-0.334929	-0.0641897	-0.966195	1.165412
Zr	4d	-0.283589	-0.11	-0.0566943	-7.409832	2.238668
	5s	0.075389	0.0	0.126655	-10.199105	1.702480
	5p	0.589141	-0.442263	0.0279435	-1.066939	1.129590
Nb	4d	-0.279637	-0.05	-0.135649	-8.440821	1.706832
	5s	-0.0514108	0.0	0.255596	-11.384021	1.666463
	5p	0.556542	-0.356295	-0.0002341	-0.103760	1.132172
Mo	4d	-0.225737	-0.3	0.0620172	-7.995133	1.777658
	5s	-0.00583137	0.0	0.300841	-7.336245	1.639917
	5p	0.291996	-0.430137	-0.104035	-3.686225	1.159781
Tc	4d	-0.273426	-0.6	-0.0066526	-9.587897	1.918066
	5s	0.36096	0.0	-0.0586205	-6.792444	1.918167
	5p	0.250957	0.395682	-0.0087319	-3.325525	1.346082
Ru	4d	-0.275832	-0.65	-0.0263914	-10.285405	2.102697
	5s	0.101063	0.0	0.447116	-5.332608	1.749643
	5p	0.340287	-0.305231	-0.0034723	-3.307153	1.348322
Rh	4d	-0.196561	-0.65	0.0104368	-11.756644	2.458187
	5s	0.154133	0.0	0.0066741	-7.850495	1.811796
	5p	0.310707	-0.188177	-0.0213308	-3.007906	1.398452
Pd	4d	-0.271731	-0.6	0.0060285	-11.963518	2.353691
	5s	0.0620014	0.0	0.026682	-9.714059	1.828354
	5p	0.453413	0.0931707	0.0503075	-2.035281	1.333352
Ag	4d	-0.164907	-0.03	-0.0062719	-9.591083	2.843549

	5s	0.0109149	0.0	-0.0065794	-8.083960	1.798462
	5p	0.115614	0.802485	0.167717	-2.934333	1.266649
Cd	5s	-0.0607687	0.0	0.141815	-7.252341	1.846689
	5p	0.376719	0.238867	-0.0309814	-0.744865	1.141823
In	5s	-0.219385	0.0	-0.0098312	-13.040909	1.963283
	5p	-0.0194965	-0.586746	0.0994688	-4.507143	1.685138
	5d	0.313545	-0.28	0.0168649	-0.805666	1.050000
Sn	5s	-0.175182	0.0	-0.0454629	-19.970428	2.551510
	5p	-0.0780287	-0.509075	-0.0320651	-7.367059	1.893784
	5d	0.126111	-0.06	-0.145941	-2.077548	1.100000
Sb	5s	-0.175435	0.0	-0.0147626	-18.371244	2.307407
	5p	-0.124946	-0.62785	-0.0091175	-7.350148	2.179752
	5d	0.308727	-0.55	0.160287	0.909033	1.256087
Te	5s	-0.248939	0.0	0.0115389	-21.930653	2.434144
	5p	-0.11232	-0.155533	-0.0082051	-9.480374	2.182459
	5d	0.318432	0.06	0.301323	0.978922	1.373076
I	5s	-0.269575	0.0	-0.050615	-20.949407	2.159500
	5p	-0.141833	-0.0338735	0.0084766	-12.180159	2.308379
	5d	0.282119	0.3	0.307713	-0.266596	1.691185
Xe	5s	-0.310965	0.0	-0.0020195	-19.090498	2.715140
	5p	-0.161979	-0.230267	0.0017246	-11.249471	2.312510
	5d	0.19049	-0.23	0.0327039	-0.497097	1.855707
Cs	6s	-0.00713637	0.0	-0.13126	-4.041706	1.225688
	6p	0.20637	0.249431	-0.01	-1.394193	0.823818
Ba	6s	-0.140366	0.0	0.0352001	-5.900000	1.528102
	6p	0.187741	2.22475	-0.0926576	-2.133395	0.991572
	5d	0.113897	-0.23	0.0147995	-1.514900	1.500000
La	5d	-0.378201	-0.3	-0.0777542	-8.958783	2.875048
	6s	-0.0673201	0.0	0.107168	-11.877410	1.731390
	6p	0.541364	-0.469967	-0.0239967	-0.601717	1.303590
Ce	5d	-0.419892	-0.3	-0.0638958	-7.381991	2.870000
	6s	-0.0610774	0.0	0.133515	-8.537781	1.725197
	6p	0.376634	-0.553966	-0.019832	-3.017508	1.309804

Pr	5d	-0.412865	-0.276923	-0.0543909	-7.280875	2.872308
	6s	-0.0604017	0.0	0.134944	-8.504806	1.729767
	6p	0.381948	-0.546278	-0.0198184	-2.873159	1.315495
Nd	5d	-0.405838	-0.253846	-0.0448861	-7.179760	2.874615
	6s	-0.0597259	0.0	0.136373	-8.471830	1.734337
	6p	0.387261	-0.538591	-0.0198048	-2.728809	1.321186
Pm	5d	-0.398811	-0.230769	-0.0353812	-7.078644	2.876923
	6s	-0.0590501	0.0	0.137803	-8.438855	1.738907
	6p	0.392574	-0.530903	-0.0197912	-2.584460	1.326877
Sm	5d	-0.391784	-0.207692	-0.0258764	-6.977529	2.879231
	6s	-0.0583743	0.0	0.139232	-8.405879	1.743478
	6p	0.397888	-0.523216	-0.0197776	-2.440110	1.332567
Eu	5d	-0.384758	-0.184615	-0.0163715	-6.876413	2.881538
	6s	-0.0576986	0.0	0.140661	-8.372904	1.748048
	6p	0.403201	-0.515528	-0.019764	-2.295761	1.338258
Gd	5d	-0.377731	-0.161538	-0.0068667	-6.775298	2.883846
	6s	-0.0570228	0.0	0.142091	-8.339929	1.752618
	6p	0.408514	-0.507841	-0.0197504	-2.151411	1.343949
Tb	5d	-0.370704	-0.138461	0.0026382	-6.674182	2.886154
	6s	-0.056347	0.0	0.14352	-8.306953	1.757188
	6p	0.413827	-0.500153	-0.0197369	-2.007062	1.349640
Dy	5d	-0.363677	-0.115384	0.012143	-6.573067	2.888462
	6s	-0.0556712	0.0	0.144949	-8.273978	1.761758
	6p	0.419141	-0.492466	-0.0197233	-1.862712	1.355331
Ho	5d	-0.35665	-0.0923072	0.0216479	-6.471951	2.890769
	6s	-0.0549955	0.0	0.146379	-8.241003	1.766328
	6p	0.424454	-0.484778	-0.0197097	-1.718363	1.361022
Er	5d	-0.349623	-0.0692302	0.0311527	-6.370836	2.893077
	6s	-0.0543197	0.0	0.147808	-8.208027	1.770899
	6p	0.429767	-0.477091	-0.0196961	-1.574013	1.366713
Tm	5d	-0.342596	-0.0461533	0.0406576	-6.269720	2.895385
	6s	-0.0536439	0.0	0.149237	-8.175052	1.775469
	6p	0.435081	-0.469403	-0.0196825	-1.429664	1.372403

Yb	5d	-0.335569	-0.0230763	0.0501624	-6.168604	2.897692
	6s	-0.0529681	0.0	0.150667	-8.142076	1.780039
	6p	0.440394	-0.461716	-0.0196689	-1.285314	1.378094
Lu	5d	-0.328542	0.0000007	0.0596673	-6.067489	2.900000
	6s	-0.0522924	0.0	0.152096	-8.109101	1.784609
	6p	0.445707	-0.454028	-0.0196553	-1.140965	1.383785
Hf	5d	-0.340957	0.1	0.017655	-7.181755	2.638729
	6s	-0.0273193	0.0	0.22715	-10.626891	2.194333
	6p	0.33515	-0.448616	-0.0069771	-1.603430	1.427467
Ta	5d	-0.303963	0.05	-0.0620136	-8.481353	2.018969
	6s	-0.157077	0.0	0.0988501	-13.073088	1.996498
	6p	0.60186	-0.339438	-0.047254	0.655254	1.407714
W	5d	-0.256771	0.37	-0.0192494	-9.501505	2.155885
	6s	0.0620898	0.0	0.254364	-11.093016	1.892022
	6p	0.492738	-0.34192	0.0236479	-1.420389	1.458186
Re	5d	-0.317231	-0.6	-0.0322139	-11.189119	2.262783
	6s	0.138901	0.0	0.111757	-12.685198	2.187549
	6p	0.339733	0.658686	-0.133516	-3.851981	1.636996
Os	5d	-0.284611	-0.65	-0.0095346	-10.382841	2.509631
	6s	0.213168	0.0	0.0346183	-8.731460	2.173991
	6p	0.280972	0.135022	-0.0208758	-3.546379	1.597888
Ir	5d	-0.246934	-0.65	0.0051977	-11.018475	2.756134
	6s	0.207338	0.0	-0.0123672	-9.349164	2.117548
	6p	0.183032	-0.0977957	-0.0079864	-3.603762	1.680343
Pt	5d	-0.272439	-0.6	-0.0204828	-12.047728	2.704492
	6s	0.0673756	0.0	0.113953	-10.482306	2.329136
	6p	0.192595	-0.0203212	0.140803	-3.778297	1.623286
Au	5d	-0.0641082	-0.6	-0.0154462	-9.578599	3.241287
	6s	0.0469154	0.0	0.147934	-7.688552	2.183171
	6p	0.252503	0.0614126	0.104807	0.883399	2.084484
Hg	6s	-0.0983345	0.0	-0.0352252	-11.538066	2.244504
	6p	0.156289	-0.537512	0.0205401	-2.532581	1.470848
Tl	6s	-0.229422	0.0	-0.0255975	-17.319333	2.294231

	6p	0.131098	-0.71334	0.0901364	-4.460584	1.731592
Pb	6s	-0.229551	0.0	-0.389346	-24.055207	2.960592
	6p	-0.0880527	0.783825	0.343712	-5.893816	1.953130
Bi	6s	-0.217501	0.0	0.0160425	-19.843840	2.788267
	6p	-0.107739	-0.6	0.0248659	-7.297456	2.277039
Po	6s	-0.209233	0.0	-0.0046813	-20.205380	3.314810
	6p	-0.184264	-0.810916	-0.0100437	-8.476927	2.389456
At	6s	-0.3055	0.0	-0.0287369	-17.050229	2.220421
	6p	-0.171085	-0.253207	-0.0007993	-9.499822	2.408112
	5d	0.23825	0.25	0.280581	-0.096063	1.500000
Rn	6s	-0.352454	0.0	-0.0001712	-21.000000	3.109394
	6p	-0.119897	-0.0302388	-0.000528	-10.496406	2.541934
	5d	0.21167	-0.23	-0.320602	-1.415056	1.790000

An STO-3G expansion is used for 1s levels (H and He) and for all d levels .

An STO-4G expansion is used for the ns and np levels with $n = [2, 5]$.

An STO-6G expansion is used for the 6s and 6p levels.

$^{(a)}$ It is noted that for s-functions, κ_A^l is not a fitted parameter but always set to zero.

References

(S1) 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* **1998**, *58*, 7260–7268.

(S2) 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.

(S3) Grimme, S.; Bannwarth, C.; Shushkov, P. 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 ($Z = 1 - 86$). *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.

(S4) 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.

(S5) 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.

(S6) Bühl, M.; Kabrede, H. Geometries of Transition-Metal Complexes from Density-Functional Theory. *J. Chem. Theory Comput.* **2006**, *2*, 1282–1290.

(S7) 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.

(S8) 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.

(S9) Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 4285–4292.

(S10) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184–32215.

(S11) Kozuch, S.; Martin, J. M. L. Halogen Bonds: Benchmarks and Theoretical Analysis. *J. Chem. Theory Comput.* **2013**, *9*, 1918–1931.

(S12) Setiawan, D.; Kraka, E.; Cremer, D. Strength of the Pnicogen Bond in Complexes Involving Group Va Elements N, P, and As. *J. Phys. Chem. A* **2015**, *119*, 1642–1656.

(S13) Marshall, M. S.; Burns, L. A.; Sherrill, C. D. Basis set convergence of the coupled-cluster correction, $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$: 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.

(S14) Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **2016**, *116*, 5105–5154.

(S15) 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.

(S16) Anacker, T.; Friedrich, J. New accurate benchmark energies for large water clusters: DFT is better than expected. *J. Comput. Chem.* **2014**, *35*, 634–643.

(S17) Lao, K. U.; Schäffer, R.; Jansen, G.; Herbert, J. M. Accurate description of intermolecular interactions involving ions using symmetry-adapted perturbation theory. *J. Chem. Theory Comput.* **2015**, *11*, 2473–2486.

(S18) 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.

(S19) Sure, R.; Grimme, S. Comprehensive Benchmark of Association (Free) Energies of Realistic Host–Guest Complexes. *J. Chem. Theory Comput.* **2015**, *11*, 3785–3801.

(S20) 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.

(S21) Kesharwani, M. K.; Karton, A.; Martin, J. M. L. Benchmark ab Initio Conformational Energies for the Proteinogenic Amino Acids through Explicitly Correlated Methods. Assessment of Density Functional Methods. *J. Chem. Theory Comput.* **2016**, *12*, 444–454.

(S22) Kozuch, S.; Bachrach, S. M.; Martin, J. M. Conformational Equilibria in Butane-1,4-diol: A Benchmark of a Prototypical System with Strong Intramolecular H-bonds. *J. Phys. Chem. A* **2014**, *118*, 293–303.

(S23) 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.

(S24) 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.

(S25) Goerigk, L.; Karton, A.; Martin, J. M. L.; Radom, L. Accurate quantum chemical energies for tetrapeptide conformations: why MP2 data with an insufficient basis set should be handled with caution. *Phys. Chem. Chem. Phys.* **2013**, *15*, 7028–7031.

(S26) Csonka, G. I.; French, A. D.; Johnson, G. P.; Stortz, C. A. Evaluation of Density Functionals and Basis Sets for Carbohydrates. *J. Chem. Theory Comput.* **2009**, *5*, 679–692.

(S27) 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.

(S28) Marianski, M.; Supady, A.; Ingram, T.; Schneider, M.; Baldauf, C. Assessing the Accuracy of Across-the-Scale Methods for Predicting Carbohydrate Conformational Energies for the Examples of Glucose and α-Maltose. *J. Chem. Theory Comput.* **2016**, *12*, 6157–6168.

(S29) Dohm, S.; Hansen, A.; Steinmetz, M.; Grimme, S.; Checinski, M. P. Comprehensive Thermochemical Benchmark Set of Realistic Closed-Shell Metal Organic Reactions. *J. Chem. Theory Comput.* **2018**, *14*, 2596–2608.

(S30) Karton, A.; Goerigk, L. Accurate reaction barrier heights of pericyclic reactions: Surprisingly large deviations for the CBS-QB3 composite method and their consequences in DFT benchmark studies. *J. Comput. Chem. 36*, 622–632.

(S31) Goerigk, L.; Sharma, R. The INV24 test set: how well do quantum-chemical methods describe inversion and racemization barriers? *Can. J. Chem.* **2016**, *94*, 1133–1143.

(S32) Karton, A.; O'Reilly, R. J.; Chan, B.; Radom, L. Determination of Barrier Heights for Proton Exchange in Small Water, Ammonia, and Hydrogen Fluoride Clusters with G4(MP2)-Type, MPn, and SCS-MPn Procedures–A Caveat. *J. Chem. Theory Comput.* **2012**, *8*, 3128–3136.

(S33) Karton, A.; O'Reilly, R. J.; Radom, L. Assessment of Theoretical Procedures for Calculating Barrier Heights for a Diverse Set of Water-Catalyzed Proton-Transfer Reactions. *J. Phys. Chem. A* **2012**, *116*, 4211–4221.

(S34) Hait, D.; Head-Gordon, M. How Accurate Is Density Functional Theory at Predicting Dipole Moments? An Assessment Using a New Database of 200 Benchmark Values. *J. Chem. Theory Comput.* **2018**, *14*, 1969–1981.