Conformational energy benchmark for longer n-alkane chains

Sebastian Ehlert, Stefan Grimme, and Andreas Hansen*

Mulliken Center for Theoretical Chemistry, Institute for Physical and Theoretical Chemistry, University of Bonn, Beringstr. 4, 53115 Bonn, Germany

E-mail: hansen@thch.uni-bonn.de

Abstract

We present the first benchmark set focusing on the conformational energies of highly flexible, long *n*-alkane chains, termed ACONFL. Unbranched alkanes are ubiquitous building blocks in nature, so the goal is to be able to calculate their properties most accurately to improve the modeling of, e.g, complex (biological) systems. Very accurate DLPNO-CCSD(T1)/CBS reference values are provided, which allow for a statistical meaningful evaluation of even the best available density functional methods. The performance of established and modern (dispersion corrected) density functionals is comprehensively assessed. The recently introduced r^{2}SCAN-V functional shows excellent performance, similar to efficient composite DFT methods like B97-3c and r^{2}SCAN-3c, which provide an even better cost-accuracy ratio, while almost reaching the accuracy of much more computationally demanding hybrid or double hybrid functionals with large QZ AO basis sets. In addition, we investigated the performance of common wavefunction methods, where MP2/CBS surprisingly performs worse compared to simple D4 dispersion corrected Hartree–Fock. Furthermore, we investigate the performance of several semiempirical and force field methods, which are commonly used for the generation of conformational ensembles in multilevel workflows or in large scale molecular dynamics studies. Outstanding performance is obtained by the recently introduced general force field, GFN-FF, while other commonly applied methods like the universal force field yield large errors. We recommend the ACONFL as a helpful benchmark set for parameterization of new semiempirical or force field methods and machine learning potentials as well as a meaningful validation set for newly developed DFT or dispersion methods.

1 Introduction

Conformers are defined as distinct minima on a molecular potential energy surface with fixed covalent topology and can be converted into each other by rotations about formally single bonds. Their structure and relative (conformational) energies are of great impor-

tance in organic molecules[1] and for biological activity.[2] This is especially true for open-chain compounds, as they usually feature many possible internal rotation axes. Such flexible molecules are key for targeting compounds with specific spatial properties and understanding how folding of biological polymers and peptides is controlled.[3] Often, many conformers cover a rather narrow energy range, so that these systems exist as a thermally populated mixture of conformers at room or physiological temperature. Since measured equilibrium properties correspond to the Boltzmann average of the microstates, representing all relevant molecular structures is essential for a reliable prediction of molecular properties.[4,5]

Unbranched *n*-alkanes, which are ubiquitous in nature, are the simplest examples of this class of molecules, so the goal is to be able to calculate their properties most accurately to improve the modeling of complex (biological) systems such as membrane proteins[6] and to allow for more reliable protein ligand docking libraries.[7] *n*-alkanes and aliphatic chains in general are of particular importance, not only as basic building blocks of organic chemistry and part of fossil fuels, but also as components of lipids and polymers.

The existence of multiple conformers for *n*-alkanes for $n > 3$ has been known since the seminal work of Pitzer published 85 years ago.[8] Since then, a number of experimental and theoretical studies have been carried out with the aim of determining conformation energies, i. e. the energy difference between two different conformers, of *n*-alkanes as accurately as possible. Shorter *n*-alkanes are typical subjects of experimental studies investigating conformational enthalpies[9] or low-energy conformers.[10,11] Also theoretical investigations of torsional and conformational energies feature mainly short alkane chains.[12–14] For longer alkane chains (n$>$ 10), previous theoretical studies focus mainly on the lowest lying conformer,[15–18] investigating the balance between repulsive hydrogen contacts and attractive London-dispersion to predict the change from the linear zig-zag conformation to a hairpin like, closed conformer.

The most comprehensive theoretical studies on conformational energies of *n*-alkanes

up to *n*-hexane were carried out by Gruzman et al.,[19] augmented in a later by Martin with a detailed potential energy surface investigation of *n*-pentane beyond equilibrium structures.[20] However, the conformational ensemble of short *n*-alkanes chain is not fully representative for longer *n*-alkanes, which have particular relevance for modeling biological systems, e. g. in lipids, due to the much stronger attractive London-dispersion.[16–18]

While *n*-alkanes are seemingly simple systems for wavefunction theory (WFT) due to the large HOMO-LUMO gap and absence of static correlation, the need for large and diffuse basis sets to accurately capture long-range dynamic correlation effects, i. e. London dispersion, and to reduce the residual intramolecular basis set superposition error (BSSE), which cannot be removed by standard counterpoise correction schemes, posed a computational challenge. Moreover, since the differences in conformational energies should be described with an accuracy of at least 0.1 kcal/mol to allow a correct assessment of the conformational order or to calculate the Boltzmann populations reasonably well at room temperature,[5] sophisticated WFT methods such as CCSD(T) are essential for a reliable theoretical reference.

To provide this level of accuracy, the cumulative medium and long range intramolecular nonconvalent interactions (NCI) in long saturated chains need to be described accurately at the same footings.[19,21–23] This balance is problematic for density functional theory (DFT) as has been, e.g., shown for 1,3 interactions in branched alkanes by one the present authors[24] even if London dispersion is captured by a suitable dispersion correction. To assess the description of intramolecular NCIs in semilocal density functionals and semiempirical methods, various related conformational benchmarks sets were devised, e.g., for melatonin,[25] butane-1,4-diol,[26] RNA backbone models,[27] amino acids,[28] and many more, for example composed in the GMTKN55 database.[29] The latter also includes the ACONF set, which comprises 15 relative energies of n-butane, n-pentane and n-hexane conformers taken from the work of Gruzman at al.[19]

Nowadays, the long-range NCI problem for DFT, semiempirical quantum mechan-

ics (SQM) and also force fields methods is largely solved,[30–32] but this has not yet been extensively assessed for conformer ensembles of longer *n*-alkane chains. Moreover, especially in folded *n*-alkane chains, many NCI contacts of H atoms due to Pauli repulsion become important and existing theoretical studies focusing on shorter *n*-alkanes or just equilibrium conformer structures could not evaluate these interactions comprehensively. The recently introduced combined tools GFN-FF,[33] GFN*n*-xTB,[34–36] CREST,[37] and CENSO[5] have filled a gap in the field of quantum chemical modeling, specifically for generating conformer ensembles of larger molecules. Here, we make use of these new capabilities to generate suitable conformer ensembles of longer *n*-alkanes, which we have combined into a new benchmark set termed ACONFL (see 3. Recent developments of accurate low-order scaling local coupled cluster methods[38] enabled us to generate high level theoretical reference values close to the basis set limit suitable for a statistical meaningful evaluation of much more approximate methods.

After summarizing the computational details in the next section, the generation of our new benchmark set and its reference values will be described followed by an extensive evaluation of various FF, SQM, DFT and WFT methods. Finally, general conclusions and method recommendations will be given.

2 Computational details

Conformer ensembles were obtained with the advanced conformer rotamer ensemble sampling tool[5,37,39] (`crest`, version 4) employing the GFN-FF[40] method and default settings. Subsequently, selected conformers (see 3.1) were re-optimized at the B97-3c[41] level of theory utilizing the `Turbomole` program package version 7.5.1.[42,43]

The `ORCA` program package version 5.0.1[38,44] was used to perform the calculations with double hybrid functionals, MP3, DLPNO-CCSD and the meta-GGA B97M, while the Hartree-Fock (HF), second order Møller-Plesset perturbation theory (MP2) and local

coupled cluster and calculations were carried out employing ORCA 4.2.1[45]. All other DFT calculations were executed with Turbomole 7.5.1. MP2D[46] calculations were conducted with Psi4.[47] κOO-MP2 ($\kappa = 1.1$)[48] and MP2.5[49] were evaluated with QChem 5.4.2.[50] The resolution of identity (RI) method was employed to accelerate the evaluation of Coulomb (RIJ) and exchange integrals (RIJK).[51,52] Except for the "3c" methods, which use the respective stripped and optimized basis sets, Ahlrichs' type quadruple-ζ def2-QZVPP[53] basis sets with matching auxiliary basis sets for RIJ and RIJK[54,55] were applied in the DFT calculations. The numerical quadrature grid option *DefGrid3* and *TightSCF* convergence criteria were applied as implemented in ORCA 5.0.1, while the *m4* grid was used in the Turbomole calculations.

The RI and frozen core approximations for the correlation part as well as *TightSCF* convergence criteria for the HF part as implemented in ORCA 4.2.1 were employed. The domain based, local pair natural orbital coupled cluster method[56] in its ORCA 4.2.1 closed-shell, sparse maps[57] iterative triples[58] implementation (DLPNO-CCSD(T1)) together with *VeryTightPNO* threshold settings (i.e. ORCA 4.2.1 *TightPNO* settings with *TCutMKN*, *TCutPNO*, and *TCutPairs* tightened to 10^{-4}, 10^{-8}, and 10^{-6}, respectively) was applied. An aug-cc-pVTZ/aug-cc-pVQZ[59] complete basis set (CBS) extrapolation according to the scheme introduced by Helgaker and Klopper[60] was carried out separately for the HF and correlation energy for all MP2 and parts of DLPNO-CCSD(T1) energies. The same extrapolation scheme was used for the MP2/CBS and PWPB95-D4/CBS energies. For all other DLPNO-CCSD(T) energies, a special CBS extrapolation scheme (see Section 3.2) was employed.

DFTB calculations were conducted with the DFTB+ program package (version 21.1),[61,62] the *3ob* parameterization[63–66] was used for the third-order DFTB Hamiltonian, while the *mio* parameterization[67–69] was used with the second-order DFTB Hamiltonian. The LC-DFTB Hamiltonian was applied with the *ob2-base* parameterization.[70] The GFN1-xTB,[34] GFN2-xTB,[35] and GFN-FF[40] calculations were carried out using the xtb program (version 6.4.1)[36,71]. The MOPAC program (version 2016)[72] was used to perform PM6-D3H4[73]

and PM7 calculations.[74] The SMIRKS Native Open Force Field (SMIRNOFF) based methods[75] were used to evaluate the SMIRNOFF99Frosst-1.1.0[76], OpenFF-1.0.0[77] and OpenFF-2.0.0[78] as implemented in the OpenFF toolkit.[79] The UFF[80] and the MMFF94[81,82] were evaluated with RDKit.[83] Calculations with the OpenFF toolkit and RDKit were driven via QCEngine.[84]

All tested DFT methods were evaluated in combination with one out of the following three London dispersion corrections, D3,[85,86] D4,[32,87] or VV10[31,88,89] (also called NL or V). The two former were applied together with the rational (Becke–Johnson) damping function,[90–92] except for the M06-L functional, where zero (Chai–Head-Gordon) damping[93] was employed. Furthermore, revised damping parameters proposed by Smith et al.[94] and the optimized power damping function[95] were tested if available for the respective functionals. D3 and D4 corrections were calculated with the `s-dftd3` (version 0.5.1)[96] and `dftd4` (version 3.3.0)[97] standalone programs and consistently include three-body Axilrod–Teller–Muto (ATM)[98,99] dispersion contributions. The non-local density-dependent VV10 dispersion correction was calculated non-selfconsistently as implemented in `Turbomole 7.5.1` or in case of B97M-V, `ORCA` 5.0.1.

3 The ACONFL test set

We present a new benchmark set termed ACONFL (Alkane CONFormers Large) to evaluate QC, SQM, and FF methods concerning their performance in predicting alkane conformer energies. It extends the well-established and commonly used ACONF benchmark set by employing longer *n*-alkanes and more diverse conformer ensembles. While the largest alkane used in the ACONF benchmark is *n*-hexane, the ACONFL set is composed of *n*-dodecane (ACONF12 subset), *n*-hexadecane (ACONF16 subset), and *n*-icosane conformers (ACONF20 subset), including 50 conformational energies in total.

Table 1: Tested semi-empirical and force field methods

method	dispersion	reference
FF		
smirnoff99Frosst-1.1.0	LJ	75,76
OpenFF-1.0.0	LJ	75,77
OpenFF-2.0.0	LJ	75,78
UFF	LJ	80
MMFF94	Buf-14-7	81,82
GFN-FF	D4	40
SQM		
GFN1-xTB	D3(BJ)	34
GFN2-xTB	D4	35
PM7	D2	74
PM6	D3(BJ)	73
DFTB3	D3(BJ), D4	63–66
DFTB2	D4	67–69
LC-DFTB2	D3(BJ), D4	70

3.1 Construction of the test set

The conformer potential energy surface of an unbranched alkane is characterized by torsional twists that lead from linear chains to highly deformed structures dominated by intramolecular dispersion forces. At temperatures less than 300 K, short alkanes ($n = 4 - 8$) in the gas phase are well-known to prefer the linear all-trans conformation. However, as the length of the alkane grows, there is a point where the attractive NCIs will cause the chain to "self-solvate" into a folded conformer.[15] A cross-gauche-cross rotation combination introduces an energetically unfavorable syn-pentane-like conformation. In addition, the chain ends are not parallel in this conformation, thus reducing the possible stabilization due to van der Waals attraction. A hairpin conformation with four gauche rotations minimizes the number of strained bonds. This allows an energetically favorable parallel arrangement of the chain ends, yielding the suggested global minimum for longer alkanes.[16–18] The zig-zag-hairpin stability turning point appears to be around hexadecane.

Since the conformational ensembles become nearly continuous in energy for longer alkanes, compiling a benchmark set out of the lowest conformers up to a certain energy

Table 2: Tested methods and dispersion corrections

method	D3	D4	VV10	reference
Composite ("3c")				
HF-3c	✓	—	—	100
PBEh-3c	✓	—	—	101
B97-3c	✓	—	—	41
r^2SCAN-3c	—	✓	—	102
(meta-)GGA				
PBE	✓	✓	✓	103,104
TPSS	✓	✓	✓	105,106
B97M	✓	✓	✓	89,107,108
r^2SCAN	✓	✓	✓	109–111
M06L	✓	✓	—	112
(rs-)(meta-)Hybrid				
B3LYP	✓	✓	✓	113,114
PBE0	✓	✓	✓	115,116
PW6B95	✓	✓	✓	117
M06-2X	✓	—	—	118
MN12-SX	—	✓	—	119
ωB97M	✓	✓	✓	89,108,120
ωB97X	✓	✓	✓	93,121,122
Lh20t	—	✓	—	123
Double-hybrid				
B2PLYP	✓	✓	✓	24
PWPB95	✓	✓	✓	124
DSD-BLYP	✓	✓	—	124
revDSD-BLYP	—	✓	—	125
WFT				
HF	✓	✓	✓	
MP2	—	—	—	
MP3	—	—	—	
MP2D	✓	—	—	46
κOO-MP2	—	—	—	48,126
MP2.5	—	—	—	49
DLPNO-CCSD	—	—	—	

threshold is not practicable since even with the CCSD(T) at the estimated basis set limit to reliably predict conformational energies smaller than about 0.1 kcal/mol. Therefore, we created a conformational ensemble at the GFN-FF level of theory using the version four conformer search as implemented in CREST and selected conformers in a 5–6 kcal/mol

energy window with a decent (i.e., clearly distinguishable at the CCSD(T) level of theory (vide infra)) equidistant spacing of the conformational energies. This approach keeps the total number of conformers in the ACONFL set reasonably small while still including as much of the diversity of the complete conformational ensemble as possible. Those conformers were re-optimized at the B97-3c level of theory and used as the starting point for performing the reference calculations. B97-3c provides sufficiently accurate geometries for our purpose, although the overall accuracy of the geometries is secondary due to the de facto continuum of structures in the conformers for these highly flexible systems and the use of the same structures also for the reference CC calculations. In total, 53 single point calculations are required to evaluate the complete ACONFL, and 50 conformational energies with respect to the respective energetically lowest conformers are overall assessed for the three subsets, ACONF12, ACONF16, and ACONF20. Compared to the ACONF set with an mean absolute conformational energy of 1.83 kcal/mol the complete ACONFL set has a higher mean of 4.62 kcal/mol.

The ACONF12 subset shown in Fig. 1 contains twelve relative conformational energies, with the lowest conformer being the linear n-dodecane molecule and the mean absolute conformational energy being 4.28 kcal/mol. This set was already successfully used in several studies to test the performance of new DFT methods[102,127] as well as in a recent perspective on the description of conformational ensembles.[5] The numbering of the conformers results from the initial conformational search rather than the final energetic ordering.

For the ACONF16 subset in Fig. 2 17 relative conformational energies are included with the folded n-hexadecane as the energetically lowest conformer, which is in line with previous studies,[17,18,128] but the linear conformation being the second lowest is only slightly higher energy by 0.09 kcal/mol. The mean absolute conformational energy of ACONF16 is 3.98 kcal/mol. Finally, the ACONF20 subset contains 21 relative conformational energies (cf. Fig. 3). For this set, the hairpin-like conformer of n-icosane is with

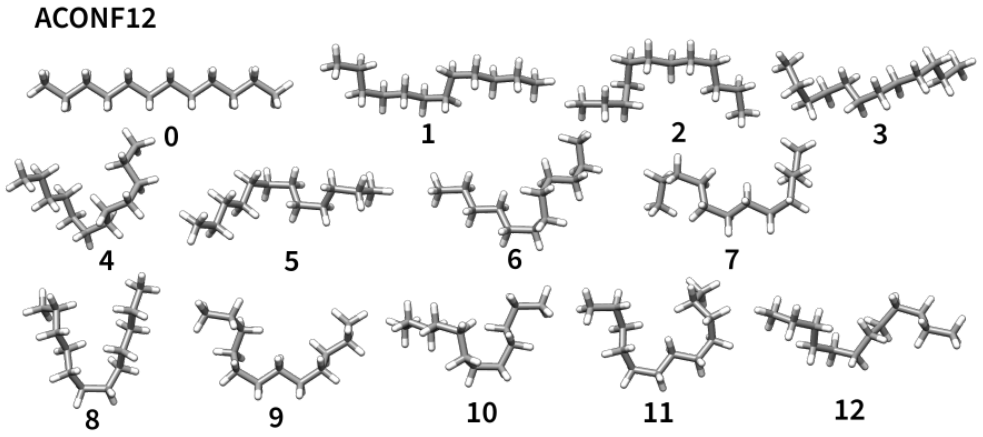


Figure 1: The 13 n-dodecane conformers of the ACONF12 subset. The conformer **0** is the lowest conformer, the numbering of the conformers does not necessarily correspond to their energetic order.

1.31 kcal/mol already significantly more stable the linear conformation. Here, the mean absolute conformational energy is 5.32 kcal/mol. These subsets allow for assessing a balanced description of dispersion and repulsion by the tested methods, since the shorter n-dodecane features incomplete attractive intramolecular NCIs to favor a closed form, while the longer n-icosane chains with favor a closed hairpin-like conformation due to cummulated effect of intramolecular NCIs, and the intermediate n-hexadecane set refers to the point, where both forms are very close in energy. It should be noted that the mean signed error (MSE) statistical descriptor depends on whether the lowest conformer is the linear or folded structure. Therefore, it is more meaningful to analyse the MSE, e.g. looking for a systematic of the linear form, for each subset separately rather than for the complete ACONFL as the MSEs of the subsets could inevitably cancel each other due to different signs.

3.2 Generation of the reference values

Previous studies of alkane conformers proved that effects beyond CCSD(T) are not important for the accurate description of conformational energies.[19] Since canonical CCSD(T) is

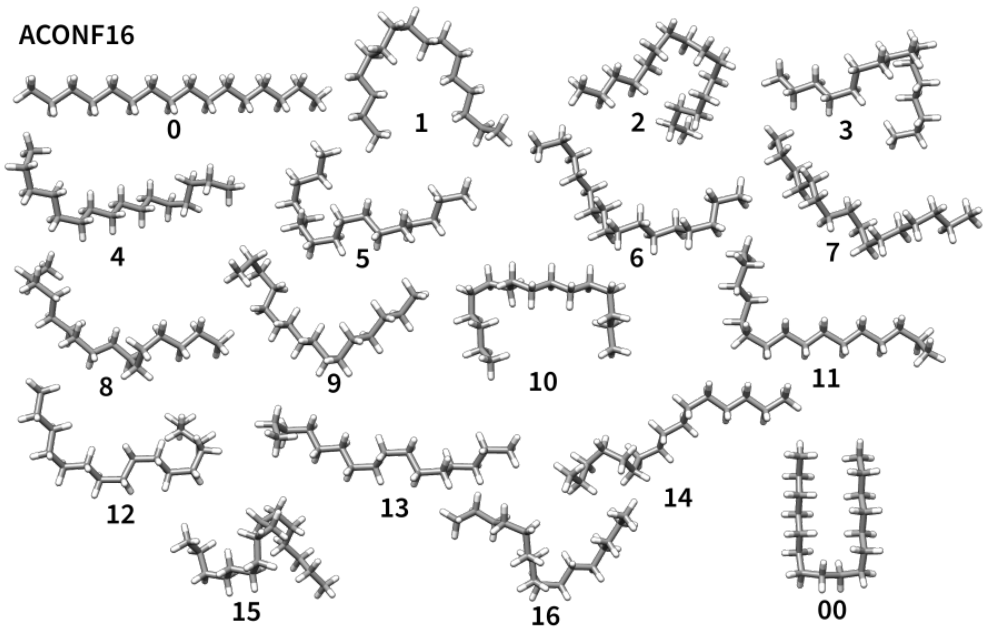


Figure 2: The 18 n-hexadecane conformers of the ACONF16 subset. The conformer **0** is the lowest conformer, the numbering of the conformers does not necessarily correspond to their energetic order.

computationally prohibitive for the target systems, the DLPNO-CCSD(T1) method with very tight PNO thresholds is used to approximate the canonical result as close as possible. For the extrapolation to the complete basis set limit[60,129] the aug-cc-pVTZ and aug-cc-pVQZ basis sets were used, dubbed aT and aQ respectively in the following paragraphs. This level of theory was shown to serve as reliable reference in various NCI benchmarks.[33,130,131] To also verify the accuracy the suggested reference protocol for the ACONFL set, we evaluate ACONF and compare to the highly accurate W1hval[19,132,133] reference values of the latter. The deviation of DLPNO-CCSD(T1)/CBS(aTaQ) from the W1hval conformational energies is shown in Fig. 4 on the left. With an negligible MSE of 0.04 kcal/mol and an error range of only 0.10 kcal/mol, sufficient accuracy for ACONFL can be expected with our reference protocol.

However, the full basis set extrapolation is still too expensive to be applicable for the ACONF16 and ACONF20 subsets. Therefore, we resort to a CBS extrapolation scheme based on focal-point analysis[134] for the latter. Following Marshall et al.,[131] the respective

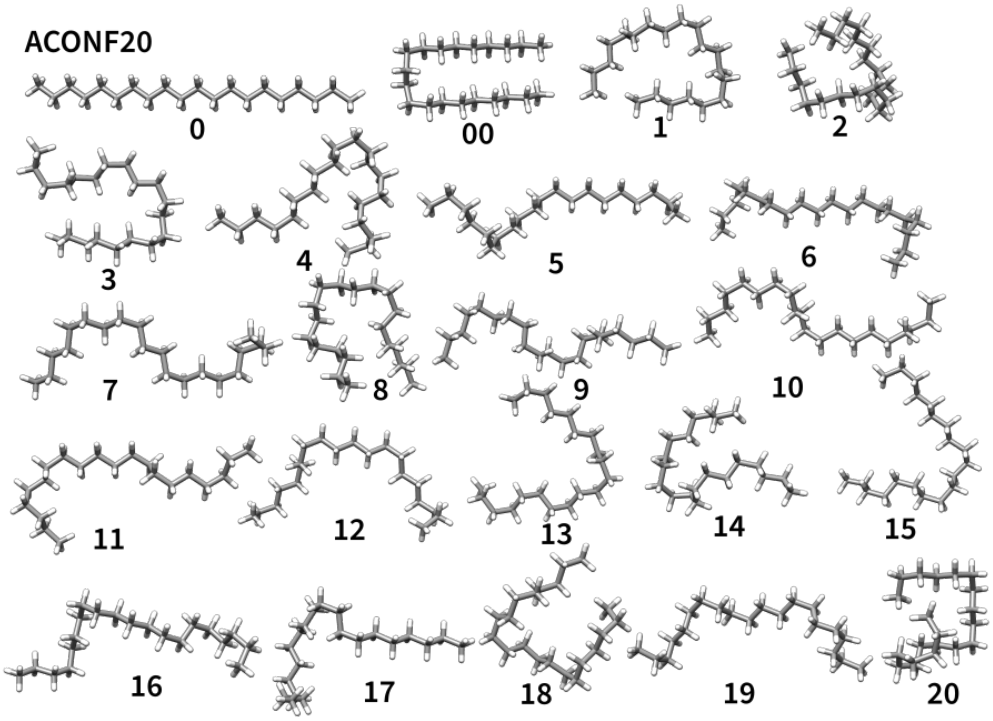


Figure 3: The 22 n-icosane conformers of the ACONF20 subset. The conformer **00** is the lowest conformer, the numbering of the conformers does not necessarily correspond to their energetic order.

"δCBS" basis set extrapolation scheme is given by

$$\delta\text{CBS} = E(\text{MP2/CBS(aTaQ)}) + E_\text{c}(\text{DLPNO-CCSD(T1)/aT}) - E_\text{c}(\text{MP2/aT})\,, \qquad (1)$$

where E_c is the correlation energy part of the total energy E. Additionally, we introduce a similar but multiplicative scheme dubbed xCBS, which represents a refined variant of the multiplicative CBS* scheme,[27,41] and is defined in the following way:

$$x\text{CBS} = E(\text{HF/CBS(aTaQ)}) + E_\text{c}(\text{MP2/CBS(aTaQ)}) \cdot \left(\frac{E_\text{c}(\text{DLPNO-CCSD(T1)/aT})}{E_\text{c}(\text{MP2/aT})} \right) . \qquad (2)$$

The xCBS protocol is typically less sensitive to more severe MP2 errors compared to the δCBS protocol. To estimate the additional error introduced by the more approximate basis set extrapolation, we compared the two schemes with the full CBS(aTaQ) conforma-

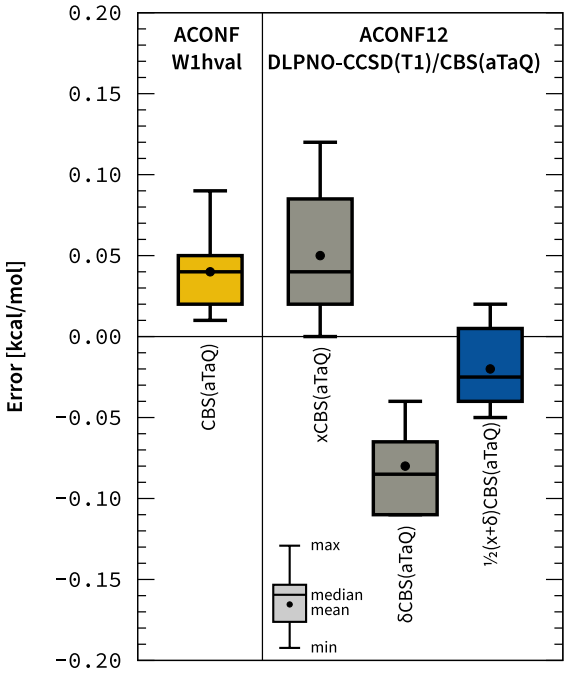


Figure 4: Deviation of the reference method DLPNO-CCSD(T1)/CBS(aTaQ) against W1hval for the ACONF benchmark set as well as deviation of CBS extrapolations on the ACONF12 benchmark set compared to a CBS(aTaQ) extrapolation scheme. The average conformational energy for the ACONF benchmark is given as 1.83 kcal/mol.

tional energies for the ACONF12 subset (see Fig. 4, on the right). While the xCBS(aTaQ) yields a slightly positive MSE of 0.05 kcal/mol, the δCBS(aTaQ) a slightly negative MSE of -0.08 kcal/mol. Hence, we find that the arithmetic mean of both schemes agrees exceptionally well with the full CBS(aTaQ) scheme. Therefore, this average was chosen as reference for the ACONF16 and ACONF20 subsets, for which the full CBS(aTaQ) extrapolation was computationally unfeasible.

The maximum residual error of the ACONFL reference conformational energies, resulting from the local DLPNO approximations, the basis set incompleteness and intramolecular superposition errors, and the additonal error from the focal point analysis for the larger subsets is conservatively estimated to be 0.35 kcal/mol. This uncertainty of the reference values is largely averaged in the analysis of the statistical descriptors for the entire ACONFL set. The square root of the sum of the squares of the estimated maximum

error divided by the number of conformational energies, yielding 0.05 kcal/mol for the ACONFL set, can be used as an estimate for statistically distinguishable values of the analyzed descriptors (see Sec. 4). With the given accuracy of the reference values, we are thus able to distinguish statistically significant errors of any method above 0.05 kcal/mol.

4 Results and discussion

In this section, the performance of all tested methods for the ACONFL set is presented and discussed. Specifically, DFT and the respective dispersion corrections, as well as WFT methods, are assessed in subsection 4.1, while SQM and FF methods are evaluated in subsection 4.2. Finally, a performance analysis in terms of computation times vs. accuracy is given in subsection 4.3. To assess the methods we will mainly discuss the mean absolute error (MAE), the analysis of other statistical quantities, like the mean signed error (MSE), standard deviation (SD), and the error range were investigated as well, but will only be discussed if they show deviating trends from the MAE. The consistency of the conformational ordering is measured by the Pearson r_p and Spearman correlation coefficients r_s, besides the MAE for the conformational energies and correctly identifying the lowest-lying conformer. For the precise definition of the employed statistical measures see the supporting information.

4.1 Assessment of DFT and WFT methods

For the following discussion, we selected five (meta-)GGAs, eight hybrids, and four DHDFs, which represent either commonly used or best performing members[29] of the respective functional rungs.[135] Established functionals like PBE,[103,104] TPSS,[105,106] and B3LYP[113,114] are included as well as modern functionals like B97M,[89,107,108] r^{2}SCAN,[109–111] and revDSD-BLYP.[125] In the hybrid class of functionals we have included global hybrids like PBE0,[115,116] range-separated hybrids like ωB97M,[89,108,120] screened exchange hybrids like MN12-SX,[118]

as well as local hybrids like Lh20t[123] to access a broad range of different construction strategies in this functional class. We also evaluated wavefunction methods like HF and MP2 in the overall comparison. Finally, we include several composite electronic structure methods of the "3c" scheme, namely B97-3c[41] (GGA), r^{2}SCAN-3c[102] (meta-GGA), PBEh-3c[101] (hybrid), and HF-3c[100] (HF), which use a tailored basis set, in combination with D3 or D4 dispersion correction and the geometrical counter-poise correction (gCP)[136,137] or a short-range basis correction (SRB)[41] to allow efficient yet accurate calculations (for a detailed overview on the "3c" type of methods we refer to Refs. 138 and 102). All methods tested here were combined with correction schemes to capture long-range dispersion interactions, which are absent in semi-local DFT.[30] We apply the D3 dispersion correction[85,86] with the rational Becke–Johnson (BJ) and zero Chai–Head-Gordon (0) damping schemes, the recently developed D4 dispersion correction,[32,87] and the nonlocal dispersion correction via the VV10 functional in its non-selfconsistent variant.[31,88,89] In case D4 damping parameters were not available, we determined them following the procedure described in Ref. 32.

Before assessing the general performance on the ACONFL, we will investigate the influence of different London dispersion corrections for a selected number of methods, including those which are available with D4, D3, and VV10. First, we want to stress that the MAE for all dispersion corrected methods is well below 1 kcal/mol, while MAEs of non-dispersion corrected functionals are significantly higher yielding an average MAE larger than 2 kcal/mol. Therefore, we will generally only consider dispersion corrected functionals in the following discussion. Dispersion interactions are crucial for the correct description of the investigated alkane conformers and due to their electronically simple structure semi-classical geometry dependent models should be sufficient. To check the influence of the dispersion correction we choose twelve methods for which D3, D4 and VV10 are available. For the tested methods shown in Fig. 5, we find that in seven cases the D4 corrected variant performs best, while for three methods D3 results in the best per-

forming method, and only in two cases the VV10 corrected functional yields the lowest MAE. Similarly, the average MAE for D4 corrected methods is with 0.29 kcal/mol lowest compared to an average MAE of 0.36 and 0.42 kcal/mol for D3 and VV10 corrected methods, respectively. We find a generally better performance for D4 compared to D3, which is most likely related to the improved parameterization strategy introduced together with D4.[32] Investigating different damping functions for D3, we find a worse performance with the zero-damping and usually an on-par performance with re-parameterized damping functions, for a full comparison see the ESI. This can be seen for a method like r^{2}SCAN where the same parameterization strategy was employed for all three dispersion corrections. Notably, the performance of VV10 with r^{2}SCAN is remarkably good and with an MAE of only 0.18 kcal/mol the best performing method in the (meta-)GGA class while outperforming all tested hybrid functionals. Further, we find for HF-D4 especially good performance with an MAE of 0.14 kcal/mol compared to its D3 and VV10 variant with larger errors, indicating that the consistent parameterization of the dispersion correction is crucial. Overall the D4 dispersion correction shows to be a reliable choice over a wide range of functionals in agreement with previous studies.[32,33,139]

To reduce the complexity of the further discussion and focus on the difference in the methods rather than dispersion corrections, we will select the best dispersion correction in each case for the DFT methods discussed in the next paragraphs. The complete statistics for all corrected methods are given in the supporting information. The error spread of all tested DFT, composite DFT and wavefunction methods is shown in Fig. 6. Notably, many methods are below an error range of $\pm$1 kcal/mol, with the best method DSD-BLYP-D3(BJ) even reaching an error range of only $\pm$0.22 kcal/mol approaching the accuracy of the coupled cluster reference values.

Overall, we find the best performing (meta-)GGA to be the newly developed r^{2}SCAN-V with an MAE of only 0.18 kcal/mol, the second best is B97M-V with 0.35 kcal/mol MAE. In the hybrid class the best performing method is r^{2}SCAN0-V with an MAE of

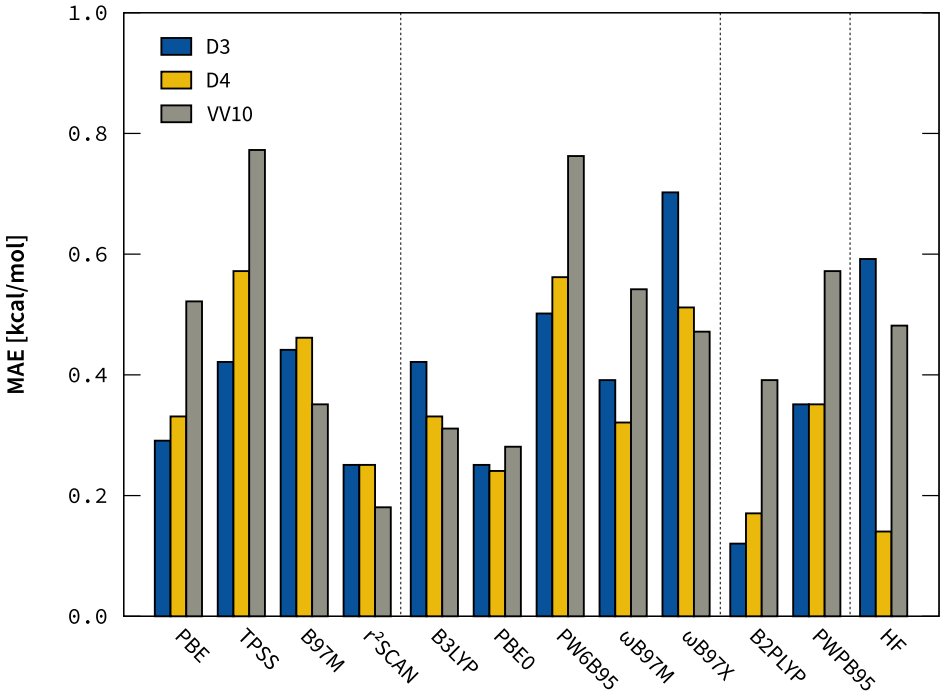


Figure 5: MAE of all twelve methods available with D4, D3, and VV10 dispersion corrections. The methods are grouped in respective rungs.

0.17 kcal/mol, which performs as good as the best (meta-)GGAs. Since the investigated systems are electronically simple, the quality of the base functional is more important than the admixture of Fock exchange here, and for reasons of computational efficiency, a good (meta-)GGA like r^{2}SCAN-V is therefore preferable. We can recover the hierarchy of Jacob's ladder[135] at the highest rung with the DHDFs, where the best-performing method on the entire ACONFL is DSD-BLYP-D3(BJ) with an MAE of only 0.06 kcal/mol.

Notably, HF-D4 performs very well with an MAE of only 0.14 kcal/mol and thus, as good as the best tested hybrid functional. However, while dispersion corrected HF performs well, we find that MP2 at the estimated basis set limit (CBS(aug-TZ/aug-QZ)) results in a large MAE value of 0.59 kcal/mol, i. e. worse than most of the assessed dispersion corrected DFT methods, except for some Minnesota-type functionals. The significance of post-MP2 contributions was already observed for *n*-hexane in the ACONF set.[19] While for MP2 and correlated WFT methods in general, large and diffuse basis

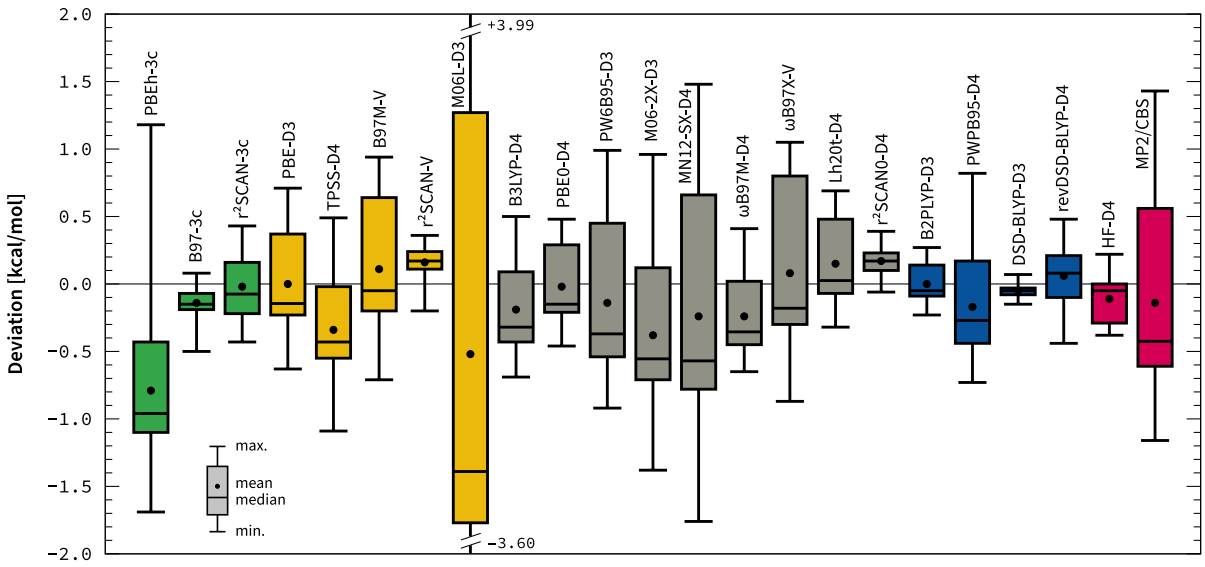


Figure 6: Deviations for the ACONFL set for selected DFT and WFT methods.

sets are necessary to fully recover long-range dispersion, the physically correct behaviour at the mean field HF level is included more conveniently by a suitable dispersion correction. Moreover, in D4 we can approximately recover three-body dispersion contributions, which would require a higher order treatment than MP2. To further analyse the rather poor performance of MP2, we compare MP2/def2-QZVPP with the recently introduced regularized κOO-MP2 (κ=1.1)/def2-QZVPP, MP2.5[49]/def2-TZVPP, and MP2D[46]/def2-QZVPP for the ACONF12 subset. With the κ regularization and orbital-optimization we only find a small improvement 0.12 kcal/mol in the MAE, as expected for closed-shell systems with large HOMO–LUMO gaps.[48,126] The MAE is reduced to 0.24 kcal/mol by mixing in third-order terms via the MP2.5 scheme containing 50% MP3 correlation energy. The full MP3/def2-TZVPP method yields an outstandingly small MAE of 0.02 kcal/mol due to a fortunate compensation of the MP3 overshooting and the residual BSSE of the triple-ζ basis set. Finally, employing the MP2D approach to correct the uncoupled HF dispersion treatment by DFT-D3 ones reduces the MAE by 0.33 kcal/mol compared to the original MP2. The remaining residual BSSE can be estimated by comparing the QZ

results with the CBS result reducing the MAE by 0.17 kcal/mol. Notably, the combination of MP2 and a dispersion correction contribution to recover the proper long-range dispersion also significantly reduces the error of DHDFs (vide supra). However, the MAE for ACONF12 with MP2D is still 0.21 kcal/mol higher than for HF-D4 due to residual basis set incompleteness and superposition errors. Comparable to the general performance of the series MP2/MP3/MP4 for NCIs,[49] DLPNO-CCSD overstabilizes the linear structure (MSE of 0.60 kcal/mol), verifying that the connected triples correction including contributions from MP4 and MP5 is essential for accurate coupled cluster results. The comparison of all tested WFT methods for the ACONF12 subset is shown in Fig. 7.

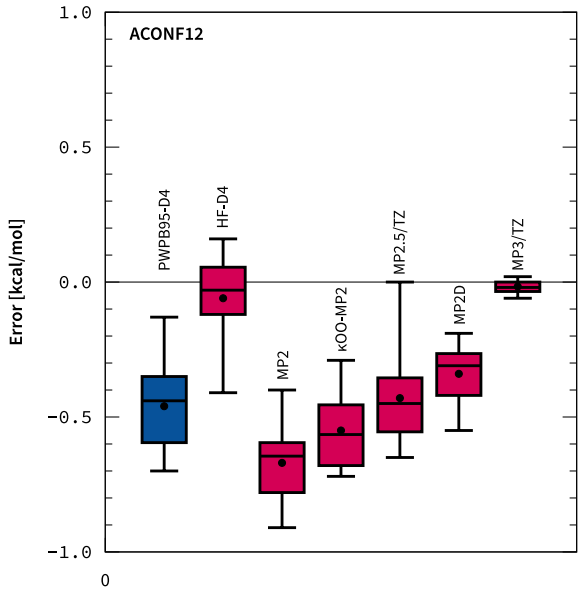


Figure 7: Comparison of wavefunction methods on the ACONF12 using a def2-QZVPP basis set if not noted otherwise.

To assess the potentially larger residual basis set incompleteness and superposition error in DHDF functionals we evaluated PWPB95-D4 with CBS(aTaQ) basis extrapolation and compared the results on the ACONF12 set with the values obtained in the def2-QZVPP basis set. The MAE reduces for this subset reduces from 0.46 to 0.32 kcal/mol, which is a statistically significant improvement. However, DHDFs are usually not extrapolated to the approximated basis set limit (i.e., CBS(aTaQ) due to the increased (about

eight times in this case) computational effort compared to the def2-QZVPP calculation. Therefore, we primarily investigated the performance with the commonly applied def2-QZVPP basis set.

Compared to the generally good performance of DFT across all functional classes, we note that only the Minnesota-type functionals tested here show significantly increased deviations, which is rather unusual. While they incorporate short and medium-range dispersion implicitly via their parameterization, the semi-local functionals still cannot fully account for long-range dispersion.[30] However, we find for most functionals of this type the combination with long-range London dispersion corrections is not beneficial for large *n*-alkanes. For example, for MN12-SX the uncorrected functional yields an MAE of 0.48 kcal/mol while that of the D4 corrected functional is with 0.76 kcal/mol significantly larger.

With the best functionals identified close to their basis set limit, we now want to investigate how much of their performance can be recovered with more cost-efficient composite electronic structure methods, namely of the "3c" construction scheme. The "3c" methods are well-suited in a multilevel model scheme to re-rank or re-optimize an ensemble created at a lower level of theory, like SQM or force fields. Both B97-3c and r^{2}SCAN-3c provide a very good description of ACONFL with an MAE of 0.15 and 0.20 kcal/mol, respectively, approaching the accuracy of the best performing methods in a quadruple-ζ basis set. B97-3c is even the best among the tested GGA functionals for this benchmark set. PBEh-3c performs somewhat worse with 0.87 kcal/mol, which can mainly be attributed to the small modified double-ζ basis set and the respective gCP error, while B97-3c and r^{2}SCAN-3c employ a larger modified triple-ζ basis set. Note that the base functional PBE already performs worse compared to modern functionals like r^{2}SCAN. Overall, the composite methods r^{2}SCAN-3c and B97-3c prove to be sufficiently accurate in a very cost-effective way. Therefore we clearly recommend their usage in multilevel workflows, e. g., for conformer ranking of flexible molecules with *n*-alkanes as building

blocks.

4.2 Assessment of SQM and FF methods

SQM and FF methods are often employed for large scale screening purposes. Due to their much lower computational cost the calculation of large conformational ensembles becomes possible, including challenging tasks like determining the absolute conformational entropy.[140] However, due to approximations inherent to SQM and FF methods the accuracy is often significantly lower and hence a re-ranking of generated ensembles at a higher level of theory becomes necessary.[5] The margin of the energetic threshold to include structures and therefore the amount of structures from the lower level of theory within such refinement workflows is crucial for the overall computational efficiency. Especially, the prediction of the correct energetically most-favorable conformer is important to avoid sorting out structures with major contributions to the final ensemble. The ACONFL provides the chemically most simple yet most flexible molecules for assessing the quality of SQM and FF methods in this context. The deviations of all semiempirical methods and force fields are shown in Fig. 8 For the three best and the three worst performing methods of this category we also show their conformational energies in Fig. 9.

A widely used method is the universal force field (UFF).[80] UFF yields a good correlation ($r_p = 0.98$ and $r_s = 0.95$) but the overall MAE of 2.91 kcal/mol is large given the mean energy of 4.62 kcal/mol. Furthermore, the UFF conformational energies are systematically too small (MSE of -2.89 kcal/mol), indicating an overall too shallow potential energy surface. A strong systematic error and too small conformational energies even with a good correlation results in larger conformational ensembles, which negatively impact the computational cost of later refinement steps at a higher level of accuracy. A similar behavior is observed for the MMFF94 force field with an MAE of 3.41 kcal/mol and also a large negative MSE.

GFN-FF is another general force field which we have tested. It yields very good agree-

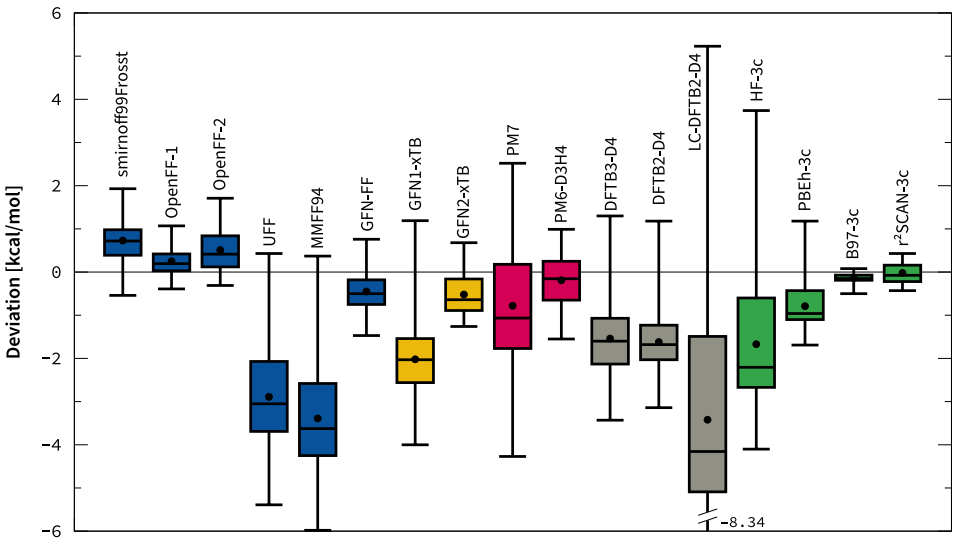


Figure 8: Comparison of all tested semiempirical quantum mechanical methods and force field methods. The "3c" composite method are included as point of reference to Fig. 6.

ment with the reference, at least for an FF, with an MAE of 0.55 kcal/mol and a good correlation of the conformer ordering ($r_p = 0.97$ and $r_s = 0.96$). Most importantly, it correctly identifies all the lowest-lying conformers in the respective subsets while getting close to the performance of some DFT methods.

From all tested force field methods, the OpenFF-1.0.0 performs best with an MAE of only 0.31 kcal/mol. Also, the SMIRNOFF99Frost and OpenFF-2.0.0 force fields yield small MAEs of 0.76 and 0.56 kcal/mol, respectively. Moreover, all SMIRNOFF methods yield an excellent Pearson correlation coefficient of 0.99. Most force fields can also correctly identify the linear form of *n*-hexadecane to be lower in energy than the lowest-lying folded structure. The individual conformational energies for the OpenFF-1.0.0 and GFN-FF as well as MMFF94 are shown in Fig. 9, emphasizing the correct conformational ordering produced by the former methods and the too shallow potential energy surface produced by the latter method.

After investigating force fields we will focus on SQM methods as the next more sophisticated level of theory explicitly including electronic structure effects, like the HF based

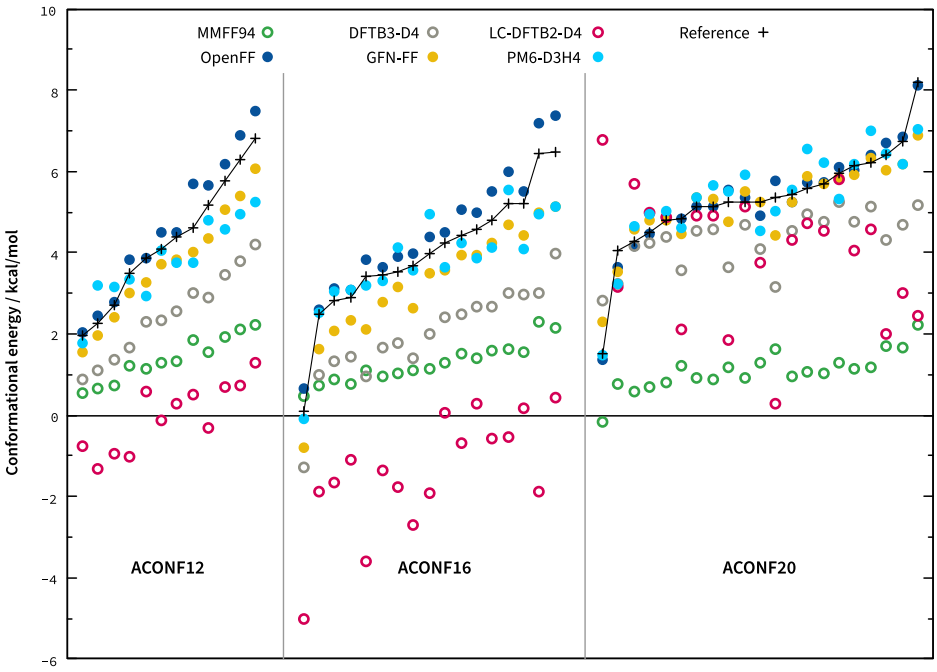


Figure 9: Conformational energies for the subsets ACONF12, ACONF16, and ACONF20 for three of the best performing and three (filled dots) of the worst performing (open dots) semiempirical and force field methods tested. The reference energies are given as black crosses, the connecting line serves only for better visibility.

NDDO methods of the PM*x* family and the DFT based tight-binding methods of either the DFTB or xTB flavor. The PM6-D3H4 method provides with an MAE of 0.55 kcal/mol a reasonably accurate description. This good performance seems to be in line with the very good results obtained by dispersion-corrected HF (MAE of 0.14 kcal/mol), on which PM6 is formally based on. However, in contrast its successor PM7 yields with 1.48 kcal/mol a much larger MAE.

From the tested tight binding methods, we find that GFN2-xTB performs best with an MAE of 0.58 kcal/mol. Compared to GFN-FF the GFN2-xTB MAE shows similar performance, however the error range is with 0.29 kcal/mol smaller than for GFN-FF (1.94 compared to 2.23 kcal/mol). Similarly, the error range obtained by GFN2-xTB is by 0.59 kcal/mol smaller than with PM6-D3H4. While the performance of DFTB2-D4 and DFTB3-D4 is quite similar to each other (MAEs of 1.66 and 1.60 kcal/mol, respectively)

the better performance in GFN2-xTB may originate from the improved description of the anisotropic electrostatics. In contrast, the GFN1-xTB method performs with an MAE of 2.06 kcal/mol worse, which is could be related to the basis set on hydrogen and the resulting worse description of repulsive NCI contacts. A remarkably weak performer is the LC-DFTB2-D4 method, which introduces spurious large errors in the conformational energies and almost no correlation of the energetic order with the reference ($r_p = 0.37$ and $r_s = 0.36$). Visual inspection of the conformational energies in Fig. 9 for LC-DFTB2-D4 shows severe errors for each of the subset, where the conformational energies are systematically too low (ACONF12, ACONF16) or spread over a wide range (ACONF20). The large MAE 3.71 kcal/mol results from the distorted potential energy surface. Whether this originates from the parameterization or is a more fundamental problem remains an open question due to lack of alternative long-range corrected DFTB methods to compare with.

Although alkanes are thought to be very prototypical, especially SQM and FFs result in a rather unusual performance order at least for the longer alkane conformers (ACONF16 and ACONF20). Overall, among the force fields and semiempirical methods tested here, GFN-FF provides the best compromise between speed and accuracy.

4.3 Performance comparison

Besides the accuracy of the method, an important factor for conformational sampling is its computational cost. For a representative number of methods, we show the computation time to evaluate the whole ACONFL benchmark set, together with their MAE. The wall times were obtained by parallel calculations using four CPU cores and are shown in Fig. 10. The evaluation of single point energies is representative for a reranking of an ensemble generated by a lower level method in a multilevel workflow, however less suitable for semiempirical methods as those are usually used in the generation of the ensemble in geometry optimizations as additional overhead from the restart or program invocation

can be already substantial compared to the total runtime.

Still, the relative time required for the single point evaluation is representative for comparing the computational efficiency of different semiempirical methods with each other. We find that the evaluation of the single point energies for SMIRNOFF methods takes 1.1 min on the entire ACONFL, while the GFN2-xTB method requires less than a second runtime. This difference results from the AM1-BCC charge calculation performed as part of the setup of the SMIRNOFF parameterization. In practice, this calculation has to be done only once per structure, subsequent energy and gradient evaluations will be significantly faster but require proper caching via the compute engine to remain feasible.

Note, the calculation of the reference values at DLPNO-CCSD(T1)/CBS level of theory took about four months cumulative wall time for the whole set.

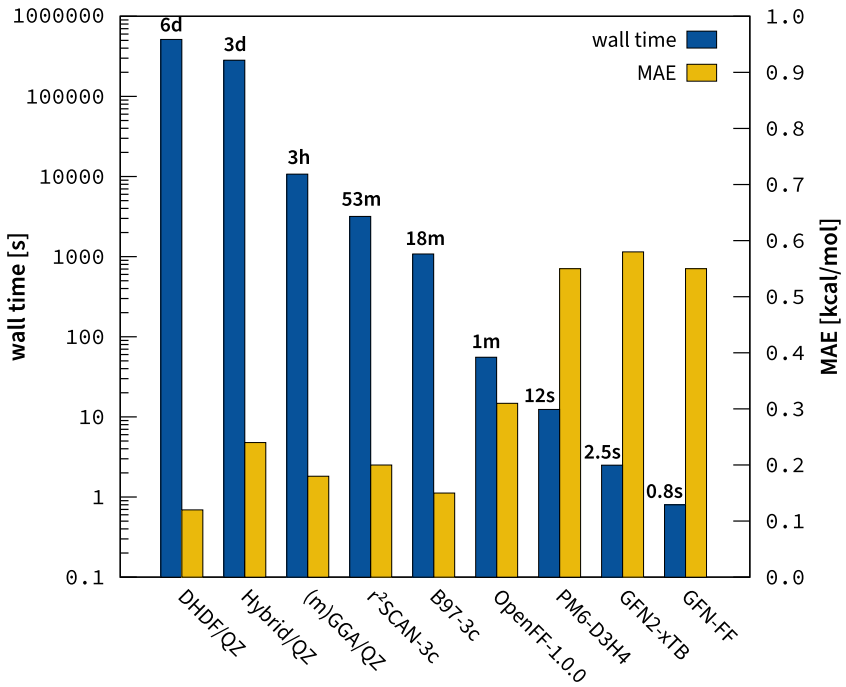


Figure 10: Wall time for evaluation of the complete ACONFL benchmark set on four Intel Core i7-7700K CPU cores. Due to the vastly different scales present over the wide range of methods assessed here we show the timings in seconds on a logarithmic scale.

5 Conclusions

We introduced the first benchmark set focusing on the conformational ensembles of long alkane chains, which are a prominent structural motif in many technically and biologically relevant molecules. This new set is termed ACONFL, indicating its relation to the ACONF benchmark introduced by Gruzman et al. in 2009,[19] which only includes alkane conformers up to n-hexane. ACONFL comprises conformational ensembles (53 conformers and 50 relative energies up to about 8 kcal/mol) of the n-alkanes $C_{12}H_{24}$, $C_{16}H_{34}$, and $C_{20}H_{42}$ that cover the transition from linear to hairpin structures as energetically lowest conformers thus providing a more realistic picture than the ACONF set. We generated reliable reference conformational energies employing high level coupled cluster theory close to the basis set limit (DLPNO-CCSD(T1)/VeryTightPNO/CBS(aug-cc-pVTZ/aug-cc-pVQZ)) allowing for a statistically meaningful evaluation for lower level methods with MAE differences larger than 0.05 kcal/mol.

Using this highly accurate reference data, we explored the performance of a hierarchy of density functionals, the "3c" family of density functional theory (DFT) composite methods, the wavefunction-based approaches HF and MP2, semiempirical approaches (SQM), as well as standard and recent force field (FF) methods. It bears pointing out that of those methods, only the latter (SQM and FF-based) are sufficiently efficient to comprehensively explore the conformational space of these flexible molecules, and are thus indispensable to accurately calculate properties like their absolute entropy.[140]

Concerning the DFT-based methods, we found that (meta-)GGA and hybrid functionals are similarly accurate. In other words, the inclusion Fock exchange does not lead to significant improvements which would justify the increased computational demands. Only DHDFs significantly reduce the error in the conformational energies further. However, even in this case it is questionable whether the small gain in accuracy (0.05 kcal/mol on average) satisfies the massively increased computational cost. The best tested method is DSD-BLYP-D3(BJ) with an MAE of 0.06 kcal/mol while the worst tested functional is

M06L-D4 with an MAE of 1.84 kcal/mol. In the ACONFL benchmark we are able to quantify the impact of dispersion, while in the smaller ACONF benchmark set, many dispersion uncorrected functionals perform only slightly worse than their dispersion corrected counterparts. In this respect, the composite DFT methods B97-3c and r^{2}SCAN-3c provide an outstanding cost/accuracy ratio as they perform on part with DFT/QZ methods for a small fraction (about 2–3 orders of magnitude faster compared to DHDF/QZ) of the computational cost.

Regarding correction-schemes in general, we want to point out that for DFT (and also HF) the application of a dispersion correction is crucial for conformational energies of longer alkane chains, which is consistent with previous studies.[17,18,128] Especially the influence of dispersion corrections on the conformational energies cannot be assessed with the smaller ACONF benchmark set[19] alone. Comparing commonly applied dispersion correction schemes, we find that methods with D4 perform on average slightly better (MAE about 0.1 kcal/mol lower) than D3 or VV10 corrected methods. Further, we notice a very good performance for HF-D4 (MAE of 0.14 kcal/mol), which can be attributed to the accurate parameterization of D4 as well as the approximate inclusion of many-body dispersion effects in the latter. While saturated systems with large gaps are usually well described by MP2, we find surprisingly poor performance for MP2/CBS, which we largely attribute to the uncoupled HF dispersion coefficients.[141] This shortcoming of the MP2 can be partially overcome by using the MP2D method, however the performance was found to be still worse compared to computationally less demanding HF-D4 method. Finally, the combination of MP3 in a triple-ζ basis set profits from fortunate error compensation, which makes DHDFs using KS-MP3 correlation[142] worth exploring in the future.

Moving to SQM and FF methods, we find that the inherent additional approximations of those methods also increase the overall error (average MAE 1.55 kcal/mol) compared to DFT significantly. However, GFN2-xTB and PM6-D3H4, the best-performing among the tested SQM methods (0.58 and 0.55 kcal/mol MAE, respectively) are sufficiently ac-

curate to retain the energetic ordering of the conformer ensemble reasonably well. Older standard FFs like UFF and MMFF94 yield generally too shallow potential energy surfaces and, in turn, much too large conformer ensembles in a given energy window. These methods thus require re-ranking and re-optimization at a higher level of theory, which makes them unsuitable in practice, especially if the global energy minimum conformer is incorrectly predicted (i. e. preference for the linear over the folded conformer for hexadecane and larger). This also raises the question of whether these common force fields are able to distinguish lipid side-chain conformations crucial for modeling biological systems in solution. Significantly higher accuracy is obtained with the recently introduced GFN-FF and the OpenFF-1.0.0 from the SMIRNOFF FF method, both outperforming all tested SQM methods, even approaching the accuracy of some hybrid DFT/QZ methods (with MAEs of 0.55 and 0.31 kcal/mol, respectively). In our experience, however, the freely available implementation of the SMIRNOFF FFs via QCEngine is not optimal, requiring an overhead of computer time by two orders of magnitude compared to GFN-FF that render their use impractical. Hence, GFN-FF provides both, fast and accurately conformational ensembles and outperformed several SQM methods on this benchmark set, which is quite surprising.

After all, due to its most favorable cost-accuracy ratio, we recommend GFN-FF for conformational searches of alkane conformers for large scale screening applications or to model extended systems with long alkyl chains. However, depending on the other details of the system in question, it may be required to move to a more robust and accurate DFT based method. Here, the efficient composite methods r^{2}SCAN-3c and B97-3c performed particularly well. Although it should be a seemingly straightforward problem for SQM and FF methods due to the simple electronic structure of alkanes, only few of the tested methods performed convincingly and thus we recommend the ACONFL as a helpful fit set for parameterization of new SQM and FF as well as machine learning potentials. Further, the ACONFL provides a meaningful validation set for newly developed DFT

and MP2-type WFT methods, especially since the accurate description of conformational energy poses a unique challenge for every investigated ensemble.

Acknowledgement

We thank J.-M. Mewes for carefully proofreading the manuscript and performing the QChem calculations. We thank C. Bannwarth for helpful discussions. We thank the reviewers for pointing out an error in the evaluation script for the conformer ordering of *n*-hexadecane. This work was supported by the DFG in the framework of the "Gottfried Wilhelm Leibniz Prize" awarded to S. G. and the SPP 2363 on "Utilization and Development of Machine Learning for Molecular Applications – Molecular Machine Learning."

Supporting Information Available

The statistical evaluation of the ACONF and ACONFL benchmark sets as well as the input geometries are provided in the supporting information

1. `aconfl.xls`: statistical evaluation of ACONF and ACONFL
2. `aconfl.zip`: geometries in xyz format
   the geometries are also available at `https://github.com/grimme-lab/aconfl`
3. `aconfl.pdf`: statistical descriptors

References

(1) Baldwin, R.; Baker, D. *Peptide Solvation and H-bonds*; Elsevier, 2006.

(2) Shmueli, U.; Wilson, A. Conformation in biology. 1982.

(3) Göttlich, R.; Kahrs, B. C.; Krüger, J.; Hoffmann, R. W. Open chain compounds with preferredconformations. *Chemical Communications* **1997**, 247–252.

(4) Tasi, G.; Izsák, R.; Matisz, G.; Császár, A. G.; Kállay, M.; Ruscic, B.; Stanton, J. F. The origin of systematic error in the standard enthalpies of formation of hydrocarbons computed via atomization schemes. *ChemPhysChem* **2006**, *7*, 1664–1667.

(5) Grimme, S.; Bohle, F.; Hansen, A.; Pracht, P.; Spicher, S.; Stahn, M. Efficient Quantum Chemical Calculation of Structure Ensembles and Free Energies for Nonrigid Molecules. *J. Phys. Chem. A* **2021**, *125*, 4039–4054.

(6) Goossens, K.; De Winter, H. Molecular dynamics simulations of membrane proteins: An overview. *J. Chem. Inf. Model.* **2018**, *58*, 2193–2202.

(7) McGaughey, G. B.; Sheridan, R. P.; Bayly, C. I.; Culberson, J. C.; Kreatsoulas, C.; Lindsley, S.; Maiorov, V.; Truchon, J.-F.; Cornell, W. D. Comparison of topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Model.* **2007**, *47*, 1504–1519.

(8) Pitzer, K. S. Thermodynamics of Gaseous Hydrocarbons: Ethane, Ethylene, Propane, Propylene, n-Butane, Isobutane, 1-Butene, Cis and Trans 2-Butenes, Isobutene, and Neopentane (Tetramethylmethane). *J. Chem. Phys.* **1937**, *5*, 473–479.

(9) Herrebout, W.; Van der Veken, B.; Wang, A.; Durig, J. Enthalpy difference between conformers of n-butane and the potential function governing conformational interchange. *J. Chem. Phys.* **1995**, *99*, 578–585.

(10) Balabin, R. M. Enthalpy difference between conformations of normal alkanes: Raman spectroscopy study of n-pentane and n-butane. *J. Phys. Chem. A* **2009**, *113*, 1012–1019.

(11) Basu, A.; Mookherjee, M.; McMahan, E.; Haberl, B.; Boehler, R. Behavior of Long-Chain Hydrocarbons at High Pressures and Temperatures. *J. Phys. Chem. B* **2022**,

(12) Smith, G. D.; Jaffe, R. L. Quantum chemistry study of conformational energies and rotational energy barriers in n-alkanes. *J. Phys. C* **1996**, *100*, 18718–18724.

(13) Allinger, N. L.; Fermann, J. T.; Allen, W. D.; Schaefer III, H. F. The torsional conformations of butane: Definitive energetics from ab initio methods. *J. Chem. Phys.* **1997**, *106*, 5143–5150.

(14) Salam, A.; Deleuze, M. High-level theoretical study of the conformational equilibrium of n-pentane. *J. Chem. Phys.* **2002**, *116*, 1296–1302.

(15) Goodman, J. M. What is the longest unbranched alkane with a linear global minimum conformation? *J. Chem. Inf. Comp. Sci.* **1997**, *37*, 876–878.

(16) Lüttschwager, N. O.; Wassermann, T. N.; Mata, R. A.; Suhm, M. A. The last globally stable extended alkane. *Angew. Chem. Int. Ed.* **2013**, *52*, 463–466.

(17) Byrd, J. N.; Bartlett, R. J.; Montgomery Jr, J. A. At what chain length do unbranched alkanes prefer folded conformations? *J. Phys. Chem. A* **2014**, *118*, 1706–1712.

(18) Liakos, D. G.; Neese, F. Domain based pair natural orbital coupled cluster studies on linear and folded alkane chains. *J. Chem. Theory Comput.* **2015**, *11*, 2137–2143.

(19) Gruzman, D.; Karton, A.; Martin, J. M. L. Performance of Ab Initio and Density Functional Methods for Conformational Equilibria of CnH2n+2 Alkane Isomers (n = 4–8). *J. Phys. Chem. A* **2009**, *113*, 11974–11983, PMID: 19795892.

(20) Martin, J. M. L. What can we learn about dispersion from the conformer surface of n-pentane? *J. Phys. Chem. A* **2013**, *117*, 3118–3132.

(21) Schleyer, P. v. R.; Williams Jr, J. E.; Blanchard, K. Evaluation of strain in hydrocarbons. The strain in adamantane and its origin. *J. Am. Chem. Soc.* **1970**, *92*, 2377–2386.

(22) Wodrich, M. D.; Corminboeuf, C.; Schleyer, P. v. R. Systematic errors in computed alkane energies using B3LYP and other popular DFT functionals. *Organic letters* **2006**, *8*, 3631–3634.

(23) Karton, A.; Gruzman, D.; Martin, J. M. L. Benchmark Thermochemistry of the C n H2 n+ 2 Alkane Isomers (n= 2- 8) and Performance of DFT and Composite Ab Initio Methods for Dispersion-Driven Isomeric Equilibria. *J. Phys. Chem. A* **2009**, *113*, 8434–8447.

(24) Grimme, S. Semiempirical hybrid density functional with perturbative second-order correlation. *J. Chem. Phys.* **2006**, *124*, 034108.

(25) Fogueri, U. R.; Kozuch, S.; Karton, A.; Martin, J. M. L. The melatonin conformer space: Benchmark and assessment of wave function and DFT methods for a paradigmatic biological and pharmacological molecule. *J. Phys. Chem. A* **2013**, *117*, 2269–2277.

(26) Kozuch, S.; Bachrach, S. M.; Martin, J. M. L. 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.

(27) 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.

(28) 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.

(29) 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.

(30) Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-corrected mean-field electronic structure methods. *Chemical reviews* **2016**, *116*, 5105–5154.

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

(32) Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A generally applicable atomic-charge dependent London dispersion correction. *J. Chem. Phys.* **2019**, *150*, 154122.

(33) Spicher, S.; Caldeweyher, E.; Hansen, A.; Grimme, S. Benchmarking London dispersion corrected density functional theory for noncovalent ion–π interactions. *Phys. Chem. Chem. Phys.* **2021**, *23*, 11635–11648.

(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) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Mul-

tipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.

(36) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding quantum chemistry methods. *WIREs Comput. Mol. Sci.* **2020**, e01493.

(37) Pracht, P.; Bohle, F.; Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169–7192.

(38) Neese, F. Software update: the ORCA program system, version 4.0. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1327.

(39) Conformer-Rotamer Ensemble Sampling Tool based on the xtb Semiempirical Extended Tight-Binding Program Package `crest`. `https://github.com/grimme-lab/crest`, Accessed: 2021-12-28.

(40) Spicher, S.; Grimme, S. Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems. *Angew. Chem. Int. Ed.* **2020**, *59*, 15665–15673.

(41) 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.

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

(43) Balasubramani, S. G.; Chen, G. P.; Coriani, S.; Diedenhofen, M.; Frank, M. S.; Franzke, Y. J.; Furche, F.; Grotjahn, R.; Harding, M. E.; Hättig, C. et al. TURBOMOLE: Modular program suite for ab initio quantum-chemical and condensed-matter simulations. *J. Chem. Phys.* **2020**, *152*, 184107.

(44) ORCA – an ab initio, density functional and semiempirical program package, V. 5.0.1, F. Neese, MPI für Kohlenforschung, Mülheim a. d. Ruhr (Germany), **2021**.

(45) ORCA – an ab initio, density functional and semiempirical program package, V. 4.2.1, F. Neese, MPI für Kohlenforschung, Mülheim a. d. Ruhr (Germany), **2020**.

(46) Rezac, J.; Greenwell, C.; Beran, G. J. Accurate noncovalent interactions via dispersion-corrected second-order Møller–Plesset perturbation theory. *J. Chem. Theory Comput.* **2018**, *14*, 4711–4721.

(47) Smith, D. G.; Burns, L. A.; Simmonett, A. C.; Parrish, R. M.; Schieber, M. C.; Galvelis, R.; Kraus, P.; Kruse, H.; Di Remigio, R.; Alenaizan, A. et al. PSI4 1.4: Open-source software for high-throughput quantum chemistry. *J. Chem. Phys.* **2020**, *152*, 184108.

(48) Lee, J.; Head-Gordon, M. Regularized orbital-optimized second-order Møller–Plesset perturbation theory: A reliable fifth-order-scaling electron correlation model with orbital energy dependent regularizers. *J. Chem. Theory Comput.* **2018**, *14*, 5203–5219.

(49) Pitoňák, M.; Neogrády, P.; Černỳ, J.; Grimme, S.; Hobza, P. Scaled MP3 non-covalent interaction energies agree closely with accurate CCSD (T) benchmark data. *ChemPhysChem* **2009**, *10*, 282–289.

(50) Epifanovsky, E.; Gilbert, A. T.; Feng, X.; Lee, J.; Mao, Y.; Mardirossian, N.; Pokhilko, P.; White, A. F.; Coons, M. P.; Dempwolff, A. L. et al. Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *J. Chem. Phys.* **2021**, *155*, 084801.

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

(52) Kendall, R. A.; Früchtl, H. A. The impact of the resolution of the identity approximate integral method on modern ab initio algorithm development. *Theor. Chem. Acc.* **1997**, *97*, 158–163.

(53) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.

(54) 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.

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

(56) 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.

(57) Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. Sparse maps – A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory. *J. Chem. Phys.* **2016**, *144*, 024109.

(58) Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D. G.; Minenkov, Y.; Cavallo, L.; Neese, F. Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [DLPNO-CCSD(T)]. *J. Chem. Phys.* **2018**, *148*, 011101.

(59) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

(60) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set convergence of correlated calculations on water. *J. Chem. Phys.* **1997**, *106*, 9639–9646.

(61) Hourahine, B.; Aradi, B.; Blum, V.; Bonafé, F.; Buccheri, A.; Camacho, C.; Cevallos, C.; Deshaye, M. Y.; Dumitrică, T.; Dominguez, A. et al. DFTB+, a software package for efficient approximate density functional theory based atomistic simulations. *J. Chem. Phys.* **2020**, *152*, 124101.

(62) "DFTB+ general package for performing fast atomistic simulations", https://github.com/dftbplus/dftbplus. Accessed: 2021-05-03.

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

(64) 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.

(65) Lu, X.; Gaus, M.; Elstner, M.; Cui, Q. Parametrization of DFTB3/3OB for Magnesium and Zinc for Chemical and Biological Applications. *J. Phys. Chem. B* **2015**, *119*, 1062–1082.

(66) Kubillus, M.; Kubar, T.; Gaus, M.; Rezac, 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.

(67) 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.

(68) Niehaus, T. A.; Elstner, M.; Frauenheim, T.; Suhai, S. Application of an approx-

imate density-functional method to sulfur containing compounds. *J. Mol. Struct. (Theochem)* **2001**, *541*, 185–194.

(69) 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.

(70) Vuong, V. Q.; Akkarapattiakal Kuriappan, J.; Kubillus, M.; Kranz, J. J.; Mast, T.; Niehaus, T. A.; Irle, S.; Elstner, M. Parametrization and Benchmark of Long-Range Corrected DFTB2 for Organic Molecules. *J. Chem. Theory Comput.* **2018**, *14*, 115–125.

(71) "Semiempirical Extended Tight-Binding Program Package xtb", https://github.com/grimme-lab/xtb. Accessed: 2021-05-03.

(72) "Molecular Orbital PACkage", https://github.com/openmopac/mopac. Accessed: 2021-12-17.

(73) S Brahmkshatriya, P.; Dobes, P.; Fanfrlik, J.; Rezac, J.; Paruch, K.; Bronowska, A.; Lepsik, M.; Hobza, P. Quantum mechanical scoring: structural and energetic insights into cyclin-dependent kinase 2 inhibition by pyrazolo [1, 5-a] pyrimidines. *Current computer-aided drug design* **2013**, *9*, 118–129.

(74) Stewart, J. J. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *Journal of molecular modeling* **2013**, *19*, 1–32.

(75) Mobley, D. L.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Slochower, D. R.; Shirts, M. R. et al. Escaping atom types in force fields using direct chemical perception. *J. Chem. Theory Comput.* **2018**, *14*, 6076–6092.

(76) Mobley, D.; Bannan, C.; Wagner, J.; Rizzi, A.; Lim, N.; Henry, M. openforcefield/smirnoff99Frosst: Version 1.1.0. 2019; https://doi.org/10.5281/zenodo.3351714.

(77) Qiu, Y.; Smith, D. G.; Boothroyd, S.; Wagner, J.; Bannan, C. C.; Gokey, T.; Jang, H.; Lim, V. T.; Lucas, X.; Tjanaka, B. et al. openforcefield/openforcefields: Version 1.0.0 "Parsley". 2019; https://doi.org/10.5281/zenodo.3483227.

(78) Wagner, J.; Thompson, M.; Dotson, D.; hyejang,; Boothroyd, S.; Rodríguez-Guerra, J. openforcefield/openff-forcefields: Version 2.0.0 "Sage". 2021; https://doi.org/10.5281/zenodo.5214478.

(79) Wagner, J.; Thompson, M.; Mobley, D.; Chodera, J.; Bannan, C.; Rizzi, A.; trevorgokey,; Dotson, D.; Rodríguez-Guerra, J.; Camila, et al. openforcefield/openff-toolkit: 0.10.1 Minor feature and bugfix release. 2021; https://doi.org/10.5281/zenodo.5601736.

(80) Rappé, A. K.; Casewit, C. J.; Colwell, K.; Goddard III, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.

(81) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.

(82) Halgren, T. A. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17*, 520–552.

(83) Landrum, G.; Tosco, P.; Kelley, B.; Martínez, R.; Riniker, S.; Gedeck, P.; Vianello, R.; NadineSchneider,; Kawashima, E.; Dalke, A. et al. rdkit/rdkit: 2021_09_3 (Q3 2021) Release. 2021; https://doi.org/10.5281/zenodo.5773460.

(84) Smith, D. G.; Lolinco, A. T.; Glick, Z. L.; Lee, J.; Alenaizan, A.; Barnes, T. A.; Borca, C. H.; Di Remigio, R.; Dotson, D. L.; Ehlert, S. et al. Quantum Chemistry Common Driver and Databases (QCDB) and Quantum Chemistry Engine (QCEngine): Automation and interoperability among computational chemistry programs. *J. Chem. Phys.* **2021**, *155*, 204801.

(85) 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.

(86) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

(87) Caldeweyher, E.; Mewes, J.-M.; Ehlert, S.; Grimme, S. Extension and evaluation of the D4 London-dispersion model for periodic systems. *Phys. Chem. Chem. Phys.* **2020**, 8499–8512.

(88) Hujo, W.; Grimme, S. Performance of Non-Local and Atom-Pairwise Dispersion Corrections to DFT for Structural Parameters of Molecules with Noncovalent Interactions. *J. Chem. Theory Comput.* **2013**, *9*, 308–315.

(89) Najibi, A.; Goerigk, L. The nonlocal kernel in van der Waals density functionals as an additive correction: An extensive analysis with special emphasis on the B97M-V and ωB97M-V approaches. *J. Chem. Theory Comput.* **2018**, *14*, 5725–5738.

(90) Becke, A. D.; Johnson, E. R. A density-functional model of the dispersion interaction. *J. Chem. Phys.* **2005**, *123*, 154101.

(91) Johnson, E. R.; Becke, A. D. A post-Hartree–Fock model of intermolecular interactions. *J. Chem. Phys.* **2005**, *123*, 024101.

(92) Johnson, E. R.; Becke, A. D. A post-Hartree-Fock model of intermolecular interactions: Inclusion of higher-order corrections. *J. Chem. Phys.* **2006**, *124*, 174104.

(93) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

(94) Smith, D. G.; Burns, L. A.; Patkowski, K.; Sherrill, C. D. Revised damping parameters for the D3 dispersion correction to density functional theory. *J. Phys. Chem. Lett.* **2016**, *7*, 2197–2203.

(95) Witte, J.; Mardirossian, N.; Neaton, J. B.; Head-Gordon, M. Assessing DFT-D3 damping functions across widely used density functionals: Can we do better? *J. Chem. Theory Comput.* **2017**, *13*, 2043–2052.

(96) "Reimplementation of the DFT-D3 program", https://github.com/awvwgk/simple-dftd3. Accessed: 2022-03-16.

(97) "Generally Applicable Atomic-Charge Dependent London Dispersion Correction", https://github.com/grimme-lab/xtb. Accessed: 2021-12-17.

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

(99) Muto, Y. *J. Phys. Math. Soc. Jpn.* **1943**, *17*, 629–631.

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

(101) 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.

(102) Grimme, S.; Hansen, A.; Ehlert, S.; Mewes, J.-M. r2SCAN-3c: A "Swiss army knife" composite electronic-structure method. *J. Chem. Phys.* **2021**, *154*, 064103.

(103) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(104) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple (vol 77, pg 3865, 1996). *Phys. Rev. Lett.* **1997**, *78*, 1396–1396.

(105) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the density functional ladder: Nonempirical meta–generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.

(106) Perdew, J. P.; Tao, J.; Staroverov, V. N.; Scuseria, G. E. Meta-generalized gradient approximation: Explanation of a realistic nonempirical density functional. *J. Chem. Phys.* **2004**, *120*, 6898–6911.

(107) Mardirossian, N.; Head-Gordon, M. Mapping the genome of meta-generalized gradient approximation density functionals: The search for B97M-V. *J. Chem. Phys.* **2015**, *142*, 074111.

(108) Najibi, A.; Goerigk, L. DFT-D4 counterparts of leading meta-generalized-gradient approximation and hybrid density functionals for energetics and geometries. *J. Comput. Chem.* **2020**, *41*, 2562–2572.

(109) Furness, J. W.; Kaplan, A. D.; Ning, J.; Perdew, J. P.; Sun, J. Accurate and numerically efficient r2SCAN meta-generalized gradient approximation. *J. Phys. Chem. Lett.* **2020**, *11*, 8208–8215.

(110) Furness, J. W.; Kaplan, A. D.; Ning, J.; Perdew, J. P.; Sun, J. Correction to "Accurate and Numerically Efficient r2SCAN Meta-Generalized Gradient Approximation". *J. Phys. Chem. Lett.* **2020**, *11*, 9248–9248.

(111) Ehlert, S.; Huniar, U.; Ning, J.; Furness, J. W.; Sun, J.; Kaplan, A. D.; Perdew, J. P.; Brandenburg, J. G. r2SCAN-D4: Dispersion corrected meta-generalized gradient approximation for general chemical applications. *J. Chem. Phys.* **2021**, *154*, 061101.

(112) Zhao, Y.; Truhlar, D. G. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.* **2006**, *125*, 194101.

(113) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(114) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. C* **1994**, *98*, 11623–11627.

(115) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

(116) Ernzerhof, M.; Scuseria, G. E. Assessment of the Perdew–Burke–Ernzerhof exchange-correlation functional. *J. Chem. Phys.* **1999**, *110*, 5029–5036.

(117) Zhao, Y.; Truhlar, D. G. Design of density functionals that are broadly accurate for thermochemistry, thermochemical kinetics, and nonbonded interactions. *J. Phys. Chem. A* **2005**, *109*, 5656–5667.

(118) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

(119) Peverati, R.; Truhlar, D. G. Screened-exchange density functionals with broad accu-

racy for chemistry and solid-state physics. *Phys. Chem. Chem. Phys.* **2012**, *14*, 16187–16191.

(120) Mardirossian, N.; Head-Gordon, M. ω B97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **2016**, *144*, 214110.

(121) Chai, J.-D.; Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **2008**, *128*, 084106.

(122) Mardirossian, N.; Head-Gordon, M. ωB97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9904–9924.

(123) Haasler, M.; Maier, T. M.; Grotjahn, R.; Gückel, S.; Arbuznikov, A. V.; Kaupp, M. A Local Hybrid Functional with Wide Applicability and Good Balance between (De) Localization and Left–Right Correlation. *J. Chem. Theory Comput.* **2020**, *16*, 5645–5657.

(124) Goerigk, L.; Grimme, S. Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals: Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2011**, *7*, 291–309.

(125) Santra, G.; Sylvetsky, N.; Martin, J. M. L. Minimally empirical double-hybrid functionals trained against the GMTKN55 database: revDSD-PBEP86-D4, revDOD-PBE-D4, and DOD-SCAN-D4. *J. Phys. Chem. A* **2019**, *123*, 5129–5143.

(126) Shee, J.; Loipersberger, M.; Rettig, A.; Lee, J.; Head-Gordon, M. Regularized second-order Møller–Plesset theory: A more accurate alternative to conventional MP2 for

noncovalent interactions and transition metal thermochemistry for the same computational cost. *J. Phys. Chem. Lett.* **2021**, *12*, 12084–12097.

(127) Bursch, M.; Neugebauer, H.; Ehlert, S.; Grimme, S. Dispersion corrected r2SCAN based global hybrid functionals: r2SCANh, r2SCAN0, and r2SCAN50. *J. Chem. Phys.* **2022**, *156*, 134105.

(128) Nalin de Silva, K.; Goodman, J. M. What is the smallest saturated acyclic alkane that cannot be made? *J. Chem. Inf. Model.* **2005**, *45*, 81–87.

(129) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N2, and H2O. *Chem. Phys. Lett.* **1998**, *286*, 243–252.

(130) Jurečka, P.; Šponer, J.; Černỳ, 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.

(131) Marshall, M. S.; Burns, L. A.; Sherrill, C. D. Basis set convergence of the coupled-cluster correction, δ MP2 CCSD (T): Best practices for benchmarking non-covalent interactions and the attendant revision of the S22, NBC10, HBC6, and HSG databases. *J. Chem. Phys.* **2011**, *135*, 194102.

(132) Martin, J. M. L.; de Oliveira, G. Towards standard methods for benchmark quality ab initio thermochemistry—W1 and W2 theory. *J. Chem. Phys.* **1999**, *111*, 1843–1856.

(133) Martin, J. M. L.; Parthiban, S. *Quantum-Mechanical Prediction of Thermochemical Data*; Springer, 2001; pp 31–65.

(134) Császár, A. G.; Allen, W. D.; Schaefer, H. F. In pursuit of the ab initio limit for conformational energy prototypes. *J. Chem. Phys.* **1998**, *108*, 9751–9764.

(135) Perdew, J. P.; Schmidt, K. Jacob's ladder of density functional approximations for the exchange-correlation energy. AIP Conference Proceedings. 2001; pp 1–20.

(136) Kruse, H.; Grimme, S. A geometrical correction for the inter-and intra-molecular basis set superposition error in Hartree-Fock and density functional theory calculations for large systems. *J. Chem. Phys.* **2012**, *136*, 04B613.

(137) Brandenburg, J. G.; Alessio, M.; Civalleri, B.; Peintinger, M. F.; Bredow, T.; Grimme, S. Geometrical correction for the inter-and intramolecular basis set superposition error in periodic density functional theory calculations. *J. Phys. Chem. A* **2013**, *117*, 9282–9292.

(138) Caldeweyher, E.; Brandenburg, J. G. Simplified DFT methods for consistent structures and energies of large systems. *J. Phys.: Condens. Matter* **2018**, *30*, 213001.

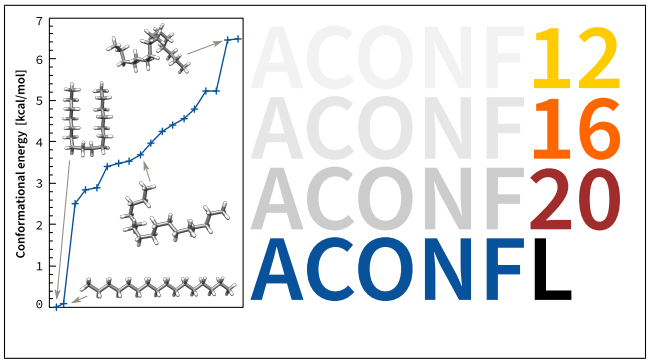
(139) Maurer, L. R.; Bursch, M.; Grimme, S.; Hansen, A. Assessing Density Functional Theory for Chemically Relevant Open-Shell Transition Metal Reactions. *J. Chem. Theory Comput.* **2021**, *17*, 6134–6151.

(140) Pracht, P.; Grimme, S. Calculation of absolute molecular entropies and heat capacities made simple. *Chemical science* **2021**, *12*, 6551–6568.

(141) Pitonak, M.; Heßelmann, A. Accurate intermolecular interaction energies from a combination of MP2 and TDDFT response theory. *J. Chem. Theory Comput.* **2010**, *6*, 168–178.

(142) Santra, G.; Semidalas, E.; Martin, J. M. L. Surprisingly Good Performance of XYG3 Family Functionals Using a Scaled KS-MP3 Correlation. *J. Phys. Chem. Lett.* **2021**, *12*, 9368–9376.

TOC Graphic


Conformational energy [kcal/mol]
0
1
2
3
4
5
6
7
ACONF12
ACONF16
ACONF20
ACONFL