# Balancing Cost and Accuracy: Method Development, Assessment, and Application of Quantum Chemical Methods

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Thomas Gasevic
aus
Dillenburg

Bonn 2025

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

| | |
|---|---|
| Gutachter/Betreuer: | Prof. Dr. Stefan Grimme |
| Gutachter: | Prof. Dr. Thomas Bredow |

| | |
|---|---|
| Tag der Promotion: | 25.02.2026 |
| Erscheinungsjahr: | 2026 |

# Publications

The following publications serve as parts of this thesis.

1. T. Gasevic, J. B. Stückrath, S. Grimme, and M. Bursch, *Optimization of the $r^2$SCAN-3c Composite Electronic-Structure Method for Use with Slater-Type Orbital Basis Sets*, J. Phys. Chem. A **126**.23 (2022) 3826, DOI: `10.1021/acs.jpca.2c02951`
2. T. Gasevic, J. B. Kleine Büning, S. Grimme, and M. Bursch, *Benchmark Study on the Calculation of $^{207}$Pb NMR Chemical Shifts*, Inorg. Chem. **63**.11 (2024) 5052, DOI: `10.1021/acs.inorgchem.3c04539`
3. T. Gasevic, M. Bursch, Q. Ma, S. Grimme, H.-J. Werner, and A. Hansen, *The p-block challenge: assessing quantum chemistry methods for inorganic heterocycle dimerizations*, Phys. Chem. Chem. Phys. **26**.18 (2024) 13884, DOI: `10.1039/d3cp06217a`
4. T. Gasevic, M. Müller, J. Schöps, S. Lanius, J. Hermann, S. Grimme, and A. Hansen, *Chemical Space Exploration with Artificial "Mindless" Molecules*, J. Chem. Inf. Model. **65**.18 (2025) 9576, DOI: `10.1021/acs.jcim.5c01364`

5a. T. Gasevic, M. Bamberg, J. Wicke, M. Bolte, A. Virovets, H.-W. Lerner, S. Grimme, A. Hansen, M. Wagner, and M. Bursch, *Confined Lewis Pairs: Investigation of the $X^- \rightarrow Si_{20}$ Interaction in Halogen-Encapsulating Silafulleranes*, Angew. Chem., Int. Ed. **63**.6 (2024) e202314238, DOI: `10.1002/anie.202314238`

5b. T. Gasevic, M. Bamberg, J. Wicke, M. Bolte, A. Virovets, H.-W. Lerner, S. Grimme, A. Hansen, M. Wagner, and M. Bursch, *Confined Lewis Pairs: Investigation of the $X^- \rightarrow Si_{20}$ Interaction in Halogen-Encapsulating Silafulleranes*, Angew. Chem. **136**.6 (2024) e202314238, DOI: `10.1002/ange.202314238`

6 M. Bamberg, T. Gasevic, M. Bolte, A. Virovets, H.-W. Lerner, S. Grimme, M. Bursch, and M. Wagner, *Regioselective Derivatization of Silylated [20]Silafulleranes*, J. Am. Chem. Soc. **145**.20 (2023) 11440, DOI: `10.1021/jacs.3c03270`

During the period of this thesis, the following further studies were carried out:

7. M. Bamberg, T. Gasevic, M. Bolte, A. Virovets, H.-W. Lerner, S. Grimme, M. Bursch, and M. Wagner, *Brominated [20]silafulleranes: pushing the limits of steric loading*, Chem. Commun. **59**.48 (2023) 7459, DOI: `10.1039/d3cc02142d`

8. M. Bamberg, T. Gasevic, T. G. Saint-Denis, J. Martinez Fernandez, S. Grimme, and T. D. Tilley, *Synthesis of Metallostannylenes from Transition Metal Polyhydride Complexes*, J. Am. Chem. Soc. (2025), DOI: `10.1021/jacs.5c14064`
9. T. Gasevic, C. Plett, L. Wittmann, I. Neira, C. Peinador, M. D. García, and A. Hansen, *Supramolecular Host-Guest Complexation Dynamics by Cost-Efficient Electronic Structure Methods*, ChemRxiv preprint (2025), DOI: `10.26434/chemrxiv-2025-3z07r` (Submitted to Chem. Eur. J., currently in second revision)

Prior to this thesis, two more works were published:

10. M. Bursch, T. Gasevic, J. B. Stückrath, and S. Grimme, *Comprehensive Benchmark Study on the Calculation of $^{29}$Si NMR Chemical Shifts*, Inorg. Chem. **60**.1 (2021) 272, DOI: `10.1021/acs.inorgchem.0c02907`
11. J. B. Stückrath, T. Gasevic, M. Bursch, and S. Grimme, *Benchmark Study on the Calculation of $^{119}$Sn NMR Chemical Shifts*, Inorg. Chem. **61**.9 (2022) 3903, DOI: `10.1021/acs.inorgchem.1c03453`

The following scientific contributions have been presented at conferences and scientific exchanges:

1. 60th Symposium on Theoretical Chemistry. Braunschweig, **2024**, Poster "IHD302: Assessing Quantum Chemistry Methods for Inorganic Heterocycle Dimerizations".
2. Exchange with the theoretical chemistry department in Bochum, **2024**, Talk "Comprehensive Computational Studies of NMR Chemical Shifts for Group 14 Elements".
3. Amsterdam Modeling Suite (AMS) Webinar, **2023**, Talk "r$^2$SCAN-3c(STO): Efficient, robust & reliable composite DFT method".
4. 58th Symposium on Theoretical Chemistry, Heidelberg, **2022**, Poster "Comprehensive Benchmark Study on the Caluclation of $^{29}$Si NMR Chemical Shifts".
5. Bunsen-Tagung 2022, Gießen, **2022**: "Optimization of the r$^2$SCAN-3c Composite Electronic-Structure Method for Use with Slater-Type Orbital Basis Sets".
6. 19th International Conference on Density Functional Theory and its Applications, Brussels, **2022**: Poster "Optimization of the r$^2$SCAN-3c Composite Electronic-Structure Method for Use with Slater-Type Orbital Basis Sets".
7. 2nd International Conference on Noncovalent Interactions (ICNI), Strasbourg, **2021**: Poster "Optimization of the r$^2$SCAN-3c Composite Electronic-Structure Method for Use with Slater-Type Orbital Basis Sets".

# Abstract

This thesis advances the field of computational chemistry through the development, systematic benchmarking, and targeted application of quantum chemical methods. A central methodological contribution is the extension of the r$^2$SCAN-3c composite density functional approach to Slater-type orbitals, thereby demonstrating that composite schemes can be successfully transferred to alternative basis set types. Due to the application of an all-electron basis set and the implementation into the Amsterdam Modeling Suite, this development also enables, for the first time, the application of explicit relativistic treatments within the "3c" composite family, expanding the scope of such methods to heavier elements and more challenging systems.

Further, comprehensive benchmark studies were conducted across diverse datasets, including $^{207}$Pb Nuclear Magnetic Resonance (NMR) chemical shifts (PbS50), inorganic heterocycle dimerizations (IHD302), and artificially generated "mindless" molecules (MB2061). These investigations provided a rigorous assessment of density functional approximations, semiempirical quantum mechanical methods, and machine-learning interatomic potentials. The results revealed systematic strengths and limitations across the different classes of methods, underscoring that accuracy is not universal but depends critically on both the targeted property and the underlying chemical environment. In this context, the recently developed semiempirical quantum mechanical g-xTB method and modern machine-learning potentials, such as UMA-sm, have emerged as particularly promising, offering hybrid DFT-level accuracy at a fraction of the computational cost. However, challenges regarding robustness and error distribution remain for machine-learning approaches.

The practical relevance of the drawn insights was demonstrated through applications to silafullerane systems, where theoretical investigations complemented experimental studies by rationalizing unusual bonding motifs, predicting hydrogenation mechanisms, and accurately reproducing measured $^{35}$Cl NMR shifts. These case studies highlight the dual explanatory and predictive role of theory in advancing chemical understanding.

Taken together, the results of this thesis demonstrate that no single "best" method exists in quantum chemistry. Instead, reliability arises from carefully validated methodological choices tailored to the problem at hand, supported by systematic benchmarking and informed by practical applications. The findings emphasize the importance of composite methods, advanced semiempirical models, and machine learning potentials in balancing accuracy and computational efficiency, while also addressing the need for rigorous stress testing using challenging benchmark sets. Ultimately, this work establishes strategies for extending the reach, reliability, and robustness of computational chemistry, laying a foundation for future methodological advancements and for bridging the gap between theoretical predictions and experimental practice across an increasingly broad chemical space.

# Contents

## CHAPTER 1

# Introduction

Chemistry stands at the heart of the natural sciences, uniting the realm of atoms and molecules with observable phenomena in biology, materials science, and environmental science. From ancient alchemy[13] to modern pharmaceuticals,[14] chemists have consistently sought to understand how atoms combine, how bonds form, and how molecular structures influence reactivity. Today, these challenges have to be solved across various disciplines, e.g., in designing catalysts to accelerate clean-energy reactions,[15–17] engineering drug candidates to achieve biological targets,[18–20] and crafting materials whose properties emerge from precise atomic arrangements.[21,22] Yet, the complexity of chemical systems often exceeds intuition, requiring quantitative approaches that can predict and ultimately guide the behavior of molecules before they are synthesized.

Over the past decades, computational chemistry has risen in popularity as a powerful tool complementing experimental approaches (cf. Figure 1.1).[23–28] Typically, chemical research projects begin with a defined target system or property of interest. Candidate structures are then either synthesized experimentally or explored computationally. Experimental and theoretical results work in tandem to verify findings and make informed, novel adjustments until the desired objective is achieved. Theoretical chemistry combines mathematical equations with the rules of physics to describe chemical problems. This is fundamentally based on the Schrödinger equation,[29] which defines the relationship between the total energy of a system and its quantum-mechanical wave function, using the Hamiltonian operator. However, the equation cannot be solved analytically for molecules with multiple electrons, as the electron-electron interaction term creates a high-dimensional correlation problem. Therefore, approximations are required. As a subfield of theoretical chemistry, computational chemistry translates the Schrödinger equation into manageable numerical problems and applies physically motivated approximations. This enables the optimization of molecular geometries, as well as the computation of

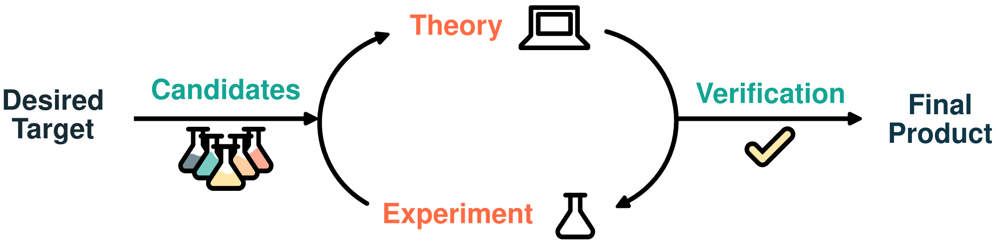


Figure 1.1: General framework for the modern interplay between theoretical and experimental chemistry.

reaction energies, spectroscopic properties, and numerous other features. Modern software typically only requires a rough guess of the molecular geometry, the system charge, and the spin multiplicity as input, making high-level calculations accessible to a broad scientific audience. However, this ease of use belies a key challenge: There is no feasible, universal method that can describe *every* chemical system correctly. Instead, computational chemists must choose from a hierarchy of approximations, each offering a specific balance of computational cost and accuracy. This leaves the open and ever-evolving question of which method is best suited to a given chemical problem.[30–34]

Answering this question has become a central focus of ongoing method development. The ultimate goal is to design approaches that not only improve theoretical accuracy but also perform reliably and efficiently across a wide variety of chemically relevant systems. To achieve this, benchmarking has become a key component of computational chemistry.[23,35] Through systematic comparison to experimental data or highly accurate theoretical references, benchmarking studies help to quantify how well a given method performs in practice. Typical benchmark targets include thermochemical quantities such as reaction energies and barrier heights,[36–38] as well as spectroscopic observables like NMR chemical shifts[11,12,39] or vibrational frequencies,[40–42] among many others. These properties are especially valuable because they are experimentally measurable, chemically meaningful, and sensitive to the electronic structure of a molecule. The insights gained from such benchmarks not only help to identify strengths and weaknesses of existing methods but also support the development of new approaches, including fitting procedures used in force fields (FFs) and electronic structure methods, as well as data-driven machine learning approaches, such as machine learning interatomic potentials (MLIPs). Furthermore, these tests help to select suitable methods for specific problems by highlighting which approximations perform best under given conditions.

The broader goal behind these efforts is to establish computational protocols that can be applied with confidence across diverse areas of chemistry. When such methods reliably guide synthetic strategies, rationalize experimental observations, and provide insights into systems that are challenging or inaccessible to direct measurements, they not only accelerate research but also deepen our understanding of chemical phenomena. In this way, computations move beyond interpretation and become tools of prediction and discovery. Their value is particularly evident in joint theoretical and experimental studies, where complementary insights enable more comprehensive analyses than either approach could achieve alone.

Through the careful development, evaluation, and application of computational methods, this work aims to contribute to a more reliable and insightful foundation for theoretical chemistry, not only complementing but also actively advancing experimental discovery. The structure of this thesis reflects this objective and is organized as follows: The current chapter outlines the scientific landscape and defines the overarching goal of assessing and advancing electronic structure methods for challenging systems. Chapter 2 provides the theoretical background, outlining the core principles of electronic structure theory, from mean-field approximations to advanced correlation treatments and practical quantum-chemical workflows. Chapter 3 focuses on method development, discussing the design and implementation of the "Swiss army knife" $r^2$SCAN-3c[43] for the use with Slater-type orbitals. Chapters 4 to 6 present comprehensive benchmarking results, including the analysis of $^{207}$Pb NMR chemical shifts, dimerization energies of structures with $p$-block elements, and so-called "mindless" reaction energies, identifying systematic trends and key sources of error. Chapters 7 and 8 then turn to the application of quantum chemistry (QC) to understand experimental results, exploring the computational modeling of the unique $Si_{20}$ silafullerane species. Finally, Chapter 9 summarizes the key findings and offers an outlook on future directions for improving the reliability and applicability of computational methods.

CHAPTER 2

# Theory

## 2.1 General

A significant portion of computational chemistry focuses on striking a balance between computational cost and accuracy. To achieve this, various methods can be applied, where the most commonly used include wave function theory (WFT), density functional theory (DFT), semiempirical quantum mechanical (SQM) methods, force fields (FFs), and machine learning interatomic potentials (MLIPs) (cf. Figure 2.1). This chapter outlines the theoretical foundations of the methods used in this work. Large portions of the discussion are based on Ref. [44]. All equations are given in atomic units.

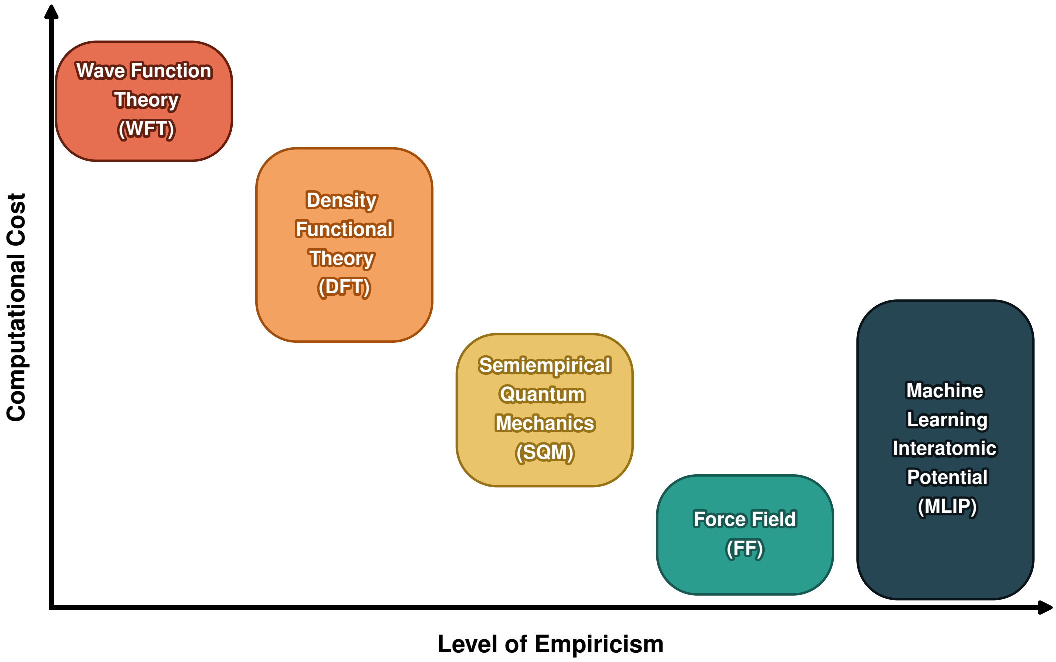


Figure 2.1: Comparing the computational cost and the level of empiricism of computational methods for chemical problems.

## 2.2 Electronic Structure Theory

Understanding and predicting the properties of molecules and materials at the quantum level typically requires the electronic energy of a system. For most applications, the non-relativistic, time-independent Schrödinger equation[29] is applied:

$$\hat{H}\Psi = E\Psi, \tag{2.1}$$

where $\hat{H}$ is the Hamiltonian operator, $\Psi$ the many-electron wave function, and $E$ the total energy. The Hamiltonian operator includes the operators for the kinetic energy of the nuclei $\hat{T}_{\mathrm{n}}$ and of the electrons $\hat{T}_{\mathrm{e}}$, as well as for the nucleus-electron attraction $\hat{V}_{\mathrm{ne}}$, the electron-electron repulsion $\hat{V}_{\mathrm{ee}}$ and nucleus-nucleus repulsion $\hat{V}_{\mathrm{nn}}$:

$$\hat{H} = \hat{T}_{\mathrm{n}} + \hat{T}_{\mathrm{e}} + \hat{V}_{\mathrm{ne}} + \hat{V}_{\mathrm{ee}} + \hat{V}_{\mathrm{nn}}. \tag{2.2}$$

This equation can be simplified using the Born-Oppenheimer approximation, which decouples the motion of electrons and nuclei based on their vastly different masses and velocities. Since electrons move much faster than nuclei, the nuclear positions can be treated as fixed when solving the electronic problem. As a result, $\hat{V}_{\mathrm{nn}}$ becomes a constant for a given nuclear arrangement. Consequently, the total Hamiltonian can be partitioned into separate nuclear and electronic contributions, with the electronic Hamiltonian formally defined as:

$$\hat{H}_{\mathrm{e}} = \hat{T}_{\mathrm{e}} + \hat{V}_{\mathrm{ne}} + \hat{V}_{\mathrm{ee}} + V_{\mathrm{nn}} = \hat{h} + \hat{V}_{\mathrm{ee}} + \hat{V}_{\mathrm{nn}}, \tag{2.3}$$

with the following contributions:

$$\hat{T}_{\mathrm{e}} = -\sum_{i}^{N_{\mathrm{elec}}} \frac{1}{2}\nabla_i^2; \quad \nabla^2 = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) \tag{2.4}$$

$$\hat{V}_{\mathrm{ne}} = -\sum_{i}^{N_{\mathrm{elec}}}\sum_{A}^{M_{\mathrm{nuc}}} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|}, \tag{2.5}$$

$$\hat{V}_{\mathrm{ee}} = \sum_{i}^{N_{\mathrm{elec}}}\sum_{j>i}^{N_{\mathrm{elec}}} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \tag{2.6}$$

$$\hat{V}_{\mathrm{nn}} = \sum_{A}^{N_{\mathrm{nuc}}}\sum_{A<B}^{N_{\mathrm{nuc}}} \frac{1}{|\mathbf{R}_A - \mathbf{R}_B|}, \tag{2.7}$$

where $\hat{h}$ is the one-electron operator, that includes $\hat{T}_{\mathrm{e}}$ and $\hat{V}_{\mathrm{ne}}$, $i$ and $j$ are the indices of the individual electrons in the system, $N_{\mathrm{elec}}$ the number of electrons, $M_{\mathrm{nuc}}$ the number of nuclei, $Z_A$ the nuclear charge of the atom $A$, $\mathbf{r}_i$ the coordinates of the electron $i$, and $\mathbf{R}_A$ the coordinates of the nucleus $A$. For a molecule containing multiple electrons and nuclei, the equation cannot be solved analytically because the electron-electron interaction term creates a non-separable, high-dimensional correlation problem. To address the challenges of solving the Schrödinger equation, electronic structure theory offers a plethora of approximations. These methods differ in how they represent the wave function (or other fundamental quantities, such as the electron density), how they account for electron correlation, and what kind of scaling behavior and accuracy they exhibit. The core classes of methods include WFT like Hartree-Fock (HF)[45–47] theory and the coupled cluster (CC) method,[48] electron density-based methods like DFT,[49,50] further SQM methods,[51,52] classical FF models,[53,54] and emerging data-driven MLIPs.[25] Each of these approximations reflects a distinct trade-off between physical rigor, computational cost, and number of empirical parameters. As many explanations involve integrals over the wave function $\Psi$, they are here expressed using Dirac's *bra-ket* notation:[55]

$$\langle\Psi| \equiv \Psi^*; |\Psi\rangle \equiv \Psi \tag{2.8}$$

$$\langle\Psi|\Psi\rangle = \int \Psi^*\Psi \mathrm{d}\mathbf{x} \tag{2.9}$$

$$\langle\Psi|\hat{H}|\Psi\rangle = \int \Psi^*\hat{H}\Psi \mathrm{d}\mathbf{x} \tag{2.10}$$

The bra $\langle n|$ represents the complex conjugate of a wave function labeled by the quantum number $n$. Conversely, the ket $|m\rangle$ denotes a wave function with quantum number $m$. Together, the braket $\langle n|\hat{O}|m\rangle$ implies integration over all coordinates and determines the expectation value of the operator $\hat{O}$. In the following subchapters, the main concepts and practical implications of the most common methods are discussed, beginning with Hartree-Fock theory.

### 2.2.1 Hartree-Fock (HF) Theory

The HF method is a cornerstone of wave function-based quantum chemistry, serving as the starting point for many post-HF correlation techniques. It was developed over several decades beginning in the 1920s and 30s, with important contributions by Douglas Hartree, who first introduced a self-consistent field (SCF) approach,[45] and Vladimir Fock, who extended it to include the antisymmetry requirement of fermionic wave functions *via* Slater determinants.[46] The central approximation of HF theory is that the total electronic wave function $\Psi$ can be represented by a single Slater determinant constructed from one-electron molecular orbitals (MOs) $\phi$ (cf. eq. (2.11)), each defined as the product of a spatial and a spin function, represented as $\phi_i(a)$ where $i$ denotes the spatial and $a$ the spin part.

$$\Psi \approx \Phi_{\mathrm{HF}} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_2(1) & \dots & \phi_N(1) \\ \phi_1(2) & \phi_2(2) & \dots & \phi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(N) & \phi_2(N) & \dots & \phi_N(N) \end{vmatrix} ; \langle\phi_i|\phi_j\rangle = \delta_{ij} \tag{2.11}$$

This ansatz ensures that the wave function is antisymmetric with respect to the exchange of any two electrons, in accordance with the Pauli exclusion principle,[56] which states that no two electrons can occupy the same set of quantum numbers. HF employs a Slater determinant as a trial wave function and determines the optimal molecular orbitals by minimizing the total electronic energy using the SCF approach. This optimization is carried out using the variational principle, which guarantees that the expectation value of the energy, evaluated with any trial wave function, is always greater than or equal to the exact ground-state energy. In Hartree-Fock theory, the energy is minimized with respect to the spin orbitals, subject to the constraint that they remain orthonormal. The energy expectation value of the wave function, evaluated with the HF Hamiltonian, leads to the variation of the total energy through the Fock operator:

$$\hat{F}_i = \hat{h}_i + \sum_{j}^{N_{\mathrm{elec}}} (\hat{J}_j - \hat{K}_j). \tag{2.12}$$

The application of this operator to the final spin orbitals yields the HF energy expression ($E_{\mathrm{HF}}$):

$$E_{\mathrm{HF}} = \sum_{i}^{N_{\mathrm{elec}}} \epsilon_i - \frac{1}{2} \sum_{i}^{N_{\mathrm{elec}}} \sum_{j}^{N_{\mathrm{elec}}} (J_{ij} - K_{ij}) + E_{\mathrm{nn}}, \tag{2.13}$$

$$\epsilon_i = \langle \phi_i | \hat{F}_i | \phi_i \rangle = h_i + \sum_{j}^{N_{\mathrm{elec}}} (J_{ij} - K_{ij}), \tag{2.14}$$

where $\epsilon_i$ are the orbital energies, $\phi_i$ the molecular orbitals, $\hat{J}_j$ and $\hat{K}_j$ are the Coulomb and exchange operators, $E_{\mathrm{nn}}$ accounts for the nucleus-nucleus repulsion, and $\hat{h}_i$ is the one-electron operator that includes $\hat{T}_{\mathrm{e}}$ and $\hat{V}_{\mathrm{ne}}$. Notably, the total energy of HF is not a simple sum over orbital energies $\epsilon_i$, since the Coulomb and exchange contributions are double-counted in that sum. This is corrected by subtracting half of the total Coulomb and exchange interaction from the final energy expression. HF is a mean-field theory, meaning that each electron interacts with the average field generated by all other electrons. While both $J_{ij}$ and $K_{ij}$ arise from the electron-electron interaction, $J_{ij}$ represents the classical Coulomb repulsion between electrons $i$ and $j$, whereas $K_{ij}$ reflects the exchange interaction, which stems from the antisymmetry of the wave function and leads to an effective attractive interaction between electrons of the same spin.

### 2.2.2 Electron Correlation Methods

As electrons interact with each other, their motions are not independent but correlated. The resulting electron correlation energy $E_{\mathrm{corr}}$ is generally defined as the difference between the exact non-relativistic electronic energy $E_{\mathrm{exact,non\text{-}rel}}$ and the energy obtained from HF theory $E_{\mathrm{HF}}$:

$$E_{\mathrm{corr}} = E_{\mathrm{exact,non\text{-}rel}} - E_{\mathrm{HF}} \tag{2.15}$$

Electron correlation can be discussed from two complementary viewpoints. The first distinguishes between *Fermi and Coulomb correlation.* The correlation arising from symmetry requirements of the wave function is known as *Fermi correlation*, while the correlation arising from mutual Coulomb repulsion between electrons is called *Coulomb correlation.*[57] HF theory accounts for Fermi correlation through the exchange term in the wave function, but not the Coulomb correlation.

Another, equally common viewpoint distinguishes between *dynamic and static correlation. Dynamic correlation* refers to the instantaneous interactions between electrons, for example, between those that occupy the same spatial orbital. *Static correlation*, on the other hand, is essential for systems where a single Slater determinant is no longer an adequate approximation to the ground-state wave function. This typically occurs in cases with near-degenerate orbitals, such as bond dissociations, diradicals, or transition metal complexes. In such cases, multiple electronic configurations contribute significantly to the ground state, and a single-determinant approach such as HF is insufficient. Despite these limitations, HF provides a variational upper bound to the exact ground-state energy and serves as a well-defined reference point for quantifying electron correlation. To recover the missing correlation energy, different post-HF methods have been developed.

One of the simplest post-HF approaches is the second-order Møller-Plesset perturbation theory (MP2),[58] which treats electron correlation as a perturbation to the HF reference and yields the following additive energy correction in a single, non-iterative step:

$$E_{\mathrm{MP2}} = \sum_{i<j}^{\mathrm{occ}} \sum_{a<b}^{\mathrm{virt}} \frac{\left|\langle \phi_i \phi_j | \phi_a \phi_b \rangle - \langle \phi_i \phi_j | \phi_b \phi_a \rangle\right|^2}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}, \tag{2.16}$$

where $\epsilon$ denotes the orbital energies, $\phi$ the molecular orbitals, $i/j$ the occupied orbitals, and $a/b$ virtual orbitals. While MP2 partly captures dynamic correlation computationally cost-effectively, it remains perturbative in nature and can become unreliable in systems with small HOMO-LUMO gaps, where the denominator approaches zero. In such cases, the MP2 correlation energy may diverge toward $-\infty$, yielding erratic results. More sophisticated and systematically improvable approaches include Coupled Cluster theory, which is based on an exponential ansatz for the wave function:

$$|\Psi_{\mathrm{CC}}\rangle = e^{\hat{T}} |\Phi_{\mathrm{HF}}\rangle , \tag{2.17}$$

where $\hat{T} = \hat{T}_1 + \hat{T}_2 + \ldots$ is the cluster operator composed of excitation operators acting on the HF reference wave function $|\Phi_{\mathrm{HF}}\rangle$. As a result of the exponential form of the ansatz, even a truncated expansion can partly account for higher-order excitations, enabling an accurate treatment of dynamic correlation. By truncating at different levels, the CC method can be systematically improved. In particular, CC theory, which includes single and double excitations, as well as perturbative triples, is commonly referred to as canonical CCSD(T) and is widely regarded as the "gold standard" in computational chemistry, frequently used as a reference for benchmarking purposes. To make such high-level WFT methods applicable to larger molecular systems, various approximate schemes have been developed that significantly reduce the computational cost while aiming to retain the accuracy of canonical CCSD(T). Examples for these are the domain-based local pair natural orbital coupled-cluster theory (DLPNO-CCSD(T)),[59–62] and the pair natural orbital local coupled-cluster theory (PNO-LCCSD(T)).[63–68] Both approaches exploit that electron correlation is mainly a local phenomenon, meaning that electrons at larger distances only contribute little to the correlation energy. Therefore, only relevant orbital pairs are identified and computed at the CC level, while the rest is computed at the MP2 level, leading to a significantly lower computational cost.

For cases requiring a proper treatment of static correlation, multi-reference methods, such as the complete active space self-consistent field (CASSCF)[69] method or the multi-reference configurational interaction (MR-CI)[70–72] method, are employed. These approaches construct the wave function as a linear combination of multiple Slater determinants, allowing for a more flexible description of electronic structure in systems where several configurations are nearly degenerate. Further details on these methods can be found in the corresponding literature.[73,74]

### 2.2.3 Density Functional Theory (DFT)

An alternative approach for the inclusion of correlation is presented by density functional theory. DFT offers an alternative framework to WFT-based methods by reformulating the many-electron problem in terms of the electron density $\rho(\mathbf{r})$, a three-dimensional scalar function, rather than the high-dimensional many-electron wave function $\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$. This approach is rooted in the two Hohenberg-Kohn theorems:[49] The first theorem establishes that, for a given system, the external potential, and thereby the total energy, is a unique functional of the ground-state electron density, which confirms the existence of an exact density functional. Notably, its explicit form remains unknown. The second theorem establishes a variational principle, stating that the ground-state energy functional $E[\rho]$ is minimized by the exact ground-state density. In this context, a functional is a mapping that takes a complete function, in this case the electron density, as input and returns a scalar quantity such as the total energy. Early attempts to approximate the total energy directly as a functional of the electron density, including all contributions from kinetic energy, external potential, Coulomb, exchange, and correlation, proved inadequate. These approaches, collectively known as orbital-free DFT, generally perform less accurately than wave function-based methods, particularly for molecular systems. The primary challenge lies in the kinetic energy functional $T[\rho]$, which lacks a reliable and transferable analytic form expressed solely in terms of the density.

To overcome this limitation, Walter Kohn and Lu Jeu Sham introduced the KS-DFT formalism in 1965,[50] which has since become the standard approach commonly referred to as only "DFT", including throughout this thesis. The density is here computed as

$$\rho(\mathbf{r}) = \sum_{i}^{N_{\mathrm{MO}}} |\phi_i(\mathbf{r})|^2, \tag{2.18}$$

with $N_{\mathrm{MO}}$ being the number of MOs. The formalism separates the kinetic energy into two components. The dominant part corresponds to the kinetic energy of a hypothetical system of non-interacting electrons $T_{\mathrm{S}}[\phi]$ that reproduces the same ground-state density as the fully interacting system, analogous to the mean-field picture of HF theory. The subscript S denotes that the kinetic energy originates from a Slater determinant. This contribution can be calculated explicitly by reintroducing molecular orbitals:

$$T_{\mathrm{S}}[\rho] = \sum_{i}^{N_{\mathrm{elec}}} \langle \phi_i | -\tfrac{1}{2}\nabla^2 | \phi_i \rangle \,. \tag{2.19}$$

The remaining part, which accounts for the kinetic energy difference between the interacting and non-interacting systems, is unknown and is absorbed into the exchange-correlation functional $E_{\mathrm{xc}}[\rho]$. As a result, the total energy functional is partitioned as:

$$E_{\mathrm{DFT}}[\rho] = \underbrace{T_{\mathrm{S}}[\rho] + E_{\mathrm{ne}}[\rho] + J[\rho]}_{\text{known}} + \underbrace{E_{\mathrm{xc}}[\rho]}_{\text{unkown}} \,. \tag{2.20}$$

Here, $E_{\mathrm{ne}}[\rho]$ is the nuclear-electron attraction, $J[\rho]$ the classical Coulomb energy, and $E_{\mathrm{xc}}[\rho]$ the exchange-correlation functional, which accounts for the difference between the true interacting and the non-interacting kinetic energy, and includes the non-classical contributions to the electron-electron interaction, namely exchange and correlation:

$$E_{\mathrm{xc}}[\rho] = (T[\rho] - T_{\mathrm{S}}[\rho]) + (E_{\mathrm{ee}}[\rho] - J[\rho]). \tag{2.21}$$

While HF and DFT share a similar formal framework, they differ primarily in the level at which the approximation is applied. In DFT, the formalism remains exact up to the point where the exchange-correlation functional is required. Since its exact form is unknown, practical implementations depend on approximations to this term, commonly referred to as density functional approximations (DFAs). In practice, $E_{\mathrm{xc}}$ is usually decomposed into an exchange part $E_{\mathrm{x}}$ and a correlation part $E_{\mathrm{c}}$, which are approximated separately. To systematically classify DFAs, John P. Perdew proposed the concept of the "Jacob's Ladder of Density Functional Approximations" in 2001.[31] This framework categorizes functionals into distinct rungs based on the ingredients they use. As one climbs the ladder, additional information about the electronic system is incorporated, typically leading to improved accuracy at the cost of increased computational effort (cf. Figure 2.2).

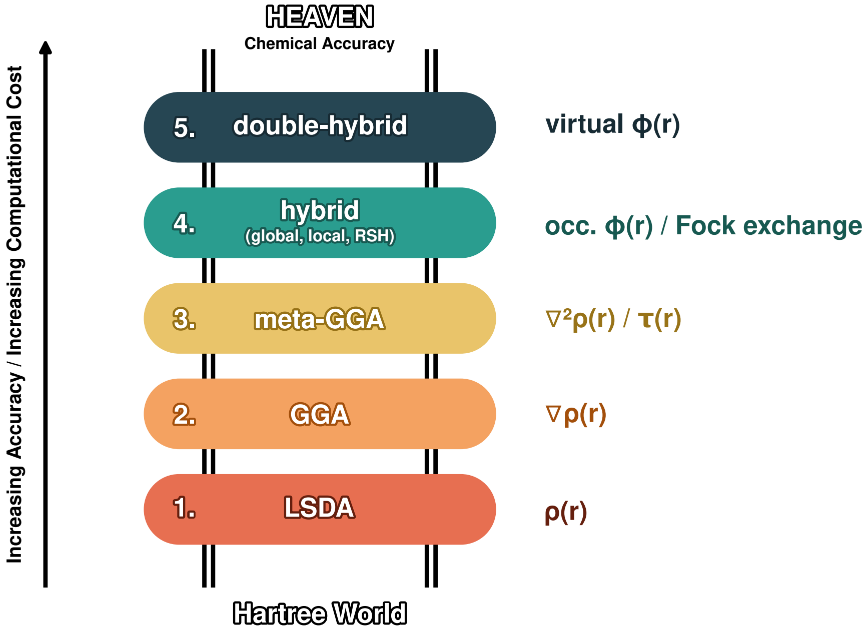


Figure 2.2: Jacob's ladder of density functional theory approximations. The respective ingredient that is added at each rung (e.g. $\rho$, $\nabla\rho$, etc.) is listed, and each higher rung contains all components of the lower rungs.

### Local (Spin) Density Approximation (LSDA)

The lowest rung of the *Jacob's Ladder* is the local density approximation (LDA), which models the exchange and correlation energy based solely on the *local* value of the electron density. It approximates the electron density $\rho$ at each point in space as that of a uniform electron gas. The correlation energy in LDAs is derived from the uniform electron gas model by interpolating between the known high- and low-density limits. For open-shell systems, the local spin density approximation (LSDA) is applied, in which individual spin densities are used instead of the total electron density. For closed-shell systems, the LSDA and LDA approaches yield the same result. Well-known examples of LSDA functionals include the parametrizations by Vosko, Wilk, and Nusair (VWN),[75] or the parameterizations by Perdew and Wang (PW).[76] Although LSDAs are computationally efficient, their accuracy is generally insufficient for systems with strongly varying electron densities, such as most molecular systems. Another significant source of error in DFT is the self-interaction error (SIE). Unlike in HF, electrons in DFT interact with the total electron density rather than being treated individually, which results in an unphysical self-interaction of each electron. Due to these limitations, LSDAs are rarely employed in modern quantum chemistry.

### Generalized Gradient Approximation (GGA)

The previously mentioned limitation on the accuracy is addressed on the second rung of the *Jacob's Ladder* by incorporating the gradient of the electron density. This leads to the generalized gradient approximation (GGA), in which the exchange-correlation energy depends not only on the local value of the density but also on how it varies in space. By including the gradient $\nabla\rho$, GGA functionals are better equipped to handle systems with inhomogeneous electron distributions, such as molecules and surfaces. This added flexibility allows GGAs to significantly improve upon LSDAs in terms of bond lengths, atomization energies, and reaction barriers. Nevertheless, they are still prone to artificial charge transfer effects. Notable examples include BLYP, which combines Becke's exchange functional[77] with the Lee-Yang-Parr (LYP) correlation functional,[78] and PBE (Perdew-Burke-Ernzerhof),[79] which is widely used due to its physically motivated, nonempirical design and good performance across a broad range of systems. Still, errors such as SIE and artificial charge transfer remain.

### meta-GGA

The third rung of the *Jacob's Ladder* is occupied by the meta-GGAs. These functionals extend the GGA framework by incorporating additional ingredients, such as the kinetic energy density $\tau(\mathbf{r})$ or the second derivative of the density with respect to the coordinates $\nabla^2\rho$, which enables a more nuanced description of electronic structure. This added flexibility helps meta-GGA DFAs better distinguish between different bonding situations and improve predictions of thermochemical and structural properties. Many meta-GGAs are designed to fulfill known exact conditions of the exchange-correlation energy, enhancing their reliability and applicability across diverse chemical systems. A prominent and widely adopted example is the r$^2$SCAN functional,[80] which adheres to nearly all theoretical constraints but does not include the fourth-order exchange gradient expansion (GE4X) to improve numerical stability and convergence. Overall, r$^2$SCAN delivers highly robust results and serves as the foundation for the r$^2$SCAN-3c(STO) method discussed in Chapter 3. While meta-GGAs represent an advancement over GGAs, they are still based on semilocal quantities and are therefore limited in their ability to capture long-range exchange interactions. In addition, they continue to suffer from self-interaction errors. These limitations provide a strong motivation to advance to the next rung of *Jacob's Ladder*.

### Hybrid DFAs

The fourth rung of the *Jacob's Ladder* is occupied by hybrid functionals. These functionals combine elements of both DFT and HF by mixing semilocal exchange, typically from a GGA or meta-GGA, with a portion of nonlocal Fock exchange. Importantly, the exact exchange in hybrid functionals is evaluated using orbitals obtained from the DFT calculation itself. Therefore, it is not identical to the exchange computed within pure HF, which uses HF orbitals throughout. This mixing introduces a nonlocal component into the functional, which reduces the SIE and helps to overcome some of the limitations of semilocal approximations, particularly for reaction barriers, charge-transfer excitations, and systems with delocalized electrons. Among the most widely used hybrid functionals are B3LYP with 20% exact exchange,[81,82] and PBE0 with 25% exact exchange.[83]

A further refinement is provided by range-separated hybrid functionals, which partition the exchange into short-range and long-range components that are combined *via* a smooth transition function. Exact exchange is applied in one of these regions, typically in the long-range part, while semilocal DFT exchange is used in the remaining part. This approach ensures a correct asymptotic behavior of the exchange potential. Notable examples are the functionals $\omega$B97X-V[84] and $\omega$B97M-V,[85] developed by Head-Gordon and co-workers, where the separation is governed by a range separation parameter that can be adjusted to improve performance for specific systems or properties. Instead of only using a short-range and long-range part, so-called local hybrids include a local mixing function that directs the exact exchange admixture. With the added flexibility, they are supposed to improve results, especially electronic excitation energies. A review on local hybrids can be found in Ref. [86].

While hybrid functionals improve upon semilocal approximations by including a portion of exact exchange, they still rely entirely on density-based models for electron correlation. To further enhance accuracy, especially for systems where dynamic correlation is significant, the fifth and final rung of the *Jacob's Ladder* introduces double hybrid functionals.

### Double Hybrid DFAs

Double hybrids (DHs) combine elements of DFT with wave function-based exchange and correlation methods. In addition to mixing semilocal and exact exchange, they incorporate a nonlocal correlation term, often derived from second-order perturbation theory (PT2), typically in the form of MP2. This additional term allows double hybrids to account more accurately for dynamic electron correlation effects, which are often poorly described by conventional functionals. The mixing ratios for exchange and correlation are usually parameterized to balance accuracy across a wide range of chemical properties. Well-known examples include B2PLYP,[87] the first ever double hybrid, as well as more recent developments such as revDSD-PBEP86[88] and Pr$^2$SCAN69,[89] which apply spin-component scaling and more advanced parameterization approaches. Although double hybrids offer high accuracy, particularly for thermochemistry, reaction energies, and noncovalent interactions, their computational cost is significantly higher than that of standard hybrid functionals. This makes them less suitable for larger systems, yet they remain a powerful tool when high accuracy is required.

### London Dispersion

DFT and HF inherently do not capture long-range London dispersion effects, which describe the attractive component of van der Waals interactions. These interactions are a part of dynamic electron correlation and lead to a long-range attraction. Since London dispersion affects most molecular systems, especially those involving inter- and intramolecular non-covalent interactions (NCIs), its absence can significantly reduce the accuracy of DFT calculations. A widely used approach to account for London dispersion is the addition of semiempirical pairwise corrections, such as DFT-D3(BJ)[90–92] and its successor DFT-D4.[93,94]

The main difference between the two models lies in the fact that D4 introduces an explicit charge dependency and includes a three-body dispersion term by default. The D4 dispersion energy is given by:

$$\begin{aligned} E_{\mathrm{disp}}^{D4} = &- \sum_{AB} \sum_{n=6,8} s_n \frac{C_n^{AB}}{R_{AB}^n} f_{\mathrm{damp}}^n (R_{AB}) \\ &- \sum_{ABC} s_9 \frac{C_9^{ABC} (3 \cos \theta_A \; cos\, \theta_B \; cos\, \theta_C + 1)}{(R_{AB} R_{AC} R_{BC})^3} \, f_{\mathrm{damp}}^{(9)} \left( \overline{R}_{\mathrm{ABC}} \right) \end{aligned} \tag{2.22}$$

where the sums run over atom pairs $AB$ and atom triples $ABC$. The $C_n$ dispersion coefficients encode environment-dependent pair-wise dispersion contributions and depend on atomic coordination numbers. For the D4 model, they also depend on the partial charges. The interatomic distances are denoted as $R_{AB}$, while $R_{ABC}$ represents the geometric mean of all pairwise distances in the atom triangle. The internal angles of this triangle are represented by $\theta_A, \theta_B, \theta_C$. The Becke-Johnson (BJ) damping function $f^{(n)}_{\text{damp}}(R_{AB})$ controls the short-range form of the dispersion correction, preventing near-singularities at small atom distances and mitigating double-counting of electron correlation at intermediate separations:[91]

$$f^{(n)}_{\text{damp}}(R_{AB}) = \frac{R^n_{AB}}{R^n_{AB} + (a_1 R^{AB}_0 + a_2)^n}, \tag{2.23}$$

with the cutoff radius $R^{AB}_0 = \sqrt{C^{AB}_8 / C^{AB}_6}$. The full model contains five fitting parameters ($s_6, s_8, s_9, a_1$, and $a_2$) which are specific to each DFA and are typically optimized against the NCIBLIND10,[95] S22x5,[96] and S66x8[97] data sets that contain high-level reference data for interaction energies of relatively small systems with NCIs.

An alternative route is the use of nonlocal correlation functionals such as VV10,[98] which adds a fully nonlocal density-dependent term that captures dispersion interactions more rigorously, without relying on atomic reference data. Therefore, it is sometimes also referred to as DFT-NL. This correction is particularly popular in so-called "dispersion-inclusive" functionals, such as $\omega$B97X-V or $\omega$B97M-V, where the nonlocal term is included in the functional design. While these approaches differ in philosophy and implementation, both have proven essential for enhancing the applicability of DFT and achieving a more physically meaningful description. It is therefore advisable to apply one of them in all practical applications. For more information on dispersion-corrected mean-field electronic structure methods, see Ref. [92].

### 2.2.4 Basis Sets

The electronic structure methods discussed so far require an explicit representation of the wave function and the electron density. However, both are generally unknown in analytical form for systems beyond the hydrogen atom. To make the problem computationally tractable, molecular orbitals are approximated as linear combinations of predefined basis functions ($\chi_\alpha$), which are typically centered on the atomic nuclei and referred to as atomic orbitals. This approach is known as the linear combination of atomic orbitals (LCAO) and is expressed as:

$$\phi_i(\mathbf{r}) = \sum_{\alpha}^{N_{\text{basis}}} c_{\alpha i} \chi_\alpha(\mathbf{r}). \tag{2.24}$$

Here, $\phi_i(\mathbf{r})$ denotes the $i$-th molecular orbital, which is expressed as a linear combination of the $N_{\text{basis}}$ basis functions $\chi_\alpha(\mathbf{r})$, with $c_{\alpha i}$ being the corresponding expansion coefficients. To obtain the total energy of a system, QC programs optimize only the expansion coefficients during the SCF procedure.

The choice of basis functions is crucial for the accuracy and efficiency of QC calculations. There are two main types of basis functions: Slater-type orbitals (STOs) and Gaussian-type orbitals (GTOs). In spherical coordinates, their general mathematical forms are:

$$\text{STO: } \chi_{\text{STO}}(r, \theta, \varphi) = N Y_l^m(\theta, \varphi) r^{n-1} e^{-\zeta r}, \tag{2.25}$$

$$\text{GTO: } \chi_{\text{GTO}}(r, \theta, \varphi) = N Y_l^m(\theta, \varphi) r^{2n-l-2} e^{-\alpha r^2}. \tag{2.26}$$

Here, $N$ is a normalization constant, $Y_l^m(\theta, \phi)$ are spherical harmonic functions, and $n$, $l$, and $m$ are the principal and angular quantum numbers. The parameters $\zeta$ and $\alpha$ define the radial decay of the function. STOs closely reproduce the correct behavior of hydrogen-like orbitals near the nucleus and at large distances. In particular, they satisfy Kato's cusp condition,[99] which ensures the correct sharp change in the slope of the wave function at the nucleus. STOs also do do not fall off as rapidly as GTOs. Nevertheless, STOs lead to complicated integrals that are computationally demanding to evaluate.

In contrast, GTOs allow for much simpler and faster evaluation of multi-center integrals due to the Gaussian product theorem, which states that the product of two Gaussian functions is another Gaussian. They are often employed in the form of contracted Gaussian-type orbitals (CGTOs), which are fixed linear combinations of several primitive Gaussian-type orbitals (PGTOs), each with its own exponent and coefficient:

$$\chi_{\text{CGTO}} = \sum_i^k a_i \chi_{\text{PGTO},i}. \tag{2.27}$$

These primitives are combined into a single, more flexible function that better approximates the shape of a Slater-type orbital. This approach captures the essential physical features of atomic orbitals while retaining the computational efficiency of Gaussians. As a result, contracted Gaussians are widely used in practice, especially since modern GTO basis sets often match or even exceed the accuracy of STO-basis sets, all while offering reduced computational cost.

To systematically improve the accuracy of a basis set, multiple CGTOs can be used to represent each valence atomic orbital. This leads to the concept of $\zeta$-quality basis sets, where the term "$\zeta$" hints at the number of basis functions included in a basis set. A *minimal basis set* is defined as the smallest set of functions sufficient to describe the electronic structure of a neutral atom. Increasing this number leads to double-, triple-, or quadruple-$\zeta$ basis sets, where each atomic orbital is described by two, three, or four basis functions, respectively. While this enhances the accuracy of the calculation, it also significantly increases the computational cost, so an appropriate balance between accuracy and efficiency must be considered. Typically, this approach is applied only to the valence shell, as the core electrons contribute relatively little to chemical bonding. While this principle can be extended to even higher levels, the majority of practical applications utilize basis sets of up to quadruple-$\zeta$ quality. Additional flexibility can be introduced through polarization functions, which are additional higher angular momentum functions, and diffuse functions, which have small exponents and extend farther from the nucleus. The latter are crucial for accurately describing loosely bound electrons, anions, and long-range interactions. When such diffuse functions are added to a basis set, it is sometimes also referred to as an *augmented* basis set.

Widely used GTO basis set families include Ahlrichs' def2 series,[100] such as def2-QZVPPD,[101] which is commonly employed in DFT calculations, and Dunning's correlation-consistent basis sets,[102–104] such as aug-cc-pVQZ, which are frequently used in WFT. Commonly used STO basis sets include the ones provided by the Amsterdam Modeling Suite (AMS),[105–109] for example, QZ4P. The naming conventions of these basis sets typically encode their main features: QZ indicates quadruple-$\zeta$ quality, P denotes the inclusion of polarization functions, and D or *aug* indicate the presence of diffuse functions.

### Basis Set Errors

Each finite basis set includes specific sources of error that can influence the accuracy of quantum chemical calculations. The two most important errors are the basis set incompleteness error (BSIE) and the basis set superposition error (BSSE).

The BSIE describes the deviation between the energy obtained at the complete basis set (CBS) limit, which corresponds to an infinite number of basis functions, and the result obtained with a finite basis set. As a consequence, absolute energies are systematically overestimated when finite basis sets are employed. It is essential to select a basis set that is sufficiently large to ensure convergence of the calculated properties with respect to basis set size. It should be noted that different levels of electronic structure theory exhibit distinct convergence behaviors: HF and DFT energies converge exponentially toward the CBS limit, whereas correlation methods show a much slower, cubic convergence.

In contrast, the BSSE originates from the artificial sharing of basis functions between different fragments of a molecule or complex, which can lead to an artificial lowering of the computed energy relative to the fragments. This effect is particularly pronounced in the calculation of interaction energies, since each fragment effectively "borrows" basis functions from its partner, resulting in an artificially enlarged basis set in the complex compared to the isolated fragments. Consequently, interaction energies are systematically overestimated. The magnitude of the BSSE diminishes with increasing basis set size, as larger basis sets approach saturation and more closely approximate the CBS limit. For addressing the BSSE, the most commonly used approach is the Boys-Bernardi counter-poise (CP) correction,[110] which compensates for intermolecular BSSE by calculating the energy of each fragment using the full basis set of the entire complex. This provides an estimate of the artificial stabilization introduced by basis set sharing. However, it frequently leads to an overcorrection and necessitates additional calculations, which may significantly increase the computational cost. The standard CP correction also only accounts for the intermolecular but not for the intramolecular BSSE, which can still affect geometries and relative energies within a single molecule. To overcome this limitation, the geometrical counter-poise (gCP) correction was developed by Kruse and Grimme in 2012.[111] As an additive and purely geometry-based correction to the electronic energy, it is designed to estimate and compensate for BSSEs but also incompleteness effects without requiring additional quantum chemical calculations. The corresponding correction term is given by:

$$E_{\mathrm{gCP}} = \sigma \cdot \sum_{b \neq a}^{\mathrm{atom}} e_a^{\mathrm{miss}} \cdot f_{\mathrm{dec}}(R_{ab}), \tag{2.28}$$

where $\sigma$ is a global scaling factor, $e_a^{\mathrm{miss}}$ the atomic energy difference between a very large reference basis set and the chosen target basis set (serving as a measure of the basis set incompleteness error), and $f_{dec}(R_{ab})$ a decay function depending on the inter-atomic distances $R_{ab}$. Overall, the gCP correction offers a highly cost-efficient approach to account for both inter- and intramolecular BSSE, making it well-suited for calculations with small to medium-sized basis sets.

### Effective Core Potentials

Since core electrons typically do not participate directly in chemical bonding and their orbitals remain relatively unchanged across different environments, they can be replaced by so-called effective core potentials (ECPs), sometimes also referred to as pseudopotentials (PPs). These potentials effectively replace the explicit treatment of tightly bound core electrons with a parametrized potential that approximates their effect on the valence electrons. This approximation introduces minimal loss of accuracy while significantly reducing the computational cost due to the lower amount of electrons.

In practice, ECPs are constructed by fitting to all-electron calculations, often incorporating relativistic effects, which become increasingly important for heavier elements. With that, they indirectly include a part of relativistic effects without any additional cost (*vide infra*). Commonly used ECP libraries include the def2-ECPs, which are typically paired with Ahlrichs' def2 basis sets, as well as the Stuttgart-Cologne ECPs (e.g., ECP10MDF)[112] developed by Dolg and co-workers, which are often used in combination with Dunning-type basis sets.

### 2.2.5 Relativistic Effects

Up to this point in the thesis, all approximations have been based on the non-relativistic Schrödinger equation, with relativistic effects included only implicitly through ECPs. For most applications, this approach offers an excellent compromise between accuracy and computational efficiency. However, when heavier elements are involved, typically starting from the fourth period of the periodic table, relativistic effects can become significant and may even dictate key physical and chemical properties. Well-known examples include the inert pair effect in thallium compounds, the liquid state of mercury, and the color of gold, all of which have been attributed to relativistic origins.[113] In heavy atoms, inner-shell electrons experience strong Coulomb attraction to the nucleus, which leads to velocities that can approach the speed of light. This results in an increase of the relativistic electron mass and a contraction of $s$-orbitals, while $d$- and $f$-orbitals expand and become more diffuse. These changes in orbital shapes influence the entire electronic structure and can affect bonding, spectroscopy, and reactivity. Therefore, relativity can heavily influence properties that are sensitive to regions close to the nucleus, such as Nuclear Magnetic Resonance (NMR). For these cases, an explicit treatment of relativistic effects becomes essential. A rigorous starting point for such a treatment is the time-independent Dirac equation,

$$\left(c\,\boldsymbol{\alpha}\cdot\mathbf{p}+\boldsymbol{\beta}mc^2+\mathbf{V}\right)\Psi = E\Psi, \tag{2.29}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the $4\times4$ Dirac parameter matrices, $\mathbf{p} = -i\nabla$ is the relativistic momentum operator, $m$ the relativistic electron mass, $c$ the speed of light, and $\mathbf{V}$ an (external) electrostatic potential. The four-component spinor $\Psi$ contains a large and a small component, with the latter contributing only marginally to the total wave function. This observation motivates two-component approximations that decouple the small component, thereby reducing computational cost while retaining the dominant relativistic contributions. Several approaches implement this principle in QC. The Douglas–Kroll–Hess (DKH)[114–117] transformation achieves decoupling through a sequence of unitary transformations, with accuracy improving systematically as the order of the expansion increases. The Zeroth Order Regular Approximation (ZORA)[118–120] offers a computationally efficient alternative, albeit without the same systematic improvability. The exact two-component (X2C) method[121–126] performs a mathematically exact one-step decoupling of the Dirac Hamiltonian, eliminating the truncation errors inherent to finite-order DKH while maintaining high efficiency. Depending on the application, it is possible to include only scalar-relativistic (SR) effects, which adds only a moderate computational cost, or to include both SR and spin-orbit (SO) effects, which are more demanding but essential for certain spectroscopic and magnetic properties, as discussed in Chapter 4.

### 2.2.6 "3c" Composite Methods

Although high accuracy is always desirable, it is inevitably accompanied by substantial computational cost. The "3c" approach was developed as a strategy to balance these competing demands by combining an electronic structure method with a relatively small basis set, while compensating for inherent deficiencies through three targeted corrections. The concept was first realized in HF-3c,[127] which applies HF theory with the MINIX minimal basis set. To correct for its known shortcomings, three additional terms are included: the D3(BJ) dispersion correction, the gCP correction to mitigate BSSE, and the short-range bond (SRB) correction, which primarily refines covalent bond lengths. HF-3c was designed to yield reliable geometries and noncovalent interaction energies, making it a versatile tool for large-scale applications where speed is critical. Building on the success of HF-3c, subsequent developments in the 3c family shifted to DFT as the underlying electronic structure method. The first of these, termed PBEh-3c,[128] employs the PBE-based[79] hybrid functional with 42% Fock exchange, combined with the double-$\zeta$ mSV(P) basis set, and retains the D3(BJ) and gCP corrections. This configuration generally offered an improved accuracy over HF-3c. Further refinement was introduced with B97-3c,[129] which incorporated an adapted triple-$\zeta$ mTZVP basis set, along with the D3(BJ) and SRB corrections. This provided a better balance between basis set flexibility and computational efficiency, leading to improved performance for both geometries and energetics. The more recent development r$^2$SCAN-3c[43] is based on the r$^2$SCAN[80] meta-GGA functional in combination with the mTZVPP basis set, the D4 dispersion correction, and the gCP correction. This method is likely one of the most commonly applied 3c methods, delivering an exceptional accuracy-to-cost ratio across diverse chemical applications, with particularly strong performance for conformational energies. Motivated by the fact that STOs describe atomic orbitals more reasonably than GTOs (cf. Section 2.2.4), Chapter 3 introduces r$^2$SCAN-3c(STO) as an alternative to the standard r$^2$SCAN-3c, employing an STO basis set in place of the usual GTO basis. Further 3c methods include HSE-3c,[130] B3LYP-3c,[131] and $\omega$B97X-3c,[132] which all follow the same design principle.

### 2.2.7 Semiempirical Quantum Mechanical (SQM) Methods and Machine Learning Interatomic Potentials (MLIPs)

Electronic structure calculations based on high-level quantum chemical methods can be prohibitively expensive for large systems or extensive sampling of the potential energy surface. In such cases, semiempirical quantum mechanical (SQM) methods offer a practical alternative by approximating the electronic Hamiltonian and introducing empirical parameters fitted to reference data. This reduces computational cost while retaining a physically motivated foundation. A widely used collection of SQM methods is the geometry, frequency, noncovalent interactions (GFN) family,[34,133] developed within the extended tight-binding (xTB) formalism. The first generation, GFN1-xTB,[134] offers a broadly applicable parametrization for all elements up to radon. GFN2-xTB[135] improves upon this by introducing more sophisticated multipole electrostatics and replacing the D3 London dispersion correction with a self-consistent D4 scheme. These enhancements significantly increase the method's accuracy, particularly for intermolecular interactions and transition-metal complexes. In addition to these electronic structure-based models, GFN-FF[54] applies the same parametrization philosophy to a purely classical force field, enabling simulations of very large systems where an explicit quantum mechanical treatment would be impractical.

An alternative approach to achieving lower computational costs for large systems is the use of machine learning interatomic potentials (MLIPs). They replace explicit electronic calculations with surrogate models trained on high-quality reference data. AIMNet2[136] is a recent atom-in-molecule neural network that combines message passing with rich atomic descriptors to achieve transferable predictions for energies and forces. Another example is the small Universal Model for Atoms (UMA-sm)[137], now also known as UMA-s-1, which aims to provide a single, general model across elements and bonding types by learning shared atomic representations and many-body interactions. While MLIPs can achieve near *ab initio* accuracy at a fraction of the cost, their performance depends critically on the training data. Poor coverage of relevant bonding motifs, charge states, or reaction pathways can lead to unphysical results, and long-range effects are often underestimated unless modeled explicitly. Both SQM methods and MLIPs are valuable tools in everyday computational workflows. Although more approximate than high-level quantum chemical approaches, they can provide rapid estimates of structures, energetics, and other properties, making them especially useful in the preliminary stages of research. Their speed enables extensive screening and exploration of chemical space before committing resources to more accurate, but significantly more expensive, calculations.

## 2.3 General Workflow

Depending on the property of interest, differently designed computational workflows must be applied. The process typically begins with obtaining an initial molecular geometry, for example, from crystallographic data or manual molecular modeling, which is then optimized to ensure a stable structure. For this purpose, optimization algorithms such as steepest descent can be employed to locate ground-state structures. These methods operate by evaluating the gradient of the potential energy surface and iteratively following it toward a local energy minimum. In contrast, locating transition states is considerably more challenging and requires specialized algorithms such as the Growing String Method (GSM)[138] or the Nudged Elastic Band for Transition States (NEB-TS).[139,140] For flexible molecules, it is often necessary to explore the conformational space to account for multiple relevant structures. This can be achieved with automated conformer sampling tools such as the global optimizer algorithm (GOAT)[141] or the conformer-rotamer ensemble sampling tool (CREST).[142,143] Since these approaches are computationally demanding, they are typically combined with SQM methods. The resulting conformer ensembles are then refined at higher levels of theory, for example DFT. This can be done automatically with tools, such as the commandline energetic sorting (CENSO) program.[144] Employing such a hierarchical sequence of methods with different levels of theory is commonly referred to as a *multi-level approach*. Once reliable geometries are available, energies and other molecular properties can be computed. When working with conformational ensembles, an important decision is whether the property of interest should be averaged over the ensemble, for instance using Boltzmann weighting (*vide infra*), or whether considering only the lowest-energy conformer is sufficient. The subsequent sections will provide a brief outline of how different types of properties are addressed within this general framework.

## 2.4 Energy Quantities and Analysis Methods

In computational chemistry, energies play a central role in describing molecular structure, stability, and reactivity. Electronic structure methods provide access to the total electronic energy, which can then be extended with enthalpic, entropic, and solvation contributions to yield thermodynamic properties such as the Gibbs free energy. Beyond these total values, energy decomposition schemes allow for the partitioning of interaction energies into chemically interpretable contributions. This section summarizes the main energy quantities and analysis techniques employed in this work.

### Gibbs Free Energy

Electronic structure calculations provide the electronic energy of a molecule, which corresponds to the expectation value of the electronic Hamiltonian at a given nuclear geometry. However, this quantity alone is not sufficient for direct comparison with experimental thermochemistry. Reaction barrier heights or reaction energies are usually reported as Gibbs free energies ($G$) that, apart from the electronic energy ($E_{\mathrm{el}}$), also include thermostatistical contributions ($G_{\mathrm{TRV}}$) and solvation contributions ($\delta G_{\mathrm{solv}}$) at a specific temperature ($T$):

$$G_{\mathrm{tot}} = E_{\mathrm{el}} + G^{T}_{\mathrm{TRV}} + \Delta G^{T}_{\mathrm{solv}}. \tag{2.30}$$

Reaction energies and barrier heights are then obtained from the differences in Gibbs free energy between products and reactants, or between transition states and reactants, respectively:

$$\Delta G = G_{\mathrm{products/TS}} - G_{\mathrm{reactants}}. \tag{2.31}$$

This formulation allows for a direct comparison with experimentally measured thermodynamic and kinetic data. It should be noted, however, that the accuracy of the computed values strongly depends on the chosen level of theory and the solvation model employed.

### Thermostatistical Contributions

In general, the Gibbs free energy is related to the enthalpy $H$, temperature $T$ and entropy $S$ by:

$$G = H - TS. \tag{2.32}$$

Both enthalpic and entropic contributions can be derived from statistical mechanics. The central quantity in statistical mechanics is the partition function ($q$), which connects the microscopic properties of individual states to macroscopic thermodynamic observables. It is defined as the sum over all possible states $s$ with corresponding energies $\epsilon_s$:

$$q = \sum_{s} e^{-\epsilon_s/k_{\mathrm{B}}T}, \tag{2.33}$$

where $k_{\mathrm{B}}$ is the Boltzmann constant. Within the rigid-rotor harmonic oscillator (RRHO) approximation, the partition function can be evaluated analytically and factorized into independent translational, rotational, vibrational, and electronic contributions:

$$q_{\mathrm{tot}} = q_{\mathrm{trans}}\, q_{\mathrm{rot}}\, q_{\mathrm{vib}}\, q_{\mathrm{elec}}. \tag{2.34}$$

Accordingly, the total enthalpy and entropy can be expressed as sums over these individual contributions:

$$H_{\mathrm{tot}} = H_{\mathrm{trans}} + H_{\mathrm{rot}} + H_{\mathrm{vib}} + H_{\mathrm{elec}}, \tag{2.35}$$

$$S_{\text{tot}} = S_{\text{trans}} + S_{\text{rot}} + S_{\text{vib}} + S_{\text{elec}}. \tag{2.36}$$

In this framework, translational motion is described by an ideal gas model, rotational motion by the rigid-rotor approximation, and vibrational motion by the harmonic oscillator model. Since the electronic energy is evaluated separately from the thermostatistical contributions $G^T_{\text{TRV}}$, the latter comprise only translational (T), rotational (R), and vibrational (V) contributions, together with the zero-point vibrational energy $E_{\text{ZPVE}}$.

However, the harmonic oscillator model can become unreliable for low-frequency vibrational modes (typically below 100 $\text{cm}^{-1}$), which can lead to unphysically too large entropic contributions. To mitigate this, such modes are better described as a mixture of harmonic vibrations and free internal rotations. This refinement is referred to as the modified RRHO (mRRHO) approach,[145,146] where "m" denotes the modification. By smoothly interpolating between vibrational and rotational treatments of low-frequency modes, the mRRHO scheme yields more realistic entropy and free energy values. In this work, all thermostatistical contributions were consistently computed using the mRRHO approach, ensuring that $E_{\text{ZPVE}}$, enthalpic, and entropic effects were accounted for.

### Solvation Free Energy

When reactions take place in solution, it is essential to account for solvation effects, as they can influence both the molecular geometry and the relative energies. Two general approaches exist to include solvation: explicit and implicit models. In explicit solvation, solvent molecules are placed around the solute and treated directly in the calculation. While this provides a detailed description of solute-solvent interactions, it is computationally very demanding. Consequently, implicit solvation models are more commonly applied. These models describe the solvent as a continuum surrounding the solute. For the solvation contribution, the interaction between the solute and the reaction field of the continuum is then evaluated. For SQM methods, such as the GFN family, commonly-used implicit models include the generalized Born and surface area (GBSA)[147–149] and analytical linearized Poisson-Boltzmann (ALPB)[150] approaches. For DFT-based approaches, popular solvation models are the conductor-like polarizable continuum model (CPCM),[151] the conductor-like screening model (COSMO),[152,153] COSMO for real systems (COSMO-RS),[154,155] and the solvation model based on density (SMD).[156] The solvation contribution to the free energy is computed as the difference between the total free energy in solution ($G_{\text{solv}}$) and in the gas phase ($E_{\text{gas}}$):

$$\Delta G^T_{\text{solv}} = G_{\text{solv}} - E_{\text{gas}}. \tag{2.37}$$

$E_{\text{gas}}$ is just the electronic energy, while the $G_{\text{solv}}$ includes the electronic energy as well as the solvation contribution. Since the solvation process involves a change of standard state going from the gas phase at 1 atm to 1 $\text{molL}^{-1}$ in solution, an additional concentration-induced free-energy shift has to be considered:

$$RT \ln V_{\text{M}} = 1.89 \text{ kcal mol}^{-1}, \tag{2.38}$$

with $R$ being the ideal gas constant, $T$ the temperature, here $T = 298$ K, and $V_{\text{m}}$ the molar volume of an ideal gas.

### Boltzmann Weighting

Since molecules typically exist in multiple conformations rather than a single one, many properties are influenced by the contributions of all relevant conformers. To quantify these contributions, their relative populations can be estimated using Boltzmann weights ($w$):

$$w_i = \frac{\exp\left(-G_i/RT\right)}{\sum_j^{N_\text{conf}} \exp\left(-G_j/RT\right)}, \tag{2.39}$$

where $G_i$ denotes the Gibbs free energy of conformer $i$, $R$ is the ideal gas constant, and $T$ the temperature. Boltzmann-averaged properties are then obtained by applying the corresponding Boltzmann weights to the computed values of each conformer, followed by summation over the entire ensemble:

$$\overline{X} = \sum_i^{N_\text{conf}} w_i X_i, \tag{2.40}$$

where $\overline{X}$ is the Boltzmann weighted property, $N_\text{conf}$ the number of conformers, and $X_i$ the respective property of conformer $i$. This procedure is especially important for predicting properties of flexible molecules, as most properties heavily depend on the underlying geometry. Boltzmann weighting has many applications and was also employed in this work to improve the accuracy of computed NMR chemical shifts, as discussed in Chapter 4.

### Interaction and Association Free Energy

After outlining how molecular energies arise from electronic, thermal, and solvation contributions, together with conformational effects, the focus can now shift toward the energetic interactions between distinct molecular entities. The *interaction energy* ($E_\text{int}$) quantifies the energetic stabilization (or destabilization) that occurs upon forming a complex from its isolated fragments:

$$E_\text{int} = E_{AB} - E_A - E_B. \tag{2.41}$$

Here, $E_{AB}$ denotes the total energy of the complex, while $E_A$ and $E_B$ correspond to the total energies of the individual fragments constrained to the geometries they adopt within the complex. As such, $E_\text{int}$ reflects the energy change associated with bringing the fragments together without allowing any structural relaxation. This quantity is therefore sometimes also referred to as the *vertical interaction energy*, analogous to a vertical excitation in electronic structure theory, where geometries are held fixed. In contrast, if the geometries of the fragments are relaxed from their optimized isolated structures to those they adopt in the complex, the resulting energy difference also includes these deformation or relaxation contributions. When such geometric relaxation is included, the energy difference is sometimes referred to as the *adiabatic interaction energy*, which proves a more realistic description of the binding process. Beyond purely electronic contributions, additional effects such as thermal, entropic, and solvation terms can be incorporated to yield the *association free energy* ($\Delta G_\text{assoc}$), which represents the overall thermodynamic driving force for complex formation:

$$\Delta G_\text{assoc} = G_{AB} - G_A - G_B. \tag{2.42}$$

### Energy Decomposition Analysis (EDA) and Local Energy Decomposition (LED)

The interaction energy of a molecular complex can be partitioned into physically interpretable contributions, providing deeper insight into the nature of intermolecular interactions. For example, in the energy decomposition analysis (EDA) scheme[157] as implemented in the TURBOMOLE program package,[158] the total interaction energy is separated into terms such as Pauli repulsion ($\Delta E_{\text{rep}}$), electrostatics ($\Delta E_{\text{elstat}}$), exchange ($\Delta E_{\text{exch}}$), orbital relaxation ($\Delta E_{\text{orb-relax}}$), correlation ($\Delta E_{\text{correl}}$), and London dispersion ($\Delta E_{\text{disp}}$):

$$E_{\text{int,EDA}} = \Delta E_{\text{rep}} + \Delta E_{\text{elstat}} + \Delta E_{\text{exch}} + \Delta E_{\text{orb-relax}} + \Delta E_{\text{correl}} + \Delta E_{\text{disp}}. \tag{2.43}$$

Each of these captures a distinct physical effect: electrostatics reflects the classical Coulomb interaction between fragments, Pauli repulsion arises from the antisymmetry of the wave function, orbital relaxation accounts for a part of polarization and charge transfer, and correlation and dispersion describe electron correlation beyond the mean-field theory.

An alternative is the local energy decomposition (LED) approach,[159,160] which provides access to a similar decomposition but within the DLPNO-CCSD(T) framework:

$$E_{\text{int,LED}} = \Delta E_{\text{HF-elprep}} + \Delta E_{\text{elstat}} + \Delta E_{\text{exch}} + \Delta E_{\text{orb-relax}} + \Delta E_{\text{disp}} + \Delta E_{\text{non-disp}}. \tag{2.44}$$

It includes the electronic preparation energy at the HF level ($\Delta E_{\text{HF-elprep}}$) that describes the repulsive part of the exchange interaction and therefore represents Pauli repulsion, as well as the electrostatic interaction ($\Delta E_{\text{elstat}}$), the attractive exchange energy ($\Delta E_{\text{exch}}$), orbital-relaxation ($\Delta E_{\text{orb-relax}}$), London dispersion ($\Delta E_{\text{disp}}$), and non-dispersive effects ($\Delta E_{\text{non-disp}}$). Here, the $\Delta E_{\text{non-disp}}$ term represents a correction for deficiencies in the description of permanent electrostatic interactions within the HF method. In this approach, the locality of the DLPNOs enables these contributions to be assigned to specific orbital domains or fragment pairs. Overall, the EDA emphasizes the physical interpretation of interactions at lower levels of theory, while LED extends these concepts to the coupled-cluster domain, thereby offering both accuracy and interpretability for the study of interactions.

## 2.5 Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR spectroscopy is a powerful tool for elucidating molecular structures and probing dynamics. The most important concepts within NMR are discussed in this Chapter, which is mainly based on Refs. [161, 162]. NMR relies on the interaction of nuclear spins with an external magnetic field and the influence of surrounding electrons. In a typical NMR experiment, a sample is placed in a strong static magnetic field and irradiated with radiofrequency pulses, which perturb the nuclear spin populations. Nuclei with a non-zero spin quantum number absorb this energy and subsequently relax back to equilibrium, re-emitting energy at resonance frequencies determined by their local magnetic environment, thereby producing a characteristic spectrum. Figure 2.3 shows a calculated $^1$H NMR spectrum of ethanol, highlighting two key observables: the chemical shifts $\delta$ and the $J$ coupling constants. These quantities will be discussed in detail in the following subsections. For complex molecules, analyzing the spectra can be complicated due to overlapping signals, subtle chemical shift differences, and intricate coupling patterns. In the case of heavier nuclei, the shifts are distributed along a large range, which can further hinder peak assignment and may require multiple measurements. Here, QC calculations provide

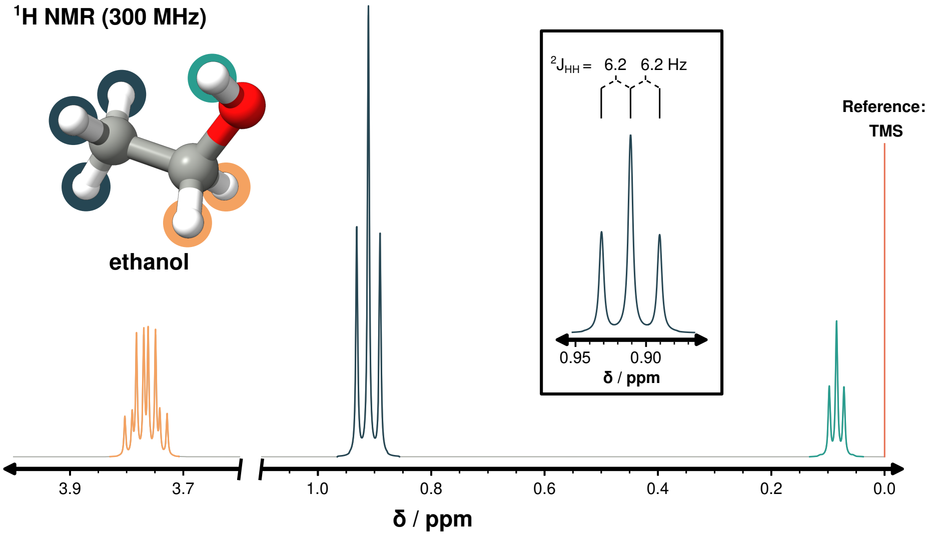


Figure 2.3: NMR spectrum of ethanol computed at the PBE0/def2-TZVP[83,100] level of theory.

valuable support by computing chemical shifts and spin-spin coupling constants. The theoretical foundation for these computations was established in the early 1950s by Norman F. Ramsey,[163,164] who showed that such observables can be expressed as second-order derivatives of the electronic energy that are explained in the following sections.

### Magnetic Shieldings and Shifts

Magnetic shielding describes the influence of the surrounding electrons on the magnetic field experienced by a nucleus in a molecule. When an external magnetic field $\mathbf{B}_0$ is applied, the electronic cloud responds by generating an induced magnetic field, which increases or decreases the local magnetic field at the nucleus $\mathbf{B}_{\text{loc}}$, and with that shields or deshields the nucleus. Only nuclei with a non-zero nuclear spin quantum number $I$ are NMR-active, with $I = \frac{1}{2}$ nuclei (e.g., $^1$H, $^{13}$C) being the most commonly measured due to their sharp, easily interpreted signals. The measurements can be taken in either the solid or solution phase, but the present work primarily focuses on the computation of NMR spectroscopy in solution.

In quantum chemistry, the magnetic shielding is deduced from the magnetic shielding tensor $\boldsymbol{\sigma}$ that relates the local magnetic field at the nucleus to the applied field:

$$\mathbf{B}_{\text{loc}} = \mathbf{B}_0 - \boldsymbol{\sigma} \cdot \mathbf{B}_0. \tag{2.45}$$

The tensor elements are defined as second derivatives of the electronic energy $E$ with respect to the nuclear magnetic moment $\boldsymbol{\mu}$ and the external magnetic field $\mathbf{B}_0$:

$$\sigma_{ij} = \frac{\partial^2 E}{\partial \mu_i \, \partial B_{0,j}}, \tag{2.46}$$

where $i$ and $j$ denote the cartesian components ($x$, $y$, and $z$). The resulting shielding tensor is a $3 \times 3$

matrix:

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}. \tag{2.47}$$

Its diagonal elements correspond to the shielding along each principal axis, while the off-diagonal elements describe cross terms between directions of $\boldsymbol{\mu}$ and $\mathbf{B}_0$. For molecules in solution, anisotropic shielding effects are averaged by rapid molecular motion. Therefore, the experimentally relevant parameter is the isotropic shielding constant, given by:

$$\sigma_{\text{iso}} = \frac{\sigma_{11} + \sigma_{22} + \sigma_{33}}{3}. \tag{2.48}$$

From perturbation theory, the shielding tensor can also be decomposed into diamagnetic ($\sigma^{\text{dia}}$) and paramagnetic terms ($\sigma^{\text{para}}$), which can be used to analyze the shielding in more detail:

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}^{\text{dia}} + \boldsymbol{\sigma}^{\text{para}}. \tag{2.49}$$

The diamagnetic term depends only on the occupied electronic states and the ground-state electron density. It is generally positive and dominates in most closed-shell systems. The paramagnetic term originates from virtual excitations from occupied to unoccupied states induced by the magnetic perturbation. It is often negative and strongly influenced by low-lying excited states. The balance between these contributions determines whether a nucleus is overall shielded or deshielded.

### NMR Chemical Shifts

In NMR experiments, results are reported as chemical shifts rather than absolute magnetic shieldings. These shifts are defined relative to a reference compound, which ensures reproducibility and comparability across measurements. Computed shifts are obtained in the same way by referencing the computed isotropic shielding $\sigma_{A,\text{iso}}$ for nucleus $A$ to the isotropic shielding of the reference compound $\sigma_{\text{ref,iso}}$ computed at the same level of theory:

$$\delta_A = \sigma_{\text{ref,iso}} - \sigma_{A,\text{iso}}. \tag{2.50}$$

Common reference compounds include tetramethylsilane for $^{1}$H, $^{13}$C, and $^{29}$Si, tetramethylstannane for $^{119}$Sn, and tetramethyllead for $^{207}$Pb NMR. Because tetramethyllead is highly toxic, safer substitutes such as lead nitrate are often used. In such cases, the secondary reference scale is calibrated against tetramethyllead through a well-established offset, ensuring that $^{207}$Pb chemical shifts are still reported relative to tetramethyllead.

Computed NMR chemical shifts are typically prone to systematic errors that can be remedied by applying an empirical linear scaling scheme to the computed NMR shifts. In this approach, the calculated shifts are corrected according to

$$\delta_{\text{calc,scaled}} = \frac{\delta_{\text{calc}} - \beta}{\alpha}, \tag{2.51}$$

where $\alpha$ denotes the slope and $\beta$ the intercept of the linear regression performed on the calculated NMR shifts with respect to the experimental reference data.

As discussed in Section 2.2.5, relativistic effects become increasingly important for heavier nuclei, especially for properties that are sensitive to the electrons in the core region, such as NMR spectroscopy. Spin-orbit coupling can give rise to the so-called heavy-atom on heavy-atom (HAHA) and heavy-atom on light-atom (HALA) effects, which alter the magnetic shielding of the nucleii and, consequently, the observed chemical shifts. Accurate prediction of such shifts, therefore, requires an explicit relativistic treatment.

### Spin-spin Couplings

Spin-spin couplings, also known as $J$-couplings, describe through-bond interactions between two nuclear spins mediated by the electronic structure of the molecule. They are responsible for the splitting patterns observed in NMR spectra and are essential for extracting connectivity information. Formally, the spin-spin coupling tensor between nuclei $A$ and $B$ is obtained as the second derivative of the electronic energy with respect to their nuclear magnetic moments:

$$J_{ij} = \frac{\partial^2 E}{\partial \mu_{A,i}\, \partial \mu_{B,j}}. \tag{2.52}$$

The isotropic $J$-coupling constant is then determined in analogy to the isotropic shielding value:

$$J_{\text{iso}} = \frac{J_{11} + J_{22} + J_{33}}{3}. \tag{2.53}$$

### Chemical Equivalence and Averaging

In solution NMR, nuclei that are chemically equivalent due to molecular symmetry or rapid molecular motions exhibit a single, averaged chemical shift. Computationally, however, a fixed-geometry calculation may produce slightly different shielding constants and spin-spin couplings for such nuclei, e.g., due to the specific conformation chosen. To achieve results that are directly comparable to those of the experiment, it is therefore important to average over magnetically equivalent nuclei and, where necessary, sample an ensemble of conformations, e.g., using the CREST program (cf. Section 2.3).[165]

### Electric Field Gradients

In NMR spectroscopy of quadrupolar nuclei with $I > \frac{1}{2}$ (e.g., $^{35}$Cl), the electric field gradient (EFG) at the nucleus influences the spin relaxation processes and therefore has a direct impact on the observed linewidth of the resonance. Large EFGs typically lead to faster relaxation and broader signals, which can severely limit spectral resolution or even render a resonance undetectable. The EFG tensor describes the variation of the electrostatic potential in the vicinity of the nucleus and, after diagonalization, is represented as a $3 \times 3$ traceless matrix:

$$\mathbf{V} = \begin{pmatrix} V_{11} & 0 & 0 \\ 0 & V_{22} & 0 \\ 0 & 0 & V_{33} \end{pmatrix}, \tag{2.54}$$

with the principal components ordered by magnitude such that:

$$|V_{33}| \geq |V_{22}| \geq |V_{11}|, \quad V_{11} + V_{22} + V_{33} = 0. \tag{2.55}$$

The asymmetry parameter $\eta$ quantifies the deviation from axial symmetry and is defined as:[166]

$$\eta = \frac{V_{11} - V_{22}}{V_{33}}; \quad 0 \leq \eta \leq 1. \tag{2.56}$$

For nuclei with a finite electric quadrupole moment $Q$, the interaction between $Q$ and the EFG leads to nuclear quadrupole coupling, which is a dominant relaxation pathway in solution and solids. The NMR linewidth $\Delta\nu$ of a quadrupolar nucleus is proportional to the inverse of the spin-spin relaxation time $T_2$, which depends on $I$, $\eta$, $Q$, the reduced Planck constant $\hbar$, the largest principal component $V_{33}$, and the isotropic correlation time $\tau_c$. Neglecting constants and assuming equal $\tau_c$ for all systems under comparison, the proportionality reduces to:

$$\Delta\nu \propto V_{33}^2 \left(1 + \frac{\eta^2}{3}\right). \tag{2.57}$$

Since $\eta$ is bound between 0 and 1, the factor $\left(1 + \frac{\eta^2}{3}\right)$ varies only between 1 and 4/3, meaning that $V_{33}^2$ is the primary determinant of quadrupolar line broadening. This behavior is briefly discussed in Chapter 8 for the $^{35}$Cl shift in chloride encapsulated silafulleranes.

## 2.6 Benchmark Set Design and Evaluation

With the introduction of approximate quantum-chemical approaches, the central question becomes which methods are most appropriate for addressing a given problem. While certain general trends can be established, the choice is often far from trivial. Post-HF methods, for example, are in most cases more reliable than plain HF, as they explicitly account for electron correlation and are systematically improvable. For DFT, however, the situation is more complex. Although the *Jacob's ladder* of DFAs provides a conceptual framework in which each rung represents an increase in theoretical sophistication, higher rungs do not universally guarantee better accuracy, and in some cases, lower-rung functionals can outperform higher ones. This variability underscores the need for systematic evaluation. To this end, benchmark sets have become an indispensable tool, as they allow for a consistent assessment of accuracy, robustness, and computational efficiency across a wide range of methods. By comparing theoretical predictions against high-quality reference data, such benchmarks provide critical guidance for method selection in both method development and practical applications. Benchmark sets are commonly designed for very specific properties, e.g., conformational energies,[167] barrier heights,[38] NMR shifts,[11,12,39] or solvation contributions.[168] In method design, this approach allows users to pinpoint certain defects in the developed method. A comprehensive, more general overview can be obtained by using benchmark collections such as the GMTKN55[36] or the GSCDB138[37], which combine benchmark sets for different properties. Designing a benchmark set is far from trivial and requires careful consideration of multiple interdependent factors. This work is mainly focused on benchmarking DFT, but the mentioned points can also be extended to other methods.

### Desired Property

The first step in designing any benchmark set is to determine which property should be evaluated. In general, benchmark sets fall into two categories, depending on the source of the reference data: theoretically or experimentally motivated sets. High-level theoretical data, particularly for energetic properties, can serve as a reliable reference for assessing and parameterizing lower-level computational methods. Alternatively, experimental data can be used, for example, for NMR chemical shifts or vibrational frequencies. In these cases, however, measurement uncertainties must be carefully considered. Experimental data can also be employed to benchmark electronic energies, reaction enthalpies, or solvation contributions, although this approach requires careful back-corrections. Typically, experimentally accessible thermodynamic quantities include both enthalpic and entropic contributions. To isolate the property of interest, high-level calculations are often performed to estimate these additional terms, which are then subtracted from the experimental values to obtain the desired reference property.

#### Dataset Size and Diversity

Benchmark sets vary widely, from narrowly focused datasets of a few dozen data points to broad compilations containing thousands. This usually depends on the availability of experimental data or computational resources. In terms of diversity, it depends on what exactly should be benchmarked. It can be favorable to benchmark only one class of compounds if this is a targeted property, but to get a more general view on the performance and especially robustness, it is crucial to ensure that the benchmark covers different types of molecules.

#### Reference Geometries

Since all calculations rely on the underlying geometries, these must be generated with particular care. Experimental observables are usually associated with stationary points on the potential energy surface. Therefore, reference geometries used for comparison with experimental data should correspond to either minima or transition states and should be optimized at a sufficiently reliable level of theory. Typically, hybrid-level DFT is applied with at least a triple-$\zeta$ basis set, though higher-level optimizations may be desirable for sensitive cases. Generally, non-stationary points can also be of interest, since they allow a more complete mapping of the potential energy surface. Including geometries far from equilibrium, such as distorted structures or points along reaction coordinates, provides valuable information on the system's response to bond stretching, angle bending, and torsional motions. This broader sampling ensures that the parametrization is not only accurate in the vicinity of minima and transition states but also robust across chemically relevant distortions, thereby improving the transferability and predictive power of the resulting models.

#### Reference Values

The accuracy of reference data is one of the most important factors. High-level wave function methods, such as CCSD(T) or localized variants like PNO-LCCSD(T), extrapolated to the complete basis set limit, are commonly employed for energy benchmark sets. One underlying problem of post-HF methods, however, is the convergence behavior of the basis set. A promising strategy to accelerate the convergence of coupled cluster methods with respect to the basis set is the use of explicitly correlated approaches, often denoted as F12 methods.[169–171] The key idea is to augment the wave function ansatz with terms that depend explicitly on the interelectronic distance $r_{12}$. Since the slow basis set convergence of conventional wave function methods originates from their inability to efficiently describe the electron-electron cusp, the inclusion of explicit $r_{12}$-dependent functions provides a much more compact and accurate description of the correlation hole. As a consequence, F12 methods can achieve near-complete basis set accuracy with significantly smaller basis sets, thereby reducing both the computational cost and the residual BSIE. This can also be applied in the context of local correlation methods. For example, the PNO-LCCSD(T)-F12[63–68,172,173] approach leverages pair natural orbitals to achieve linear scaling with system size, while the F12 treatment ensures rapid basis set convergence. This synergy enables highly accurate coupled cluster calculations on large molecular systems at a fraction of the cost of canonical CCSD(T). It is therefore frequently employed for the determination of high-accuracy reference energies, as discussed in Chapters 5, 6, and 7. Nevertheless, PNO-CCSD(T)-F12 is not a "black box" method as it often demands careful parameter tuning and auxiliary basis set choices tailored to the specific compound class, requiring substantial expert knowledge for reliable application.

### Benchmarking Study & Statistical Analysis

Every benchmark set that is established requires a benchmarking study to evaluate different methods.[35] For DFT, it is important to select a representative selection of all commonly used density functionals, and, depending on the property, basis set effects can also be evaluated. For electronic energies, a quadruple-$\zeta$ basis set should be applied, especially for double-hybrids, due to their PT2 correlation term. If anions or other more complex structures are present, one needs to consider applying more diffuse basis functions.

To effectively evaluate performance and interpret benchmarking results, it is essential to employ appropriate statistical metrics. The following measures are commonly used for this purpose:

$$\text{error } e_i = x_i - x_i^{\text{ref}}, \tag{2.58}$$

$$\text{Mean signed error (MSE)} = \frac{1}{N}\sum_{i=1}^{N} e_i = \overline{e}, \tag{2.59}$$

$$\text{Mean absolute error (MAE)} = \frac{1}{N}\sum_{i=1}^{N} |e_i|, \tag{2.60}$$

$$\text{Standard deviation (SD)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \left[e_i - \overline{e}\right]^2}, \tag{2.61}$$

$$\text{Root mean square error (RMSE)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \left(e_i\right)^2}, \tag{2.62}$$

$$\text{Absolute maximum error (AMAX)} = \max\left(|e_i|\right), \tag{2.63}$$

with $x_i$ describing the computed value, $x_i^{\mathrm{ref}}$ the reference value and $N$ the dataset size.
The mean signed error (MSE) (also referred to as the mean error, ME) reveals systematic over- or underestimation, while the mean absolute error (MAE) measures the average unsigned error and thus reflects overall accuracy. It should be noted that the MAE is often incorrectly termed the mean absolute deviation (MAD) in the benchmarking literature. This should not be done as the MAD describes the average of absolute deviations from a reference point, usually the mean of all datapoints. A short discussion about this topic can be found in the Supporting Information of Ref. [174]. The standard deviation (SD), defined without Bessel's correction, captures the spread of errors around the mean, providing insight into error consistency and, with that, about the robustness. The root mean square error (RMSE) emphasizes larger deviations more strongly than the MAE due to squaring, making it sensitive to outliers. Finally, the absolute maximum error (AMAX) highlights the worst-case error within the dataset, which is particularly relevant for identifying problematic cases.

If the methods are evaluated based on the performance on different benchmark sets at the same time, it is important to consider the mean absolute value ($|\Delta E|$) of a benchmark set. For example, conformational energies can be quite low with energies up to 10 kcal·mol$^{-1}$, while reaction energies can rise up to over 100 kcal·mol$^{-1}$. To enable a balanced comparison across benchmark sets that differ in their characteristic energy scales, the weighted total mean absolute deviation in version 2 (WTMAD-2) was introduced alongside the GMTKN55 set:[36]

$$\mathrm{WTMAD\text{-}2} = \frac{1}{\sum\limits_i^N N_i} \cdot \sum_i^N N_i \cdot \frac{\overline{|\Delta E|}_{\mathrm{total}}}{\overline{|\Delta E|}_i} \cdot \mathrm{MAE}_i . \tag{2.64}$$

In this expression, $N$ denotes the total number of benchmark subsets, while $N_i$ is the number of data points in subset $i$. The term $\mathrm{MAE}_i$ represents the mean absolute error of the method for subset $i$. The weighting factor is determined by relating the average of all absolute reference energies within the respective subset ($\overline{|\Delta E|}_i$) to the average of all $|\Delta E|$ values across the complete dataset ($\overline{|\Delta E|}_{\mathrm{total}}$). This weight is multiplied by the number of data points in the subset to account for its relative size. Finally, the WTMAD-2 is obtained by summing over all subsets and dividing by the total number of data points across all benchmark sets. This statistic enables a more consistent comparison across datasets with very different energy scales.

### Data Availability

To ensure reproducibility and facilitate future studies, benchmark data should be openly available and easily accessible in electronic form. Essential information includes Cartesian coordinates, charges, multiplicities, reference values, and details on the level of theory. Transparency in data presentation also supports applications beyond benchmarking, such as training machine-learning models.

Designing a benchmark set is therefore a multifaceted endeavor that goes far beyond simply collecting data. Every choice, from the selection of molecules and reference methods to the statistical treatment of results, influences the conclusions that can ultimately be drawn. Small oversights in one step can propagate and compromise the reliability of the entire study. While well-designed benchmarks provide powerful and lasting insights into the performance of computational methods, their construction requires a high level of care, consistency, and expertise. It is precisely this complexity that makes benchmark design a challenging but essential foundation for meaningful method assessment.

## CHAPTER 3

# Optimization of the $r^2$SCAN-3c Composite Electronic-Structure Method for Use with Slater-Type Orbital Basis Sets

Thomas Gasevic,[†] Julius B. Kleine Büning, Stefan Grimme, and Markus Bursch

*Received: April 28, 2022*
*First published: June 2, 2022*

Reprinted in Appendix A (adapted) with permission from:
T. Gasevic, J. B. Stückrath, S. Grimme, and M. Bursch, *Optimization of the $r^2$SCAN-3c Composite Electronic-Structure Method for Use with Slater-Type Orbital Basis Sets*, J. Phys. Chem. A **126**.23 (2022) 3826, DOI: `10.1021/acs.jpca.2c02951`

**Own contributions**

- Conceptualization
- Data Curation
- Formal Analysis & Investigation
- Software – Implementing $r^2$SCAN-3c(STO) into the Amsterdam Modeling Suite (AMS)
- Visualization
- Writing – Original Draft Preparation, Review & Editing

---

[†] Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany
[‡] Max-Planck-Institut für Kohlenforschung, 45470 Mülheim an der Ruhr, Germany

The composite density functional method $r^2$SCAN-3c has already established itself as a highly efficient and reliable approach for a broad range of quantum chemical applications. Originally, it combines the meta-GGA density functional $r^2$SCAN with the well-balanced mTZVPP triple-$\zeta$ GTO basis set, the D4 London dispersion correction, and the gCP correction for residual BSSE. Extensive benchmark studies have shown that this cost-effective method can rival hybrid density functionals with large basis sets in terms of accuracy while drastically reducing computational cost.[43] However, the method was previously limited to GTOs, which, despite their practical advantages, have known shortcomings in correctly representing the near-nuclear behavior and long-range decay of the analytical hydrogen-like atomic orbitals (cf. Chapter 2.2.4).

In this work, $r^2$SCAN-3c was adapted for use with STOs as implemented in the Amsterdam Density Functional (ADF) program of the Amsterdam Modeling Suite.[109] In analogy to the original approach, $r^2$SCAN is here combined with a specially tailored all-electron STO triple-$\zeta$ basis set, termed mTZ2P, along with STO-specific reparametrizations of the D4 (cf. Chapter 2.2.3) and gCP corrections (cf. Chapter 2.2.4). Scalar relativistic (SR) effects are, per default, accounted for *via* the all-electron basis set and ZORA, which replaces the ECPs that are used in the GTO variant. This modification retains the ability to describe core-electron effects explicitly and, with that, also allows for the optional explicit treatment of spin-orbit coupling effects. The mTZ2P basis set is derived by adapting and merging the existing DZP, TZP, and TZ2P STO basis sets available in ADF, followed by manual reoptimization of diffuse polarization functions for the hydrogen and oxygen atoms. These refinements were essential to reduce overall errors and retain efficiency. For the D4 correction, only the $s_9$ parameter, which scales the three-body dispersion, was reparametrized, reducing it from 2.00 to 1.53. For the gCP correction, the global scaling factor $\sigma$ was decreased from 1.000 to 0.879. Unlike its parent functional SCAN, $r^2$SCAN does not suffer from numerical instabilities as much. Nevertheless, it was found to require tighter settings in the ADF program, and the *NumericalQuality* parameter was therefore set to *good* as a default.

To validate the STO implementation, an extensive benchmark study was conducted comprising 82 datasets, including eight geometry and 74 energy test sets, covering main-group and transition metal chemistry. The total evaluation included 621 geometry data points and 4405 relative energies. In geometry optimizations, $r^2$SCAN-3c(STO) is either on par with or outperforms the GGA BP86-D4[77,175] as well as the hybrid M06-2X-D3(0)[176] DFA in combination with the TZP basis set for various molecular classes, including organometallic complexes and weakly bound systems. Compared to the original method, its accuracy in geometry optimizations is very similar.

The performance for relative energies was evaluated using the GMTKN55 database and several additional test sets. The STO-based variant achieved a WTMAD-2 of 7.15 $\text{kcal}\cdot\text{mol}^{-1}$on the GMTKN55, surpassing the original GTO-based $r^2$SCAN-3c with 7.50 $\text{kcal}\cdot\text{mol}^{-1}$and even some hybrid functionals with large quadruple-$\zeta$ basis sets. In comparison to the original approach, it shows superior performance for barrier heights and basic thermochemical properties. Notably, the improved accuracy in describing ion-$\pi$ interactions and some transition-metal complexes underlines the benefits of explicitly treating core electrons and relativistic effects. From a computational efficiency perspective, the STO implementation in ADF is highly competitive. Comparing timings within the ADF program for single-point energy calculations reveals that $r^2$SCAN-3c(STO) is approximately 2.5 times faster than the hybrid DFA PBE0/TZ2P and over ten times faster than PBE0/QZ4P, while offering similar accuracy.

In conclusion, the STO-based $r^2$SCAN-3c variant offers a robust, flexible, and accurate alternative to the original GTO implementation. It maintains or improves upon the already excellent accuracy of the method, particularly in describing noncovalent interactions and transition-metal chemistry, while enabling a more rigorous treatment of relativistic effects through the application of ZORA and an all-electron basis set. Thanks to its efficient computational performance and broad applicability, r$^2$SCAN-3c(STO) emerges as an attractive choice for high-throughput and routine quantum chemical calculations, particularly within the ADF framework.

# CHAPTER 4

# Benchmark Study on the Calculation of $^{207}$Pb NMR Chemical Shifts

Thomas Gasevic,[†] Julius B. Kleine Büning, Stefan Grimme, and Markus Bursch

*Received: December 21, 2023*
*First published: March 6, 2024*

Reprinted in Appendix B (adapted) with permission from:
T. Gasevic, J. B. Kleine Büning, S. Grimme, and M. Bursch, *Benchmark Study on the Calculation of $^{207}$Pb NMR Chemical Shifts*, Inorg. Chem. **63**.11 (2024) 5052, DOI: `10.1021/acs.inorgchem.3c04539`

**Own contributions**

- Conceptualization
- Data Curation – Collecting experimental $^{207}$Pb NMR Chemical Shifts and Crystal Structures
- Formal Analysis & Investigation – Generating Conformer Ensembles
- Visualization
- Writing – Original Draft Preparation, Review & Editing

---

[†] Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany

[‡] Max-Planck-Institut für Kohlenforschung, 45470 Mülheim an der Ruhr, Germany

After the synthesis of chemical compounds, it is essential to elucidate their structures using sophisticated methods, where NMR spectroscopy is often the method of choice. Usually, only the light NMR-active $^{1}$H and $^{13}$C nuclei are investigated as they are the most common in organic synthesis. Still, heavier nuclei, such as $^{207}$Pb, are also of high interest, as they can detect lead-containing compounds and, with that, also indicate toxicity. Since the range of shifts becomes larger as the nuclei become heavier, experiments sometimes struggle to locate the expected shift, as only a limited range can be measured. Computations can therefore aid and verify measurements, but choosing a proper computational method remains a considerable challenge. To investigate which approximation yields the best results for $^{207}$Pb NMR chemical shifts, the *PbS50* benchmark set is introduced, a comprehensive collection of 50 experimentally characterized organolead compounds with up to 187 atoms designed to probe a variety of coordination motifs, bonding scenarios, and solvent environments. The dataset encompasses seven distinct coordination types with up to seven bonding partners, as well as six common NMR solvents, with experimental $^{207}$Pb chemical shifts ranging from +10745 to -5030 ppm with a mean absolute value of 2599 ppm. A multi-level workflow was applied to create conformer ensembles for each structure using CREST at the GFN2-xTB level of theory, followed by refinement using CENSO at the r$^{2}$SCAN-3c(CPCM) level of theory, yielding a total of 282 conformers. Boltzmann-averaged NMR shifts for each compound were computed in the respective solvent that was applied in the NMR measurements using 10 different density functionals. As non-relativistic calculations can lead to significant errors for heavier elements, the SR- and SO-ZORA treatments were applied with the ZORA/TZP triple-$\zeta$ basis set. The analysis reveals that purely SR-ZORA-based $^{207}$Pb NMR calculations suffer from systematic underestimation, with MAEs of 1999-2123 ppm for all tested DFAs. The introduction of spin-orbit coupling *via* SO-ZORA significantly reduces these errors, lowering MAEs to the 429-856 ppm range. This dramatic improvement underscores the strong influence of heavy-atom on heavy-atom (HAHA) spin-orbit coupling (SOC) interactions on $^{207}$Pb chemical shifts and highlights the necessity of explicit SO treatments for accurate predictions.

Within the SO-ZORA framework, the Jacob's ladder behavior is clearly visible. GGAs perform approximately 40-60 % worse than hybrid DFAs, even after applying an empirical linear scaling correction (cf. Chapter 2.5). Hybrid DFAs such as PBE0 and mPW1PW[177] emerge as the top performers, delivering MAEs of 446 ppm and 429 ppm, respectively. Applying the empirical linear correction to the hybrid DFA results further reduces the MAEs of PBE0 and mPW1PW to 341 and 338 ppm, respectively. All systems were initially computed using the AMS program package, which offers only a limited selection of density functionals. To further extend the comparison, a subset of systems without significant SOC contributions was subsequently evaluated with the ORCA program package[178] that does not include the support for SOC in NMR calculations. For this subset, additional density functionals were tested, revealing that range-separated hybrids yield even lower MAEs than global hybrids.

Conformational contributions were evaluated by comparing Boltzmann-weighted ensemble averages to single lowest-energy conformer calculations. As the benchmark set mainly contains rigid coordination environments, the differences in the MAE remain below 5.5 ppm, validating the use of single-structure approaches in such cases. However, the most flexible organolead system in this set exhibits deviations of up to 133 ppm, depending on the conformer, indicating that more flexible structures require proper conformer sampling.

To evaluate the impact of the geometry optimization level on the workflow results, different methods were tested while computing the NMR shifts at the PBE0/SO-ZORA/TZP level. First, the importance of an explicit relativistic treatment was investigated for the geometry optimization, comparing ECPs with

SO-ZORA using r$^2$SCAN-3c. As the results are very similar, and the ECP-based approach is usually computationally more efficient, it is recommended to apply it for the geometry optimizations. When combined with the empirical linear scaling protocol, GFN2-xTB geometries yield slightly lower errors than r$^2$SCAN-3c optimizations without the linear scaling. These results demonstrate that cost-effective SQM-driven pipelines, combined with targeted empirical corrections, can achieve accuracies similar to those of efficient meta-GGA-based DFT approaches. At the same time, they require only a fraction of the computational resources, making them very attractive for high-throughput screening of organolead compounds.

In conclusion, these findings provide guidelines for establishing a practical, multi-level workflow that balances computational cost and the accuracy in the computation of $^{207}$Pb NMR chemical shifts.

## CHAPTER 5

# The $p$-block challenge: assessing quantum chemistry methods for inorganic heterocycle dimerizations

Thomas Gasevic,[†] Markus Bursch,[‡§] Qianli Ma,[¶] Stefan Grimme[†], Hans-Joachim Werner,[¶] and Andreas Hansen[†]

*Received: December 21, 2023*
*First published: April 11, 2024*

Reprinted in Appendix C (adapted) with permission from:
T. Gasevic, M. Bursch, Q. Ma, S. Grimme, H.-J. Werner, and A. Hansen, *The p-block challenge: assessing quantum chemistry methods for inorganic heterocycle dimerizations*, Phys. Chem. Chem. Phys. **26**.18 (2024) 13884, DOI: `10.1039/d3cp06217a`
T. Gasevic, M. Bursch, Q. Ma, S. Grimme, H.-J. Werner, and A. Hansen, *Correction: The p-block challenge: assessing quantum chemistry methods for inorganic heterocycle dimerizations*, Phys. Chem. Chem. Phys. **27**.16 (2025) 8572, DOI: `10.1039/d5cp90062j`

**Own contributions**

- Data Curation & Investigation – Performing DFT calculations
- Formal Analysis
- Software – Writing Python-based code for the generation of dimer structures
- Writing & Visualization

---

[†] Mulliken Center for Theoretical Chemistry, Universität Bonn, Beringstr. 4, D-53115 Bonn, Germany
[‡] Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, D-45470 Mülheim an der Ruhr, Germany
[§] FACCTs GmbH, D-50677, Köln, Germany
[¶] Institut für Theoretische Chemie, Universität Stuttgart, Pfaffenwaldring 55, D-70569 Stuttgart, Germany

Inorganic compounds composed of main-group elements from the $p$-block are gaining increasing attention due to their relevance in a wide range of applications, such as catalysis, optoelectronics, and materials design.[180–187] However, many of these compounds exhibit unconventional bonding motifs and involve heavier elements, which makes them challenging for electronic structure theory. Most established benchmark sets focus on (bio)organic molecules or transition-metal complexes, while chemically diverse $p$-block systems remain underrepresented. This gap limits both the development and the assessment of QC methods.

To address this issue, the *IHD302* benchmark set is introduced, comprising 604 dimerization energies of 302 singlet ground state, inorganic six-membered heterocycles. These structures are built from $p$-block elements (from B to Po) of groups 13-16, excluding carbon. Each heterocycle appears in two dimeric forms: one with covalent inter-ring bonding (COV) and one with non-covalent, weak donor-acceptor (WDA) interactions. The resulting dimers encompass a range of chemically diverse bonding scenarios, from weakly bound van der Waals complexes in the WDA subset to crown-shaped, covalently linked species in the COV subset. To generate such structures, monomer geometries were created using chemically intuitive valence rules and saturated with three or six hydrogen atoms to avoid open-shell or highly ionic systems. Dimers were then constructed in two ways: WDA dimers were created by rotating one monomer by 180° and separating the ring centers by twice the van der Waals radius of the heaviest atom involved. In contrast, covalent dimers were obtained by optimizing the geometries of the WDA dimers using r$^2$SCAN-3c. All systems were checked for spin symmetry breaking using unrestricted PBE0-D4 as an indication of potential multireference character. In addition, eight representative systems were tested for fractional occupation density (FOD)-based diagnostics[188–190] to confirm that static correlation effects remain negligible. High-quality reference energies were computed at the PNO-LCCSD(T)-F12 level using the cc-VTZ-PP-F12 basis set[191] and the ECP10MDF[112] pseudopotentials, while including correlation explicitly for both the valence-valence and core-valence parts. To improve accuracy, a composite scheme was employed involving a basis set correction to approximate the basis set limit using PNO-LMP2-F12/aug-cc-pwCVTZ. The final reference level of theory is referred to as PNO-LCCSD(T)/vtz-f12(corr.) and is computed as:

$$\begin{aligned} E_{\text{PNO-LCCSD(T)/vtz-f12(corr.)}} = & E_{\text{LCCSD(T)-F12/cc-VTZ-PP-F12}} - E_{\text{LMP2-F12/cc-VTZ-PP-F12}} \\ & + E_{\text{LMP2-F12/aug-cc-pwCVTZ}} \end{aligned} \tag{5.1}$$

The reference energies range from −165.1 to 9.3 kcal·mol$^{-1}$in the COV set with a mean absolute value of 95.8 kcal·mol$^{-1}$, while they range from −29.2 to −1.2 kcal·mol$^{-1}$in the WDA set with a mean absolute value of 12.4 kcal·mol$^{-1}$.

A significant challenge in the DFT benchmarking study arose from the absence of relativistic pseudopotentials for Ga, Ge, As, and Se in the def2-ECP set. As these elements are described using ECP10MDF pseudopotentials in the reference calculations, this discrepancy complicates a consistent comparison between the two approaches. This introduces errors of up to 6 kcal·mol$^{-1}$in the DFT calculations due to missing relativistic effects in the predicted dimerization energies of the covalent set. To overcome this issue, the aug-cc-pVQZ-PP-KS basis set was developed and re-contracted using atomic PBE0 calculations for these elements. It already introduces ECPs starting from Ga, instead of Kr. This modification improves consistency across the periodic table and restores the accuracy for the affected systems.

To evaluate the performance of approximate QC methods on this challenging dataset, 26 DFAs were tested. These functionals span all rungs of the Jacob's ladder and were tested with three London dispersion correction schemes: DFT-D3, DFT-D4, and DFT-NL. In addition, the performance of five composite DFT schemes and five SQM methods was evaluated. The results show systematic overbinding when the London dispersion corrections are combined with already attractive functionals, particularly for hybrid and double-hybrid functionals. Functionals explicitly trained with a London dispersion correction (e.g., $\omega$B97M-V) show relatively good performance but still overbind, while Minnesota functionals consistently underestimate dimerization energies. Importantly, potential energy surface scans comparing PBE0 results with the high-level reference method confirm that the asymptotic behavior is fully recovered when a London dispersion correction is applied. At short and medium intermolecular distances, however, significant deviations remain. These residual errors can be attributed primarily to the exchange-correlation functional rather than the dispersion correction itself.

Excellent results were obtained for both the covalent and weak donor-acceptor subsets using r$^2$SCAN-D3(BJ), $\omega$B97M-V, and r$^2$SCAN0-D4.[192] The best-performing DFA is revDSD-PBEP86-D4(2021), which yields an MAE of 2.3 and 0.7 kcal·mol$^{-1}$for the COV and WDA sets, respectively. The r$^2$SCAN-3c and $\omega$B97X-3c composite DFT methods also perform very well across both subsets, with RMSEs close to or better than those of the respective large-basis equivalent, highlighting their robustness and efficiency.

For SQM methods, such as the GFN$n$-xTB and PM$x$ families, the limitations of their parameterization become evident in the IHD302 set. PM6 and PM7 exhibit large deviations and outliers that exceed the average reference dimerization energies (MAE$_{\text{COV}}$=110.0 and 115.2 kcal·mol$^{-1}$, respectively), making them unsuitable for structure motifs covered by this benchmark. In contrast, GFN1-xTB (MAE$_{\text{COV}}$=32.5 kcal·mol$^{-1}$) and GFN2-xTB (MAE$_{\text{COV}}$=27.9 kcal·mol$^{-1}$) yield substantially better results, making them suitable for initial screening purposes. Nevertheless, they still exhibit errors that are an order of magnitude larger than those obtained using DFT.

Overall, the IHD302 benchmark set bridges a significant gap in current benchmark coverage, contributing to more reliable and transferable electronic structure methods for the description of inorganic and main-group chemistry.

# CHAPTER 6

# Chemical Space Exploration with Artificial "Mindless" Molecules

Thomas Gasevic,[†] Marcel Müller,[†] Jonathan Schöps,[†] Stephanie Lanius,[‡], Jan Hermann,[‡] Stefan Grimme,[†] and Andreas Hansen[†]

*Received: June 16, 2025*
*First published: September 2, 2025*

Reprinted in Appendix D (adapted) with permission from:
T. Gasevic, M. Müller, J. Schöps, S. Lanius, J. Hermann, S. Grimme, and A. Hansen, *Chemical Space Exploration with Artificial "Mindless" Molecules*, J. Chem. Inf. Model. **65**.18 (2025) 9576, DOI: `10.1021/acs.jcim.5c01364`

**Own contributions**

- Conceptualization
- Formal Analysis
- Investigation – Generating the MB2061 set & computing all results, excluding the reference energies
- Visualization
- Writing – Original Draft Preparation, Review & Editing

---

[†] Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany
[‡] Microsoft Research AI for Science, Microsoft Research, D-10117 Berlin, Germany

Understanding and exploring the so-called *chemical space*,[193] that describes the multidimensional landscape of all theoretically possible molecules, is a central goal in modern computational chemistry. This pursuit spans applications from materials science to drug discovery and catalysis. Despite considerable progress, large portions of chemical space remain inaccessible to experiment, necessitating computational strategies to generate and study hypothetical compounds far from known chemistry.[194–196] Such endeavors require robust tools capable of handling the diversity and unpredictability of these regions.

To probe such areas, the `MindlessGen` was developed in this work, a modular and open-source Python-based generator that creates artificial "mindless" molecules (MLMs). The software enables systematic exploration of uncharted chemical motifs, assisting both in method development and dataset generation. It begins by placing atoms randomly in space while satisfying the given stoichiometric and charge constraints. A geometric contraction step ensures the formation of compact initial guesses, which are subsequently optimized using a GFN$n$-xTB method. A graph-based algorithm checks for fragmentation by analyzing interatomic distances using covalent radii from Pyykkö and Atsumi.[197] If fragmentation occurs, only the largest connected fragment is reoptimized and rechecked for fragmentation. This iterative cycle continues until a single, connected molecule is obtained. The HOMO-LUMO gap is checked, which serves as a crude diagnostic for potential multireference (MR) character. Optionally, the workflow can generate symmetric structures, build supramolecular complexes, and perform automatic refinements at higher levels of theory (e.g., DFT).

Using this workflow, the *MB2061* benchmark set was constructed, which contains 2061 closed-shell MLMs with system sizes ranging from 8 to 20 atoms and elemental compositions that span elements from hydrogen to iodine. Noble gases and transition metals were excluded to avoid chemically inert or multireference cases, respectively. Randomly selected charges between −2 and +2 were assigned to ensure electronic closed-shell configurations. Final geometries were optimized at the DFT level using PBE0-D4/def2-TZVP, and several MR diagnostics, including FOD analysis and the SCF stability analysis,[198] were used to validate the applicability of single reference methods. Structures with unwanted fragmentations (e.g., containing $H_2O$, $CO^-$, or HCl) that were not detected by the fragment recognition step were filtered out through manual curation.

To benchmark computational methods, decomposition reactions were defined where each MLM reacts with $H_2$ to form hydrides and the diatomics for group 15 and 17 elements:

$$\square\, \mathrm{C}_{\square}\mathrm{H}_{\square}\mathrm{B}_{\square}\mathrm{N}_{\square}\ldots + \square\, \mathrm{H}_2 \longrightarrow \square\, \mathrm{CH}_4 + \square\, \mathrm{BH}_3 + \square\, \mathrm{N}_2 + \ldots \tag{6.1}$$

High-level reference reaction energies were computed using PNO-LCCSD(T)-F12.[63–68] The reaction energies have a broad distribution ranging from −1232.4 to +1607.8 kcal·mol$^{-1}$, with a mean absolute value of 306.3 kcal·mol$^{-1}$. Several classes of methods were tested, including DFAs, SQMs, FFs, and MLIPs. Further, the dataset was subdivided into focused subsets (e.g., *MB727-Light*, *MB782-s-Block*, *MB552-Heavy*).

The DFT results show a clear trend following the Jacob's ladder of DFAs, with double-hybrids performing best. $\omega$B97X-2 yields the lowest errors (MAE = 8.4 kcal·mol$^{-1}$), while r$^2$SCAN-3c (MAE = 19.6 kcal·mol$^{-1}$) shows the best balance between computational cost and accuracy. While the *Light* subset exhibited the lowest errors, increased errors were observed for the $s$-block species and for heavier elements. For example, the MAE of $\omega$B97X-2 increases from 5.9 to 13.7 kcal·mol$^{-1}$going from the *Light* to the *Heavy* subset. Surprisingly, performance differences among DFAs did not correlate consistently with the degree of empiricism or fitting strategy, challenging previous assumptions about overfitting and generalization. Conventional SQM methods struggle with this chemically diverse dataset as all tested methods exhibited very large errors (MAE > 180 kcal·mol$^{-1}$). Notably, even the GFN-FF force field surpassed some older SQM approaches in accuracy. However, the development version of the currently developed g-xTB[199] SQM method achieves much better results (MAE ≈ 35 kcal·mol$^{-1}$), approaching DFT-level accuracy. MLIPs also show a very strong performance. For instance, AimNet2 performs on par with PBE0 for the reduced subset of elements accessible to it. At the same time, UMA-sm attains an MAE of 24.0 kcal·mol$^{-1}$on the full dataset, approaching the accuracy of the best-performing composite methods. While some larger outliers remained, especially for heavier elements, these results indicate that modern MLIPs are capable of describing the energy of electronically unusual molecules with moderate accuracy.

Since the dataset includes elements beyond krypton, relativistic effects could also be assessed explicitly. X2C computations show that the commonly employed ECP-based treatments deviate only slightly, with an MAE of up to 6.2 kcal·mol$^{-1}$relative to the X2C approach with SOC, thereby supporting their suitability for benchmarking purposes.

In summary, the `MindlessGen` and the *MB2061* benchmark set provide a robust platform for generating and evaluating electronically and structurally diverse molecules. This significantly contributes to the validation and development of quantum chemical methods, particularly beyond conventional chemistry. It is well-suited for method training, benchmarking, and active learning frameworks.

## CHAPTER 7

# Confined Lewis Pairs: Investigation of the $X^{-}\rightarrow Si_{20}$ Interaction in Halogen-Encapsulating Silafulleranes

Thomas Gasevic,[†] Marcel Bamberg,[‡] Julius Wicke,[‡] Michael Bolte,[‡] Alexander Virovets,[‡] Hans-Wolfram Lerner,[‡] Stefan Grimme,[†] Andreas Hansen,[†] Matthias Wagner,[‡] and Markus Bursch,[†]

*Received: September 27, 2023*
*First published: December 7, 2023*

Reprinted in Appendix E (adapted) with permission from:
T. Gasevic, M. Bamberg, J. Wicke, M. Bolte, A. Virovets, H.-W. Lerner, S. Grimme, A. Hansen, M. Wagner, and M. Bursch, *Confined Lewis Pairs: Investigation of the $X^{-}\rightarrow Si_{20}$ Interaction in Halogen-Encapsulating Silafulleranes*, Angew. Chem., Int. Ed. **63**.6 (2024) e202314238, DOI: `10.1002/anie.202314238`

**Own contributions**

- Conceptualization
- Formal Analysis
- Investigation – Performing DFT and LED calculations
- Writing & Visualization – Original draft preparation, review & editing

---

[†] Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany

[‡] Institut für Anorganische und Analytische Chemie, Goethe-Universität Frankfurt am Main, 60438 Frankfurt am Main, Germany

After establishing the importance of careful benchmarking in quantum chemistry, particularly for the selection of density functionals, basis sets, and dispersion corrections, these validated methods are now applied to the structurally unique, halide-encapsulating silafulleranes of the type $[X@Si_{20}Y_{20}]^-$. In these clusters, a halide ion $X^-$ ($F^-$, $Cl^-$, $Br^-$, $I^-$) is confined within a dodecahedral $Si_{20}$ cage, which is saturated with different exohedral substituents Y = F, Cl, Br, I, H, Me, Et. These systems are best described as confined Lewis pairs (CLPs), where the electron-rich guest interacts strongly with an electron-deficient host, yet cannot be easily separated due to the rigid cage environment. This spatial confinement gives rise to distinct electronic features that are not accessible in classical Lewis pairs or even in conventional host-guest complexes. The present study combines high-level quantum chemical methods to evaluate the nature, strength, and energetic consequences of the $X^- \rightarrow Si_{20}$ interaction.

First, to better understand the formation process of these endohedral species, a series of silapolyquinane model systems was constructed, simulating the stepwise growth of the $Si_{20}$ cage in $[Cl@Si_{20}Cl_{20}]^-$. These fragment-based models (from $Si_5$ to $Si_{20}$) clearly support the hypothesis of a structure-directing template effect induced by the $Cl^-$ ion: The interaction energies computed at the DLPNO-CCSD(T)/*TightPNO*/def2-TZVPP level steadily increase with cluster size, from $-78.3$ kcal·mol$^{-1}$ in the smallest unit $[Cl@Si_5H_5Cl_5]^-$ to $-146.6$ kcal·mol$^{-1}$ in the full cage. This steady increase reflects the cooperative nature of the interaction, in which the growing number of silicon atoms provides increasingly favorable electrostatic stabilization. In contrast, the orbital relaxation effects per Si atom remain consistent. These findings align with the observed fact that the empty $Si_{20}Y_{20}$ cage is synthetically inaccessible without the presence of the halide ion.[200]

In the next step, different endohedral guests and exohedral substituents were evaluated. The respective interaction energies $E_{int}$ computed at the PNO-LCCSD(T)-F12b/AVTZ'/default level in gas phase span a wide range, from $-60.9$ kcal·mol$^{-1}$ for weakly bound systems like $[Cl@Si_{20}Et_{20}]^-$ to $-159.2$ kcal·mol$^{-1}$ for the strongly interacting $[Cl@Si_{20}F_{20}]^-$. The corresponding free energies of association, computed at the r$^2$SCAN0-D4[192] level in combination with the def2-QZVPPD (on $X^-$), def2-TZVPPD (on Si), and def2-QZVPP (on Y) basis sets, as well as the COSMO-RS($CH_2Cl_2$) solvation contribution, and thermostatistical corrections, follow similar trends: The most stable complexes show $\Delta G$ values below $-90$ kcal·mol$^{-1}$, whereas systems with electron-donating or bulky alkyl substituents can even become thermodynamically unstable ($\Delta G > 0$), highlighting the delicate balance between attractive electronic forces and destabilizing steric or entropic contributions. Both the EDA and LED schemes reveal that the attractive interaction is primarily electrostatic in character, but also contains significant orbital-relaxation contributions that depend sensitively on both the size of the encapsulated halide and the electronic nature of the exohedral substituents.

Complementing the energetic analysis, theoretical $^{35}$Cl NMR computations for the chloride encapsulating systems were performed at the SO-ZORA-PBE0/TZP level of theory. The calculated chemical shifts range from $-61.4$ ppm for $[Cl@Si_{20}F_{20}]^-$, which features the strongest interaction, up to 340.9 ppm for $[Cl@Si_{20}H_{20}]^-$, one of the more weakly bound systems in the study. The paramagnetic shielding component $\sigma_{para}$ was identified as the dominant factor influencing the total isotropic shielding, while showing a strong inverse correlation with both the donor-acceptor orbital gap and the interaction energy. This analysis shows that $^{35}$Cl NMR is a sensitive and reliable probe for assessing the strength of the $Cl^- \rightarrow Si_{20}$ interaction, analogous to the Gutmann-Beckett method[201,202] that correlates the $^{31}$P NMR shift for probing Lewis acidity.

On the experimental side, the study expands the library of accessible silafulleranes by synthesizing and characterizing new $[Cl@Si_{20}Y_{20}]^-$ derivatives with Y = Me, Et, Br. These compounds were isolated as salts of $[PPN]^+ = [(Ph_3P)_2N]^+$ and characterized by NMR spectroscopy. The crystal structures confirm the expected dodecahedral geometry, and show subtle but systematic variations in Si–Si and Cl–Si distances that correlate with the computationally predicted cage strain and interaction strength. For instance, $[Cl@Si_{20}Me_{20}]^-$ shows a cage slightly expanded compared to its halogenated counterparts, consistent with the reduced electrostatic attraction. The experimental $^{35}Cl$ chemical shifts ($\delta_{exp}$ = 332.7 vs. $\delta_{calc}$ = 333.3 ppm for $[Cl@Si_{20}Me_{20}]^-$, $\delta_{exp}$ = 326.3 vs. $\delta_{calc}$ = 322.3 ppm for $[Cl@Si_{20}Et_{20}]^-$, and $\delta_{exp}$ = 210.9 vs. $\delta_{calc}$ = 205.4 ppm for $[Cl@Si_{20}Br_{20}]^-$) closely match the computations, reinforcing the validity of the theoretical approach.

In summary, this combined computational and experimental study demonstrates how quantum chemical methods can yield profound chemical insight into the bonding, stability, and NMR properties of confined Lewis pairs. By revealing the cooperative effects between endohedral ions and exohedral substituents, quantifying their decomposed interaction energies, and verifying a robust connection to experimental observables, this work sets the stage for the rational design of similar clusters.

# CHAPTER 8

# Regioselective Derivatization of Silylated [20]Silafulleranes

Marcel Bamberg,[†] Thomas Gasevic,[‡] Michael Bolte,[†] Alexander Virovets,[†] Hans-Wolfram Lerner,[†] Stefan Grimme,[‡] Markus Bursch,[‡] and Matthias Wagner[†]

*Received: March 29, 2023*
*First published: May 12, 2023*

Repinted in Appendix F (adapted) with permission from:
M. Bamberg, T. Gasevic, M. Bolte, A. Virovets, H.-W. Lerner, S. Grimme, M. Bursch, and M. Wagner, *Regioselective Derivatization of Silylated [20]Silafulleranes*, J. Am. Chem. Soc. **145**.20 (2023) 11440, DOI: `10.1021/jacs.3c03270`

**Own contributions**

- Formal Analysis
- Investigation – Performing QC calculations
- Writing – Review & Editing

---

[†] Institut für Anorganische und Analytische Chemie, Goethe-Universität Frankfurt am Main, 60438 Frankfurt am Main, Germany

[‡] Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany

Building on the previously established computational framework for the study of halide-encapsulating silafulleranes, the present work extends these efforts by applying an advanced quantum chemical approach as well as experimental synthesis and analysis to investigate the regioselective derivatization of the $Si_{20}$-based siladodecahedrane scaffold. Earlier studies primarily focused on the structural characterization, halide encapsulation, and initial functionalization steps of $[Si_{20}]^-$-type frameworks, including full hydrogenation or chlorination.[203] The current study instead investigates controlled, regioselective chemical transformations of the precursor $[Cl@Si_{20}(SiCl_3)_{12}Cl_8]^-$ ($[\mathbf{1}]^-$) into selectively functionalized silafullerane derivatives.

Three key target compounds were synthesized and characterized: the fully hydrogenated species $[Cl@Si_{20}(SiH_3)_{12}H_8]^-$ ($[\mathbf{2}]^-$), the partially hydrogenated species $[Cl@Si_{20}(SiH_3)_{12}Cl_8]^-$ ($[\mathbf{3}]^-$), and the mixed-substituted methylated species $[Cl@Si_{20}(SiH_3)_{12}Me_8]^-$ ($[\mathbf{4}]^-$). These transformations were achieved using different aluminum hydride reagents under carefully optimized conditions. In particular, an excess of $i$$Bu_2AlH$ in ortho-difluorobenzene (oDFB) solvent led to quantitative Cl/H exchange yielding $[\mathbf{2}]^-$. Switching to an oDFB/$Et_2O$ solvent mixture suppressed core hydrogenation and enabled a selective conversion to $[\mathbf{3}]^-$. Finally, a combination of $Me_2AlH$ and $Me_3Al$ resulted in simultaneous Cl/H and Cl/Me exchange, affording $[\mathbf{4}]^-$ in high yield. Crystallographic analysis confirmed the expected substitution patterns and verified all structures

To rationalize the observed regioselectivity, quantum chemical calculations were performed at the $r^2$SCAN0-D4/ma-def2-QZVPP//PBEh-3c(SMD($CH_2Cl_2$))[192,204] level of theory with the COSMO-RS($CH_2Cl_2$) solvation model and $G_{\mathrm{mRRHO}}$ thermostatistical corrections. Two model systems were designed: $[\mathbf{M1}]^-$, representing substitution at exohedral $SiCl_3$ groups, and $[\mathbf{M2}]^-$, mimicking substitution at cage-bound Si–Cl sites. In both models, four vertices were modeled explicitly, while the remainder was truncated to Si–H groups. These simplified structures captured the essential electronic characteristics of their respective environments while reducing computational cost. Two literature-known mechanisms for the hydrogenation of chlorosilanes using $i$$Bu_2AlH$ were considered: the intramolecular $S_N$i-Si pathway involving a four-membered transition state, and the $S_N$2-Si mechanism proceeding *via* a trigonal bipyramidal intermediate. Calculations revealed that the $S_N$i-Si activation barrier is significantly higher at the cage site ($\Delta G^{\ddagger} = 21.4\ \mathrm{kcal{\cdot}mol^{-1}}$) compared to the exohedral site ($\Delta G^{\ddagger} = 11.5\ \mathrm{kcal{\cdot}mol^{-1}}$). Moreover, it was found that the $S_N$2-Si mechanism at the exohedral position is even more favorable ($\Delta G = 9.0\ \mathrm{kcal{\cdot}mol^{-1}}$), making it the preferred route. Upon coordination of $Et_2O$ to the aluminum hydride reagent, by forming a bulky $iBu_2AlH \cdot OEt_2$ adduct, the activation barriers are significantly increased. Under such conditions, core hydrogenation becomes inaccessible at ambient temperature ($\Delta G^{\ddagger} = 34.1\ \mathrm{kcal{\cdot}mol^{-1}}$), thereby kinetically only favoring selective derivatization of exohedral sites and leading to the formation of $[\mathbf{3}]^-$ in the experiment.

$^{35}$Cl NMR spectroscopy further supported the found regioselectivity of the substitution reactions. The experimental chemical shifts of the endohedral chloride ion provided a sensitive probe of the surrounding environment. $[\mathbf{2}]^-$ exhibits the most downfield signal ($\delta$ = 469.0 ppm), consistent with the less electronegative $SiH_3$ substituents. In contrast, $[\mathbf{4}]^-$ ($\delta$ = 457.1 ppm) and $[\mathbf{3}]^-$ ($\delta$ = 363.7 ppm) show more upfield values, whereas $[\mathbf{1}]^-$ displayed the most upfield value ($\delta$ = 274.5 ppm). These experimental findings could be accurately reproduced with quantum chemical NMR shift calculations at the same level of theory applied in the previous study. Notably, initial attempts to synthesize $[\mathbf{4}]^-$ from $[\mathbf{1}]^-$ by adding only $Me_2AlH$ without $Me_3Al$ lead to partially substituted cage species showing distinctive $^{35}$Cl NMR signals ranging from 469.1 to 456.8 ppm, indicative of a mixture of substitution patterns. This contradicts previous experiments on $[Cl@Si_{20}(H/Cl)_{20}]^-$ silafulleranes that showed severe line broadening upon symmetry breaking in the geometry, leading to undetectable signals. This discrepancy was solved by computing the EFG as its largest principal component correlates with the linewidth (cf. Chapter 2.5). It is found that the methyl substituents in [**4**]$^-$ lead to a distinctly smaller value than for the [Cl@Si$_{20}$H$_{12}$Cl$_8$]$^-$ species, explaining the persistence of sharp $^{35}$Cl signals even in asymmetric derivatives.

Taken together, this comprehensive joint experimental and theoretical investigation establishes the mechanistic foundations for regioselective substitution chemistry on Si$_{20}$-based silafulleranes. The good agreement between computed and experimental barrier heights, NMR shifts, and structural data underscores the performance of the employed quantum chemical protocol. The ability to selectively functionalize different sites on the silafullerane cage opens new avenues for the design of silicon-based nanostructures with tailored surface functionality.

## CHAPTER 9

# Summary and Outlook

The work presented in this thesis encompasses the development, evaluation, and application of quantum chemical methods, thereby addressing significant domains of modern quantum chemistry. By embracing the balance between cost and accuracy as a guiding principle, computational chemistry has not only shown the ability to interpret but also shape the frontiers of modern chemical research. The main goal of this work was to establish strategies that enable reliable theoretical insights across a broad chemical space, ideally while maintaining computational feasibility even for large and complex systems. This goal has been approached through three complementary directions: (i) the development of new methods, (ii) comprehensive benchmarking across chemically diverse and experimentally relevant data sets, and (iii) the application of these methods to designed model systems and experimentally motivated target molecules. The results obtained within these domains not only reflect the current state of the art but also clarify the practical boundaries of quantum chemical approaches, outlining possible pathways for their future advancement.

The first major contribution of this work lies in the field of method development. By extending the r$^2$SCAN-3c composite DFT method to Slater-type orbital basis sets, it was demonstrated that compact composite schemes can be successfully adapted to alternative basis set types without sacrificing the central design philosophy of balancing accuracy and efficiency. This required reparameterization of the D4 London dispersion and the gCP basis set error corrections to ensure stable and transferable performance. The resulting protocol illustrates how physically motivated approximations, combined with targeted empirical corrections, yield a versatile and computationally efficient method. Notably, the STO-based implementation also enables the combination of a composite DFT scheme with an explicit relativistic treatment due to the all-electron nature. Although ECPs already provide sufficient accuracy for many systems, a systematic exploration of explicit relativistic effects within such composite schemes remains a fascinating field of research, particularly for cases where relativistic contributions are expected to play a decisive role, such as heavy-element chemistry or systems with strong spin-orbit coupling.

The second central theme of this thesis concerned benchmarking. Three benchmark sets were constructed to evaluate the reliability of different QC methods against experimental and high-level reference data across various properties. The study of $^{207}$Pb NMR chemical shifts using the *PbS50* benchmark set provided a challenging test where the choice of method plays a decisive role. Through this analysis, it became clear that at least hybrid DFAs, such as PBE0 and mPW1PW, and the explicit treatment of relativistic effects through approximations, such as ZORA, are required to reproduce experimental values within meaningful margins of error for heavy element NMR. The investigation of inorganic heterocycle dimerization using the *IHD302* set extended this benchmarking philosophy to reaction energetics, again identifying systematic strengths and deficiencies among commonly employed methods. The Jacob's ladder of density functional theory could be observed, with double-hybrids, such as revDSD-PBEP86-D4, yielding the most accurate results. To save computation time, a composite scheme, like the aforementioned $r^2$SCAN-3c, can be applied, as it yields outstanding accuracy considering its computational cost. The third benchmarking part of the thesis extended into chemical space exploration. The generation and study of artificial "mindless" molecules provided a means of probing chemical space in an unbiased manner. While traditional sets are typically dominated by well-studied, synthetically accessible molecules, the molecules in the *MB2061* set were instead fully computer-generated. The decomposition reactions of these molecules, mediated by $H_2$, were then used to probe the robustness of computational methods under conditions far beyond those of conventional chemistry. For DFT, the results were similar to those of the *IHD302* set, with the double-hybrid $\omega$B97X-2 yielding the most accurate results and $r^2$SCAN-3c again providing the best compromise between cost and accuracy. By comparing different DFAs, it was also possible to assess the respective performance based on parameterization strategies. While it was expected that highly parameterized methods would perform worse on this set, this could not be verified. While commonly used SQM methods lack accuracy, promising results could still be achieved with the recently developed g-xTB SQM method. Surprisingly, MLIPs demonstrated strong performance, reaching the accuracy of hybrid DFT as indicated by the MAEs but being less robust, exhibiting a larger error spread. Overall, these results provide a foundation for the continued refinement of computational protocols for future studies. Notably, the benchmarks also reveal that accuracy is not a single, universal metric but depends strongly on the property of interest and the chemical context in which it is evaluated.

The third and last part of this work focused on the application of quantum chemical protocols to chemically relevant problems. Complementing the theoretical investigations presented earlier, these studies provided concrete demonstrations of the predictive power and general application of QC. $Si_{20}$ silafulleranes were investigated in the context of confined Lewis pairs with the goal to rationalize unusual bonding motifs and reactivity patterns that are difficult to characterize solely by experiment. DFT was first employed to construct a hypothetical series of structures representing the formation of the $Si_{20}$ cage. Within this series, it was shown that the endohedral $Cl^-$ ion acts as a template, with the interaction energy between the ion and the cage increasing progressively throughout the formation process. Local energy decomposition analysis uncovered that the principal attractive contributions originate from electrostatics and orbital-relaxation terms. Furthermore, it was demonstrated that $^{35}$Cl NMR shifts can serve as sensitive probes for evaluating the strength of the $Cl^- \rightarrow Si_{20}$ interaction, closely analogous to the Gutmann–Beckett method, which relies on $^{31}$P NMR shifts to probe Lewis acidity. This interaction can be tuned by modifying both the endohedral ion and the exohedral substituents. Computations identified $[Cl@Si_{20}F_{20}]^-$ as the compound exhibiting the strongest interaction energy. In a subsequent study, the mechanism of the hydrogenation of $[Cl@Si_{20}(SiCl_3)_{12}Cl_8]^-$ with $i$Bu$_2$AlH was explored experimentally as well as computationally. Two reaction pathways were considered, namely the $S_N2$-Si and the $S_Ni$-Si mechanisms, with hydrogenation being possible both at the exohedral $SiCl_3$ substituents and at the endohedral core. The analysis indicated that hydrogenation at the exohedral sites *via* an $S_N2$-Si mechanism is preferred, although reactions at the core cannot be entirely excluded. In the presence of $Et_2O$, however, hydrogenation is hindered at the core and occurs exclusively at the exohedral positions. Finally, $^{35}$Cl NMR shifts were computed for all investigated structures using the same workflow as presented in the benchmark study of $^{207}$Pb NMR shifts. Comparison with experimental data showed systematic deviations, which could be effectively corrected through the application of a linear scaling function.

Taken together, the contributions of this thesis reinforce a central conclusion. There is no single "best" method in quantum chemistry, but rather a hierarchy of approaches that must be carefully selected and combined according to the present problem. State-of-the-art method development, systematic benchmarking, and diverse applications collectively demonstrate that the most reliable outcomes arise from meticulously validated methodological choices tailored to the problem at hand. Despite these advances, several limitations remain and highlight directions for future research. As demonstrated by r$^2$SCAN-3c(STO), composite schemes offer an attractive compromise between accuracy and computational cost. However, their applicability to very large systems remains limited due to the drastically increasing computation time. In such cases, a speedup of multiple orders of magnitude can be achieved by employing SQM methods or MLIPs. This typically comes at the cost of reduced robustness, limiting reliability in specific cases. Nevertheless, g-xTB results have already demonstrated that incorporating more physically meaningful components, along with a sophisticated parameterization strategy, on a large and diverse set can deliver excellent results, approaching hybrid DFT accuracy at a fraction of the cost. A similar trend is observed for MLIPs, such as UMA-sm. While it was expected to fail for the unconventional MLMs due to its highly empirical nature, it demonstrated excellent performance. The results on the *MB2061* set have shown that the training set for UMA-sm already covers a significant portion of chemical space, which probably can be attributed to the inclusion of out-of-equilibrium structures in the training set. Still, it contains a few drawbacks, e.g., a larger standard deviation. In the future, training sets could be expanded even more to include more unique data points and greater chemical diversity, thereby ensuring even broader coverage of chemical space. For data generation, the concept of "mindless" chemical space exploration can be extended to provide more rigorous stress tests for emerging methods. In the *MB2061* set, the focus was primarily on well-behaved systems that could easily be described by DFT. By systematically generating more complex structures, such as those with unusual oxidation states and pronounced multireference character, it may aid in identifying weaknesses in both traditional and machine learning-based approaches before they emerge in practical applications. Such an approach could also be employed in self-supervised learning frameworks to enhance model generalization and accuracy. However, one must consider that the computation of references is usually the most problematic step in this process, as special problems, such as multireference cases, may require manual inspection. Apart from energy-related benchmark sets, it is also important to extend property-based benchmarking sets. The *PbS50* benchmark set not only provides a data set but also a workflow recommendation for the computation of $^{207}$Pb NMR chemical shifts. This has proven to be successful even in the computation of $^{35}$Cl NMR chemical shifts in the works of the chloride-encapsulating silafulleranes. Further, it could be applied in screening processes of unknown compounds. As an outlook for the silafulleranes, the differently substituted species identified in this work may serve as weakly coordinating anions and as versatile building blocks for reticular assemblies, opening potential applications in catalysis and materials design. Furthermore, Si-H functionalities in modified derivatives open possibilities for subsequent reactions such as hydrosilylation or dehydrogenative Si-Si coupling. These could, e.g., yield nano-sized oligomeric silafullerenes occupying the intermediate regime between small-molecule silanes and silicon-based solids.

From a broader perspective, this thesis has shown how method development, benchmarking, and application are not isolated endeavors but interconnected to form a research cycle. New methods require validation, benchmarks require interpretation, and applications motivate further improvements in accuracy and efficiency. Progress in computational chemistry is therefore best understood not as a linear sequence of innovations but as a feedback loop in which each component informs and enhances the others. Recognizing this interplay provides a framework for future research efforts and ensures that methodological progress remains closely tied to the practical needs of chemical science.

# Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Prof. Dr. Stefan Grimme, for his continuous support and inspiration throughout my studies. My time in his group has allowed me to grow not only academically but also personally, for which I am deeply thankful. I am also grateful to Dr. Andreas Hansen and Dr. Markus Bursch, whose guidance, encouragement, and supervision have significantly contributed to the success of this work.

I want to thank all remaining members of my Doctoral committee, Prof. Dr. Thomas Bredow, Prof. Dr. Sigurd Höger, and Prof. Dr. Petra Mutzel, for their time, interest, and valuable feedback on my thesis.

The success of this work would not have been possible without the support of my collaborators, to whom I extend my sincere thanks: Dr. Marcel Bamberg, Dr. Michael Bolte, Dr. Julius Kleine Büning, Prof. Dr. Marcos D. García, Dr. Jan Hermann, Stephanie Lanius, Dr. Hans-Wolfram Lerner, Dr. Jose Martinez Fernandez, Dr. Qianli Ma, Dr. Marcel Müller, Dr. Iago Neira, Prof. Dr. Carlos Peinador, Christoph Plett, Dr. Tyler G. Saint-Denis, Jonathan Schöps, Prof. Dr. T. Don Tilley, Dr. Alexander Virovets, Prof. Dr. Matthias Wagner, Prof. Dr. Hans-Joachim Werner, Julius Wicke, and Lukas Wittmann.

I would also like to thank all current and former members of the Grimme group, including Benedikt Bädorf, Dr. Fabian Bohle, Dr. Markus Bursch, Dr. Eike Caldeweyher, Robin Dahl, Sebastian Dohm, Dr. Sebastian Ehlert, Thomas Froitzheim, Marvin Friede, Johannes Gorges, Jan Kachnowicz, Abylay Katbashev, Dr. Julia Kohn, Dr. Jeroen Koopman, Lukas Kunze, Sarah Löffelsender, Dr. Jan Michael Mewes, Dr. Hagen Neugebauer, Andrea Pellegrini, Christoph Plett, Dr. Philipp Pracht, Dr. Thomas Rose, Karola Schmitz, Leopold Seidler, Christian Selzer, Tim Schramm, Dr. Sebastian Spicher, Dr. Marcel Stahn, and everyone else who I might forgot to mention. Working with such a dedicated and supportive team has been both inspiring and rewarding. I greatly enjoyed our daily visits to the Mensa and coffee breaks, which brought a welcome balance to work. I also appreciate all the social activities, from board game nights and hiking trips to visits to Phantasialand, table tennis matches, and, of course, Tim's unforgettable house parties. We also always had so much fun in my office, so a big thanks goes to my office mates Dr. Sebastian Spicher, Dr. Markus Bursch, Christoph Plett, Abylay Katbashev, Christian Selzer, Andrea Pellegrini, Christopher Staudt, Tim Schramm, Prof. Marcos D. Garcia, and Dr. Iria Díaz Arias. Furthermore, I would like to extend my sincere thanks to Benedikt Bädorf, Johannes Gorges, Tim Schramm, and Lukas Wittmann for their invaluable support in discussing topics and providing thoughtful feedback on my thesis. A special thanks goes to my favorite troublemakers, Benni, Albert, Chris, and Lukas, whose countless shenanigans brightened my days at the institute and made this time truly memorable. I would also like to express my gratitude to Claudia Kronz for her invaluable administrative support and to Jens Mekelburg for always ensuring that everything ran smoothly on the technical side.

Beyond my work group, I am deeply grateful to my friends whom I met during my time as a student, including Catrin Allar, Johannes Gorges, Lars Westhofen, Timo Dräger, Anna Rasgauski, Mia Weiland, and Saskia Tsiaparas, as well as to those outside academia, including Kaddy, Quynh, and Ngoci, as well as my PnP-group with Aline, Schromm, Pascal and Tessa, and of course many others I may not have mentioned by name, whose friendship, encouragement, and unwavering support have meant more to me than words can express. I treasure the memories of our time together, from escape rooms and cooking evenings to enduring the challenges of COVID-19, going out, playing video games, and simply chatting about anything and everything. Thank you!

Finally, I owe everything to my family. Your unconditional love, patience, and unwavering support have been the foundation of all my achievements. After leaving everything behind in Russia and coming to Germany with nothing but the hope of a better life, our journey was marked by many hardships. Yet, you always found a way to turn "rubbish into gold", ensuring that I never felt like I was missing out on anything. I know you made countless sacrifices so that my brothers and I could have a better life and more opportunities than you did, and for that, I am endlessly grateful. I am proud to be the first generation in our family to achieve this academic milestone, something that would not have been possible without your help. This thesis is dedicated to you. Thank you for everything — I love you.

Here is a Russian translation of the last section for my family:

Наконец, я всем обязан своей семье. Ваша безусловная любовь, терпение и непоколебимая поддержка стали прочной основой для всех моих достижений. Оставив всё позади в России и приехав в Германию с одной лишь надеждой на лучшую жизнь, мы прошли через множество трудностей. Но вы всегда находили способ делать невозможное возможным, так, чтобы я никогда ни в чем не нуждался. Я знаю, что вы принесли в жертву многое ради того, чтобы я и мои братья могли жить лучше и иметь больше возможностей, чем вы сами. За это я бесконечно вам благодарен. Я горжусь тем, что стал первым в нашей семье, кто достиг этой учёной степени. Это то, что никогда не случилось бы без вашей помощи. Эта диссертация посвящается вам. Спасибо за всё — я вас люблю.

# Appendix

# APPENDIX A

# Optimization of the r$^2$SCAN-3c Composite Electronic-Structure Method for Use with Slater-Type Orbital Basis Sets

Thomas Gasevic,[†] Julius B. Kleine Büning, Stefan Grimme, and Markus Bursch

*Received: April 28, 2022*
*First published: June 2, 2022*

Reprinted (adapted) with permission from:
T. Gasevic, J. B. Stückrath, S. Grimme, and M. Bursch, *Optimization of the r$^2$SCAN-3c Composite Electronic-Structure Method for Use with Slater-Type Orbital Basis Sets*, J. Phys. Chem. A **126**.23 (2022) 3826, DOI: `10.1021/acs.jpca.2c02951`

**Own contributions**

- Conceptualization
- Data Curation
- Formal Analysis & Investigation
- Software – Implementing r$^2$SCAN-3c(STO) into the Amsterdam Modeling Suite (AMS)
- Visualization
- Writing – Original Draft Preparation, Review & Editing

---

[†] Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany

[‡] Max-Planck-Institut für Kohlenforschung, 45470 Mülheim an der Ruhr, Germany

THE JOURNAL OF PHYSICAL CHEMISTRY A
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCA Article

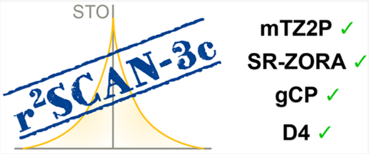
# Optimization of the r$^2$SCAN-3c Composite Electronic-Structure Method for Use with Slater-Type Orbital Basis Sets

Thomas Gasevic, Julius B. Stückrath, Stefan Grimme,* and Markus Bursch*

**Cite This:** *J. Phys. Chem. A* 2022, 126, 3826−3838 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** The "Swiss army knife" composite density functional electronic-structure method r$^2$SCAN-3c (*J. Chem. Phys.* **2021**, *154*, 064103) is extended and optimized for the use with Slater-type orbital basis sets. The meta generalized-gradient approximation (meta-GGA) functional r$^2$SCAN by Furness et al. is combined with a tailor-made polarized triple-ζ Slater-type atomic orbital (STO) basis set (mTZ2P), the semiclassical London dispersion correction (D4), and a geometrical counterpoise (gCP) correction. Relativistic effects are treated explicitly with the scalar-relativistic zeroth-order regular approximation (SR-ZORA). The performance of the new implementation is assessed on eight geometry and 74 energy benchmark sets, including the extensive GMTKN55 database as well as recent sets such as ROST61 and IONPI19. In geometry optimizations, the STO-based r$^2$SCAN-3c is either on par with or more accurate than the hybrid density functional approximation M06-2X-D3(0)/TZP. In energy calculations, the overall accuracy is similar to the original implementation of r$^2$SCAN-3c with Gaussian-type atomic orbitals (GTO), but basic properties, intermolecular noncovalent interactions, and barrier heights are better described with the STO approach, resulting in a lower weighted mean absolute deviation (WTMAD-2(STO) = 7.15 vs 7.50 kcal mol$^{-1}$ with the original method) for the GMTKN55 database. The STO-optimized r$^2$SCAN-3c outperforms many conventional hybrid/QZ approaches in most common applications at a fraction of their cost. The reliable, robust, and accurate r$^2$SCAN-3c implementation with STOs is a promising alternative to the original implementation with GTOs and can be generally used for a broad field of quantum chemical problems.



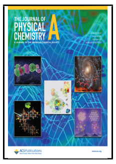
## INTRODUCTION

In recent years, Kohn−Sham density functional theory (DFT)[1] has become one of the most popular methods in quantum chemistry, mainly due to its outstanding accuracy to computational cost ratio.[2] It can be employed for a large number of problems, including molecular structures and various chemical properties, as well as reactions that facilitate research and commercial projects.[3−6] Despite its high efficiency, the limits of conventional DFT are quickly reached for calculations of large systems that contain more than 300 atoms. The emerging need for fast yet accurate low-cost methods paves the way for composite schemes. These typically include small optimized basis sets to reduce the computational cost and compensate for the resulting errors with tailored corrections. A prominent class of such composite schemes is the "3c" method family. The first 3c method was the Hartree−Fock theory-based HF-3c[7] method that contains three name-giving corrections to improve its accuracy. The same concept was later applied to DFT from which the PBEh-3c/HSE-3c[8−10] hybrid and B97-3c[11] GGA functionals resulted.

The latest addition to the "3c" family is r$^2$SCAN-3c,[12] which utilizes a well-balanced triple-ζ Gaussian-type atomic orbital (GTO) basis set, the D4 London dispersion correction,[13,14] and a geometrical counterpoise (gCP)[15] correction for remaining inter- and intramolecular basis set superposition errors (BSSE). The underlying meta-generalized-gradient approximation (meta-GGA)-type density functional r$^2$SCAN[16,17] is the regularized and restored form of the strongly constrained and appropriately normed (SCAN)[18] functional. r$^2$SCAN yields improved accuracy and a much reduced sensitivity to the employed numerical integration grid. Overall, the original GTO-based r$^2$SCAN-3c was shown to yield excellent results for the calculation of thermochemical properties as well as conformational energies for systems with main-group elements and transition metals, partly reaching the accuracy of hybrid functionals, applying basis sets of quadruple-ζ (QZ) quality.[12]

Up to this point, the "3c" composite schemes were limited to GTO basis sets, while an assessment with Slater functions is missing. They satisfy Kato's cusp condition[19] at the nucleus and possess a correct long-range behavior. In this work, we present an optimized Slater-type atomic orbital (STO) variant of the composite r$^2$SCAN-3c DFT method with a customized all-electron STO basis set which was implemented in the Amsterdam Density Functional (ADF) program[20,21] of the Amsterdam Modeling Suite (AMS).[22] The performance of r$^2$SCAN-3c is compared for both implementations (GTO vs STO) and extensively assessed on a comprehensive database

**Received:** April 28, 2022
**Revised:** May 16, 2022
**Published:** June 2, 2022



https://doi.org/10.1021/acs.jpca.2c02951
*J. Phys. Chem. A* 2022, 126, 3826−3838

consisting of 621 data points for geometrical quantities and 4405 data points for energies. This data collection includes extensive benchmark sets such as the GMTKN55[23] database as well as additional benchmark sets for noncovalent interactions (e.g., IONPI19[24]), conformational energies, and organometallic reactions (e.g., MOR41[25] and ROST61[26]). Comparisons between GTOs and STOs have already been made in a different context,[27] but this study is probably one of the most extensive ones considering the variety, chemical relevance, and amount of evaluated data points.

## ■ THEORETICAL METHODS

The composite electronic-structure method r$^2$SCAN-3c consists of five different components, some of which are interdependent. An overview is shown in Figure 1, and modifications of each component will be discussed in the following. The main building block is the underlying density functional approximation (DFA) r$^2$SCAN. In comparison to its predecessors SCAN and rSCAN,[28] it is more accurate and less sensitive to the numerical integration grid. Consequently, much finer integration grids compared to those used in other conventional DFT methods are not required anymore, which leads to faster and more robust computations. In the STO approach, the basic DFA remains unchanged and is implemented via the Libxc library.[29]

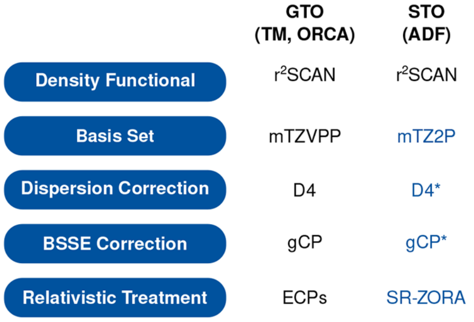


**Figure 1.** Components of r$^2$SCAN-3c applying Gaussian-type atomic orbitals (GTO) and Slater-type atomic orbitals (STO). Changes in the STO approach are marked in blue. The D4 and gCP corrections were adjusted for STOs as indicated by an asterisk.

**Basis Set Modification.** The new modified all-electron triple-ζ basis set mTZ2P includes a combination of the default STO atomic orbital basis sets DZP, TZP, and TZ2P in ADF, which are contracted for the zeroth-order regular approximation (ZORA).[30−32] It is constructed in analogy to the original GTO basis set mTZVPP to which it is compared in Table 1. The contraction schemes do not match for every element due to the different composition of the STO basis set. For example, one d- and one f-polarization function each are used for oxygen instead of two d-functions. In general, the STO basis sets include more basis functions for heavy elements compared to the original GTO basis set as the latter per default applies small-core effective core potentials (ECP) to represent the core electrons. Alterations to the underlying DZP and TZP basis sets by removing or exchanging basis functions (e.g., replacing f- by d-polarization functions) were not successful, typically increasing the obtained errors.

**Table 1. Comparison of the New mTZ2P STO Basis Set with the Original mTZVPP GTO Set[a]**

| | contraction | | underlying |
|---|---|---|---|
| element | mTZVPP | mTZ2P | STO basis |
| H | [2s1p] | [2s1p] | DZP[b] |
| He | [2s1p] | [2s1p] | DZP |
| N | [5s3p2d] | [5s3p2d] | TZ2P |
| O | [5s3p2d] | [5s3p1d1f] | TZ2P[c] |
| F | [5s3p2d] | [5s3p2d] | TZ2P |
| Ne | [5s3p2d] | [5s3p1d1f] | TZ2P |
| Si−S | [5s4p2d] | [7s5p1d1f] | TZ2P |
| Cl | [5s4p2d] | [7s5p1d1f] | TZ2P |
| Ar | [5s4p2d] | [7s5p1d1f] | TZ2P |
| Kr | [6s5p4d] | [8s7p4d1f] | TZ2P |

[a] Elements that are not listed are described by the standard TZP basis in the STO set. [b] 2p exponent is changed from 1.25 to 1.70. [c] 3d exponent is changed from 2.00 to 2.15.

Nevertheless, it was found that the respective 3d- and 2p-polarization functions of oxygen and hydrogen were initially too diffuse (too small exponents) when they are used in combination with the D4 and gCP corrections. The exponents were manually optimized using the WATER27[33] (water clusters), S22,[34] S66[35,36] (noncovalent interactions of small molecules), and HB300SPX[37] (hydrogen bonds) benchmark sets. First, the 3d exponent of oxygen was changed from 2.00 to 2.15 and subsequently the 2p exponent of hydrogen from 1.25 to 1.70. The effect on the WATER27 set is illustrated in Figure 2. Here, the optimization of the 3d exponent of oxygen already improves the results noticeably but the influence of tuning the 2p exponent of hydrogen is more substantial, drastically decreasing the tentative underestimation of the intermolecular water−water interactions.

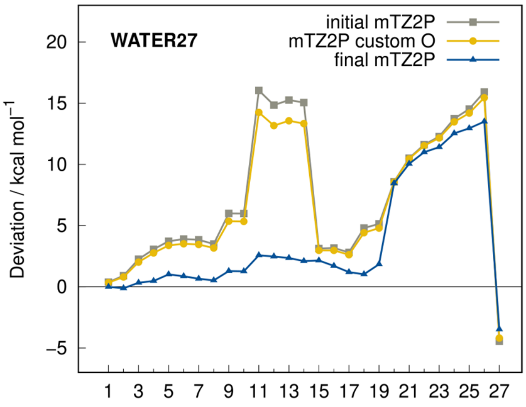


**Figure 2.** Deviations calculated with r$^2$SCAN-3c(STO) for the WATER27 interaction energy benchmark set applying the original exponents for oxygen and hydrogen (initial mTZ2P), customized exponents for oxygen (mTZ2P custom O), and customized exponents for oxygen and hydrogen (final mTZ2P).

**London Dispersion Correction (D4).** Since semilocal density functional approximations do not account for long-range electron correlation effects, they lack the description of London dispersion interactions.[38,39] In r$^2$SCAN-3c, they are included by the atomic-charge dependent London dispersion correction D4, which is calculated according to

$$E_{\text{disp}}^{\text{D4}} = -\frac{1}{2}\sum_{AB}\sum_{n=6,8} s_n \frac{C_{(n)}^{AB}}{R_{AB}^{(n)}} f_{\text{damp}}^{(n)}(R_{AB}) - \frac{1}{6}\sum_{ABC} s_9 \frac{C_{(9)}^{ABC}}{R_{ABC}^{(9)}} f_{\text{damp}}^{(9)}(R_{ABC}, \theta_{ABC}) \qquad (1)$$

where $A$, $B$, and $C$ denote atoms, $s_n$ is the scaling parameter, $C_{(n)}$ is the dispersion coefficient, $R_{AB}$ is the interatomic distance, $R_{ABC}$ is the geometrically averaged distances, $\theta_{ABC}$ is the angle in atomic triangles, and $f_{\text{damp}}^{(n)}$ is the Becke−Johnson damping function $f_{\text{BJ}}^{(n)}$:

$$f_{\text{BJ}}^{(n)}(R_{AB}) = \frac{R_{AB}^{(n)}}{R_{AB}^{(n)} + (a_1 R_0^{AB} + a_2)^{(n)}} \qquad (2)$$

with the functional specific parameters $a_1$ and $a_2$. As described in Basis Set Modification, the STO basis set mTZ2P is generally more diffuse than the GTO basis set mTZVPP and has a different long-range behavior due to the shape of the Slater-type functions. Accordingly, the manifestation of basis set superposition error (BSSE) is different for both basis sets which has a direct influence on the D4 and gCP corrections that have to be adjusted accordingly. The comparison between the GTO and STO variants of r$^2$SCAN-3c in Figure 3 shows that, by applying the same D4 and gCP corrections (original parameters of the GTO variant), the interaction energies differ by up to 1.66 kcal mol$^{-1}$ in the L7 benchmark set for noncovalent interactions of large complexes.[40,41] Here, a slight mismatch of the attractive D4 correction and the repulsive gCP is observed for the STO basis set. To partly correct this issue, the $s_9$ scaling parameter of the three-body dispersion has been set to 1.53 (vs 2.00 in the GTO approach), as it yields the lowest mean absolute deviation for the S30L benchmark set[42] (association energies of large NCI complexes). The remaining parameters $s_6$, $s_8$, $a_1$, and $a_2$, as well as the parameters in the charge-scaling functions $\beta$ and $\gamma$ (see eq 2 of ref 14) are kept unchanged. An overview of the utilized D4 parameters is listed in the Supporting Information. The effect of adjusting the $s_9$ parameter for the L7 set is depicted in Figure 3. In general, the BSSE at the TZ basis set level can be partly absorbed in the D4 parametrization.[43]

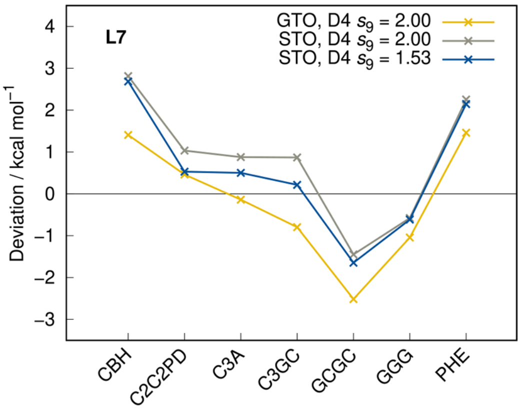


**Figure 3.** Deviations from reference values calculated with r$^2$SCAN-3c for the L7 benchmark set applying different $s_9$ D4 parameters.

**Geometrical Counterpoise Correction (gCP).** Calculations applying finite basis sets are contaminated by inter- as well as intramolecular BSSE. These errors can be corrected with a geometrical counterpoise scheme according to

$$E_{\text{gCP}} = \sigma f_{\text{damp}}^{\text{gCP}}(R_{AB}) \sum_{A}^{\text{atoms}} \sum_{A \neq B}^{\text{atoms}} E_A^{\text{miss}} \frac{\exp(-\alpha (R_{AB})^{\beta})}{\sqrt{S_{AB} N_B^{\text{virt}}}} \qquad (3)$$

where $A$ and $B$ denote atoms, $\sigma$ is a global scaling parameter, $f_{\text{damp}}^{\text{gCP}}$ is a damping function as described in the work on the PBEh-3c[8] method, $\alpha$ and $\beta$ are global fit parameters, and $S_{AB}$ is an s-type Slater overlap integral evaluated with scaled standard valence-average exponents. $E_A^{\text{miss}}$ is originally the atomic energy difference between a large, almost complete basis set and the target basis set, and $N_B^{\text{virt}}$ is the number of virtual orbitals in the target basis set. In r$^2$SCAN-3c, both $E_A^{\text{miss}}$ and $N_B^{\text{virt}}$ are used as additional free fit parameters or are set to unity. In order to correct the remaining BSSE as well as the absorbing part of the (small) basis set incompleteness error (BSIE) in the STO-based r$^2$SCAN-3c, the gCP correction was manually adjusted by optimizing the global scaling parameter $\sigma$ after the basis set and the D4 correction had been modified. The remaining parameters are not altered. For the manual optimization of $\sigma$, mainly the S22, S66, and NCIBLIND10[44] benchmark sets were analyzed to reduce overall deviations, but additional sets were also cross-checked. Figure 4 shows the impact of the gCP correction as well as the different global scaling factors for the WATER27 and ACONF12[12] benchmark sets. Although individual test sets, such as WATER27, benefit from a large scaling factor close to one, the majority is better described with a smaller value (cf. Figure 4b), which indicates that the STO

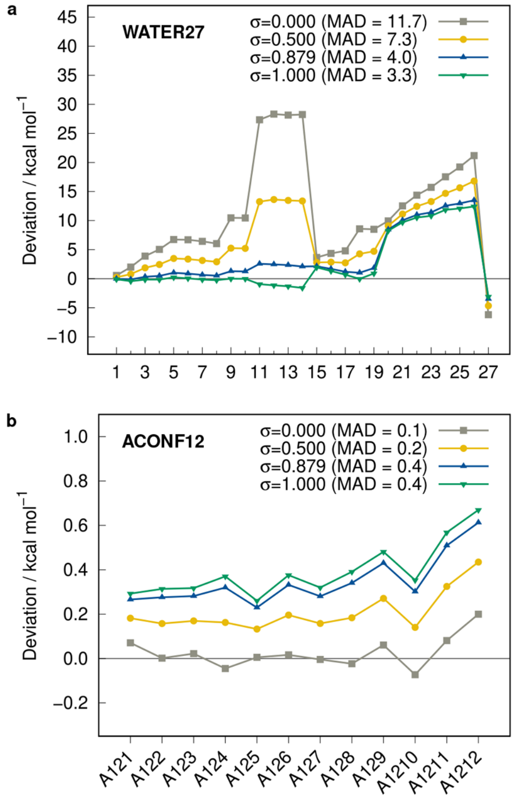


**Figure 4.** Deviations calculated with r$^2$SCAN-3c(STO) for the WATER27 (a) and ACONF12 (b) benchmark sets applying different settings for the global scaling $\sigma$ of the gCP correction. MADs are given in kcal mol$^{-1}$.

basis set mTZ2P is less prone to remaining basis set errors than the GTO basis set mTZVPP. A good balance for all tested benchmark sets is represented by the value $\sigma = 0.879$, which is applied instead of the originally used $\sigma = 1.000$.

**Relativistic Effects.** In standard quantum chemical problems, the nonrelativistic Schrödinger equation is approximately solved to obtain the final wave function. However, relativistic effects can affect the molecular geometry as well as properties, especially when heavy atoms (typically with $Z > 36$) are present.[45−48] They can be included implicitly by relativistic effective core-potentials (ECPs) that replace the core electrons. This approach has the advantage of a lower computational cost and is sufficient for most chemical problems which mainly depend on valence electrons like thermochemisty.[49] Accordingly, the original GTO-based r$^2$SCAN-3c is based on a modified Ahlrichs basis set that is constructed for default use with ECPs.

Relativistic effects can also be incorporated explicitly with the more time-consuming four-component Dirac equation that can be approximated with the zeroth-order regular approximation (ZORA).[30−32] It is based on an expansion of the full relativistic Hamiltonian with respect to a potential-dependent perturbation parameter and contains relativistic corrections at the zeroth order.[32,50] Overall, the accuracy for structures and electronic energies is typically similar to ECPs and ZORA.[51−53] Nevertheless, ZORA is less approximate, and the application of an all-electron (AE) basis set gives the flexibility to also apply explicit spin−orbit relativistic Hamiltonians, which may be crucial for very heavy elements. Further, explicit description of the core electrons can be crucial for a correct description of properties such as NMR chemical shielding tensors. Since scalar-relativistic effects are typically dominant for most applications and ZORA is the default in the ADF program package, it is also applied in the STO-based r$^2$SCAN-3c.

**Grid Study.** Any conventional DFT calculation is typically depending on a sufficiently fine numerical integration grid.[54] And even though r$^2$SCAN is already numerically more robust than its preceding functionals SCAN[18] and rSCAN,[28] the choice of a reasonable grid size is still relevant to obtain reliable results. In ADF, the grid can be controlled by the *NumericalQuality* setting, which simultaneously sets the quality of the *BECKE* integration grid and the quality of the density fitting, termed *ZLMFIT*. To determine a suitable default, different settings were assessed on a test set that includes the ACONF,[55] ACONF12,[12] L7, MOR41,[25] S30L, and S22 benchmark sets. The performance of each *BECKE*/*ZLMFIT* combination as well as a timing comparison is depicted in Figure 5. It was found that the influence of *ZLMFIT* on the accuracy is negligible and that the results mainly depend on the *BECKE* grid, which is almost converged with the *good* setting. This goes along with an increased computational cost compared to the *normal* setting but is necessary in order to make the method robust. For the same reason, we decided to also set the *ZLMFIT* to *good* in all calculations. In general, the *good* setting for the *BECKE* grid leads to a slightly lower number of points on a Lebedev grid compared to the originally used *m4* angular grid in combination with a radial grid size of 10 in the GTO-based TURBOMOLE[56,57] (TM) code (e.g., 460046 (ADF) vs 476590 (TM) for *n*-dodecane).

**Computational Details.** Single-point calculations with r$^2$SCAN-3c(STO) were performed with a development version of the Amsterdam Density Functional program ADF from the Amsterdam Modeling Suite AMS 2021.201 program package.[20,21] Consideration of molecular symmetry was turned off and the *NumericalQuality* was set to *good*. Benchmark sets that require computations of single atoms (AHB21,[58] ALKBDE10,[59] BH76,[60−62] BH76RC,[62] CHB6,[58] DIPCS10,[23] G21EA,[62,63] G21IP,[62,63] HEAVYSB11,[23] PA26,[23,59,62,64] RG18,[23] SIE4x4,[23] W4-11[65]) were calculated with the *IntegerAufbau* option to obtain integer instead of fractional orbital occupations. Scalar-relativistic effects were treated with the zeroth-order regular approximation (ZORA). For the D4 London dispersion correction and the geometrical counterpoise (gCP) scheme, the standalone programs *dftd4 3.3.0*[13,14] and *mctc-gcp*[15] were used.

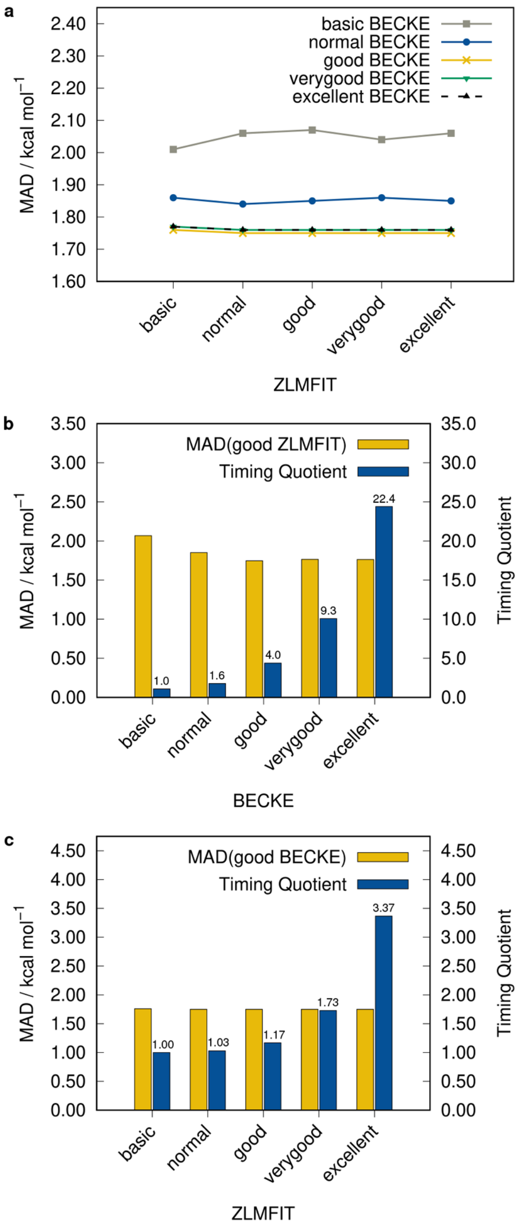


**Figure 5.** Mean absolute deviation (MAD) for the ACONF, ACONF12, L7, MOR41, S30L, and S22 benchmark sets calculated with r$^2$SCAN-3c(STO): Different *BECKE* and *ZLMFIT* settings (a) as well as timing quotients relative to the *BECKE* setting *basic* (b) and relative to the *ZLMFIT* setting *basic* (c).

Further single-point calculations were conducted with PBE,[66] TPSS,[67] SCAN,[18] r$^2$SCAN,[16,17] PBE0,[68,69] and

B3LYP[70,71] in combination with the TZP, TZ2P, or QZ4P basis sets[72] for a timing comparison. Additionally, geometry optimizations and single-point energies were calculated with BP86[73,74]-D4/TZP and M06-2X[75]-D3(0)[76,77]/TZP. The same settings as in the r$^2$SCAN-3c calculations were used in all calculations.

Results for the GMTKN55 benchmark sets calculated with the GTO version of r$^2$SCAN-3c were taken from the GMTKN55 database. The remaining test sets were computed with TURBOMOLE 7.5.1[56,57] using r$^2$SCAN-3c with grid m4 and a radial grid size of 10. The resolution of identity (RI) approximation for the Coulomb energy was used with the same reduced auxiliary basis sets developed originally for B97-3c.[78−80] Default settings were used if not stated otherwise.

## ■ RESULTS AND DISCUSSION

To compare the STO-based composite r$^2$SCAN-3c method with the original GTO-based approach, the performance of both implementations is assessed on eight geometry and 74 energy benchmark sets. The mean deviations (MDs), mean absolute deviations (MADs), standard deviations (SDs), and root-mean-square deviations (RMSDs) for each test set are listed in the Supporting Information.

**Geometries.** The performance for the calculation of covalent bond lengths is assessed on benchmark sets that contain light main-group bonds (LMGB35[81]), heavy main-group bonds (HMGB11[81]), long main-group bond lengths (LB12[81]), transition metal complexes (TMC32[82]), and small semirigid organic molecules (CCse21[83,84]). Rotational constants are evaluated with the ROT34[85,86] test set. The results are summarized in Figures 6 and 7.

In general, the GTO- and STO-based approaches of r$^2$SCAN-3c yield similar results for the molecular geometries. Both gravitate toward slightly too long bonds for light and heavy main group bonds and toward too short bonds for transition metal complexes. In general, the error spread of r$^2$SCAN-3c(STO) tends to be slightly larger than that of the GTO-based method, as described by the standard deviation of each test set, but the difference is small and the mean absolute deviation is lower with the STO variant. Overall, it is either on par with or outperforms the commonly used GGA method BP86-D4/TZP as well as the meta-hybrid method M06-2X-D3(0)/TZP, which performs well for small organic molecules. In calculations of bond angles and rotational constants, r$^2$SCAN-3c(STO) yields smaller errors than the GTO-based composite method (about 14% difference in the MAD for both cases).

To test noncovalent bonds, center-of-mass distances ($R_{\mathrm{CMA}}$) for the noncovalent interaction (NCI) benchmark sets S66x8[35,36,87] and HB300SPXx10[37] were calculated via a six-point cubic spline-interpolation ({0.90, 0.95, 1.00, 1.05, 1.10, 1.25}$R_e$) of rigid fragment potential energy curves. The results are evaluated with respect to CCSD(T)/CBS and counterpoise-corrected MP2-F12/V{T,Q}Z-F12 reference values, respectively. The MADs and SDs are shown in Figure 7.

In the S66 test set, which contains organic van der Waals and hydrogen-bonded systems, both implementations of r$^2$SCAN-3c perform equally well with very similar statistical measures. For example, both yield too long noncovalent bonds with a MD of 5.3 pm for the STO-based composite method and 5.4 pm for the GTO-based method. Im comparison, BP86-D4/TZP and M06-2X-D3(0)/TZP yield systematically too short NCI contacts with a MD of −2.5 and −8.6 pm, respectively. The performance of both variants of r$^2$SCAN-3c is remarkable, as their SD is about 1 pm smaller than that of BP86/TZP and M06-2X and the MAD is about 2.9 pm smaller than that of M06-2X-D3(0)/TZP.

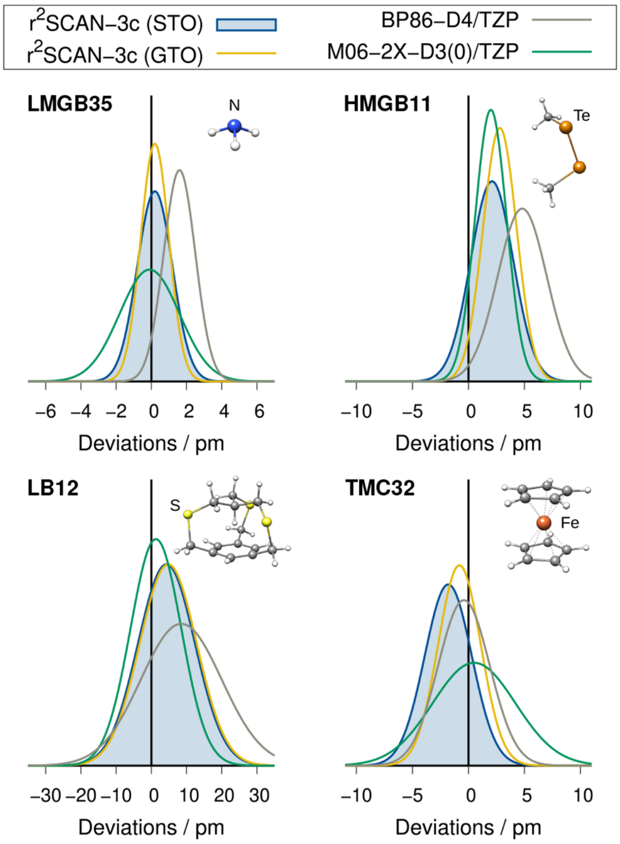


Figure 6. Gaussian error distributions for a selection of covalent bond length benchmark sets. Negative mean deviations indicate overall too short bond lengths.

For the hydrogen-bonded systems in HB300SPX, r$^2$SCAN-3c(STO) yields smaller deviations than the GTO-based method, and the respective MAD of 7.3 pm is halved compared to BP86/TZP, representing the smallest value in this study. Here, both versions of r$^2$SCAN-3c yield slightly too short H-bonds with a MD of −1.5 pm for the STO variant and −0.8 pm for the GTO variant. These values are rather small compared to that of BP86-D4/TZP (MD = −12.9 pm) and M06-2X (MD = −6.6 pm), which both drastically underestimate H-bond lengths.

Overall, r$^2$SCAN-3c(STO) yields very similar results as the GTO-based method and is, in most cases, even slightly more accurate. It reaches the accuracy of computationally much more demanding hybrid/TZ approaches and can therefore be recommended for geometry optimizations.

**Relative Energies.** The study on relative energies includes 4405 data points in a range between −363.0 and 1290.7 kcal mol$^{-1}$ with a mean reaction energy of 19.8 kcal mol$^{-1}$ covering a broad area of the chemical space with tests for thermochemistry, reaction barriers, noncovalent interactions (NCIs), and conformational energies.

**Main-Group Thermochemistry and Reaction Barriers.** The extensive GMTKN55 database contains 55 versatile benchmark sets with CCSD(T)/CBS reference data for main-group thermochemistry, kinetics, and noncovalent interactions

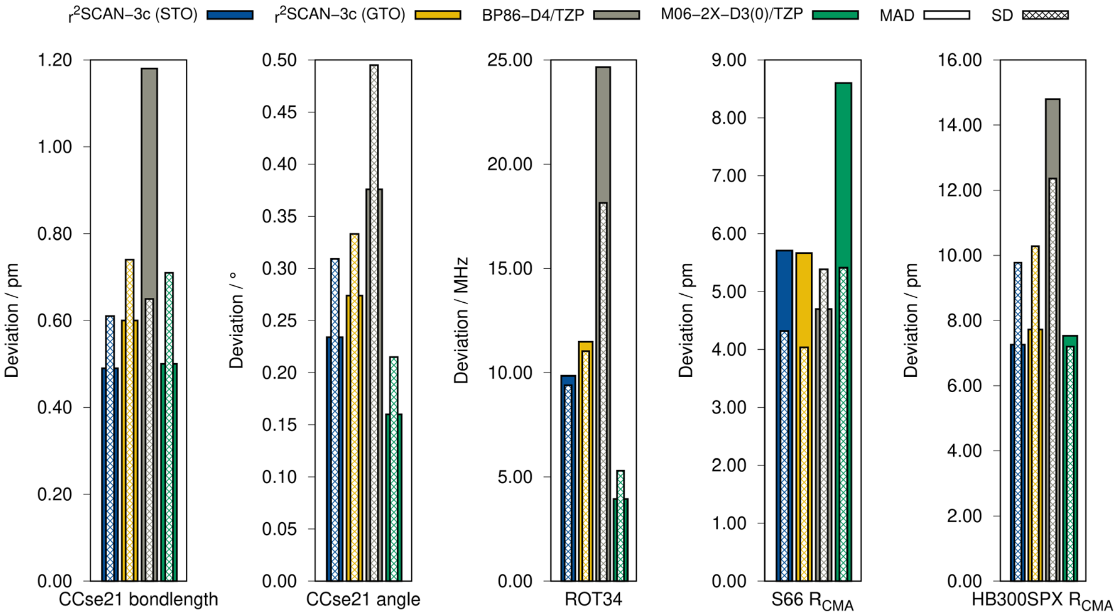


**Figure 7.** Mean absolute deviation (MAD) and standard deviation (SD) of different geometry benchmark sets calculated with both variants of r$^2$SCAN-3c, as well as BP86-D4/TZP and M06-2X-D3(0)/TZP.

and represents an ideal base to validate density functional approximations. It consists of 1505 data points that can be categorized in the five subsets basic properties and reaction energies for small systems (basic properties), reaction energies for large systems (reactions), reaction barrier heights (barriers), and inter- as well as intramolecular NCIs. Because the average energies between the test sets vary significantly, the standard weighted MAD (WTMAD-2; see Supporting Information) is taken as a statistical performance measure.[23] The results are shown in Figure 8 and Table 2.

The STO-based r$^2$SCAN-3c composite method surpasses the accuracy of the parent functional r$^2$SCAN-D4 in nearly all categories, independent of the basis set size. This is especially noticeable in the WTMAD-2 of the entire GMTKN55 database, where the STO version of r$^2$SCAN-3c (7.15 kcal mol$^{-1}$) yields the best results, followed by r$^2$SCAN-D4/QZ4P (7.41 kcal mol$^{-1}$), with TZ2P (7.97 kcal mol$^{-1}$) and TZP (8.45 kcal mol$^{-1}$).

**Table 2. Weighted Mean Absolute Deviation (WTMAD-2) of the Entire GTMKN55, As Well As Its Subclasses Computed with r$^2$SCAN-D4 and Both Variants of r$^2$SCAN-3c**

| | TZP | TZ2P | QZ4P | 3c(STO) | 3c(GTO) |
|---|---|---|---|---|---|
| entire GMTKN55 | 8.45 | 7.97 | 7.41 | 7.15 | 7.50 |
| basic properties | 5.36 | 5.22 | 5.10 | 5.19 | 6.40 |
| reactions | 7.28 | 7.86 | 7.87 | 7.23 | 6.89 |
| barriers | 15.66 | 15.10 | 14.51 | 13.07 | 14.15 |
| intermol. NCIs | 9.84 | 9.03 | 7.33 | 7.53 | 7.22 |
| intramol. NCIs | 8.20 | 6.65 | 6.16 | 5.96 | 5.67 |

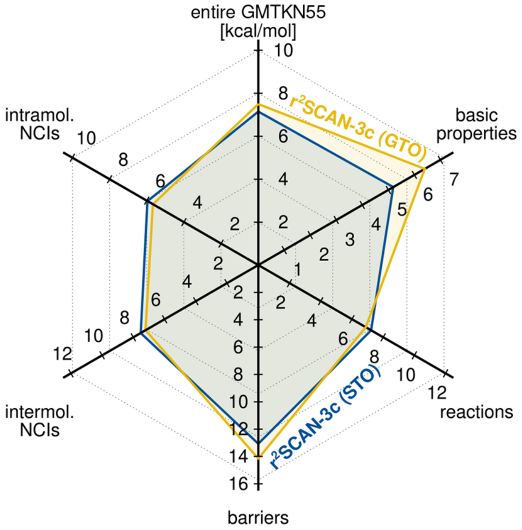


**Figure 8.** Weighted mean absolute deviation (WTMAD-2) of the GMTKN55 as well as its subclasses computed with both variants of r$^2$SCAN-3c.

The WTMAD-2 of r$^2$SCAN-3c(STO) for the entire database is also lower than that of the original GTO-based r$^2$SCAN-3c (7.50 kcal mol$^{-1}$). Notably, this performance approaches that of hybrid DFAs with large aug-def2-QZVP AO basis sets such as B3LYP-D4 (6.5 kcal mol$^{-1}$) and PW6B95-D4 (5.5 kcal mol$^{-1}$)[12] with a drastically reduced computational cost. The overall accuracy of both r$^2$SCAN-3c implementations is similar, but the GTO-based approach yields slightly more accurate results for reactions and intramolecular NCIs, while the STO-based approach yields better results for basic properties and barrier heights. This behavior is also depicted in Figure 9, which shows the difference in the MADs and SDs of both r$^2$SCAN-3c implementations. In this comparison, every positive value represents a better description by the GTO version and every negative value a better description by the STO version of r$^2$SCAN-3c. The good performance of r$^2$SCAN-3c(STO) for the basic properties presumably stems from the larger basis set that seemingly reduces the self-interaction error (SIE), as is also observed in the SIE4x4 test set.

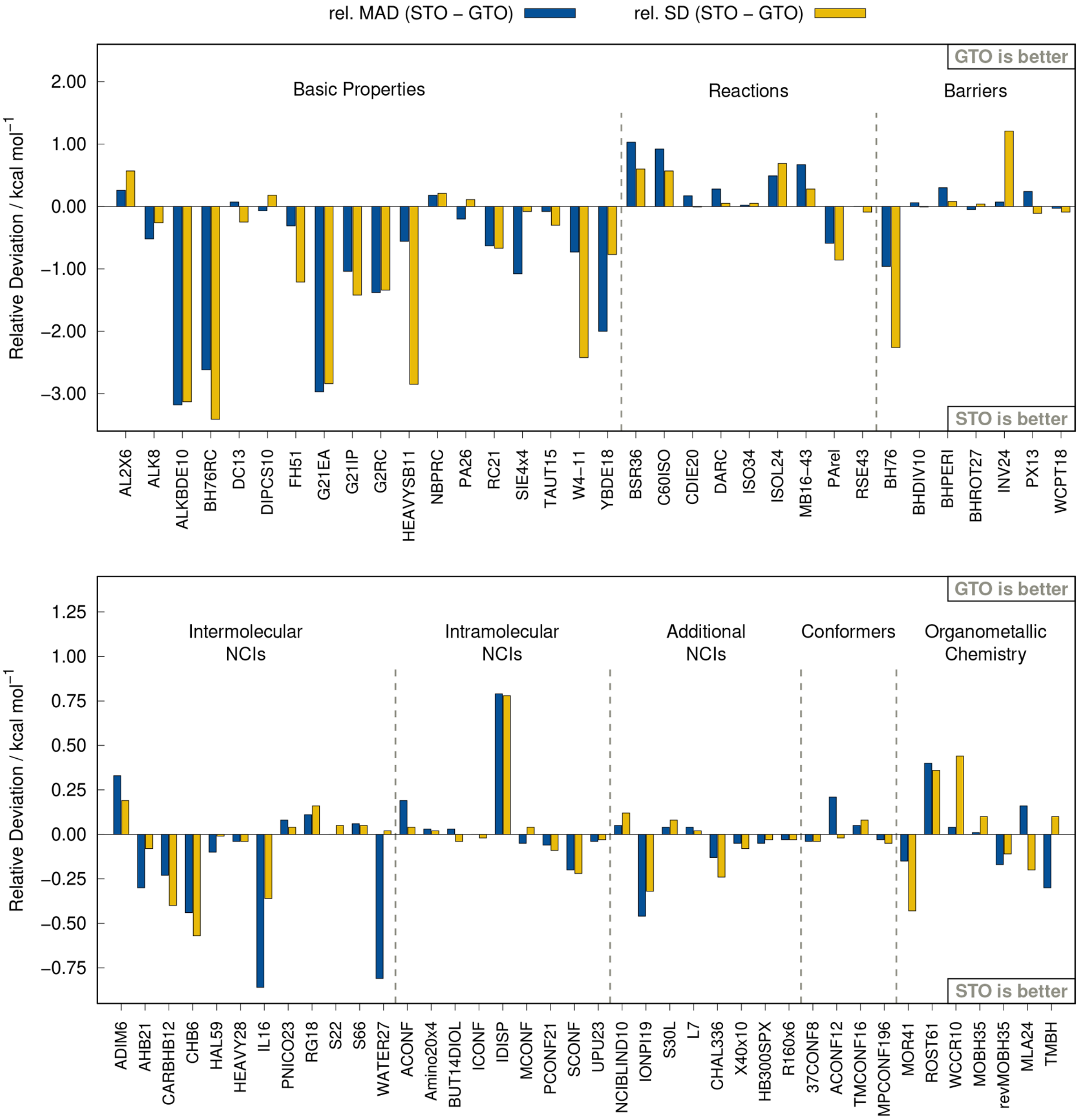


**Figure 9.** Relative mean absolute deviation (MAD) and standard deviation (SD) of the STO version of r$^2$SCAN-3c with respect to r$^2$SCAN-3c (GTO) calculated on the GMTKN55 and several other benchmark sets.

**Noncovalent Interactions.** In addition to the GMTKN55 benchmark collection, more recent benchmark sets on noncovalent interactions are assessed in this section. Among them are large complexes (S30L, L7), various chalcogen (CHAL336[88]), halogen (X40x10[89]), and hydrogen bonding sets (HB300SPX), ion−$\pi$ interactions (IONPI19), a blind test for DFT-based methods (NCIBLIND10), and repulsive intermolecular contacts (R160x6[12,90,91]). For the L7 set, average values of the respective LNO-CCSD(T) and fixed-node diffusion Monte Carlo (FN-DMC) interaction energies published by Al-Hamdani et al.[40] were used as reference.

Similar to the statistics for the GMTKN55, the MAD values for the additional NCI benchmark sets in Figure 10 show that r$^2$SCAN-3c(STO) is either on par with r$^2$SCAN-D4/QZ4P or even more accurate. In particular, in the S30L benchmark for association energies of realistic host−guest complexes, r$^2$SCAN-3c stands out as the MAD is 56% lower than that of r$^2$SCAN-D4/QZ4P. The dominant contribution to the interaction energies is London dispersion, which might indicate that the parametrization of the D4 correction in r$^2$SCAN-D4 is not optimal for STOs. Unexpectedly, it is also observed that the largest tested basis set, QZ4P, does not always yield more accurate results than the smaller basis sets. Compared to the original GTO-based r$^2$SCAN-3c, the STO approach yields overall similar results (cf. Figure 9). The largest difference is observed in the IONPI19 benchmark set, which is shown in Figure 11. Here, the STO variant of r$^2$SCAN-3c yields a 35% lower MAD of 0.83 kcal mol$^{-1}$.

**Conformational Energies.** The conformations of a molecule have a direct influence on chemical properties.[92,93]

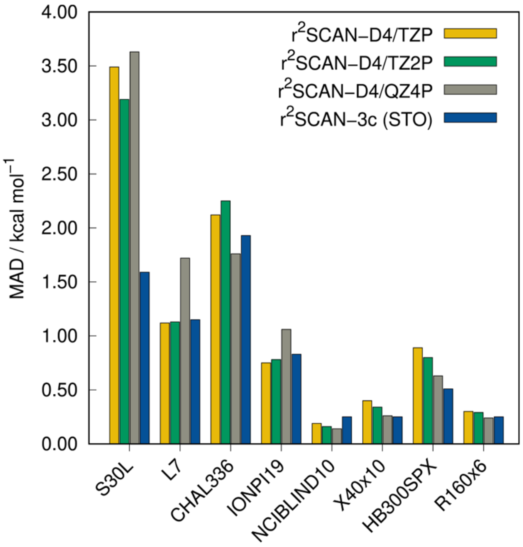


**Figure 10.** MADs of the additional NCI benchmark sets calculated with r$^2$SCAN-3c(STO) and r$^2$SCAN-D4 in combination with different STO basis sets.

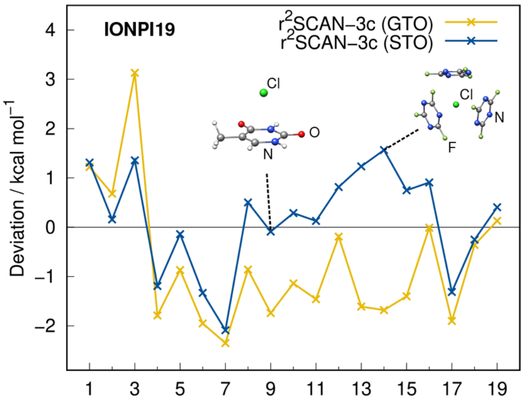


**Figure 11.** Deviations of the IONPI19 benchmark set calculated with both versions of r$^2$SCAN-3c.

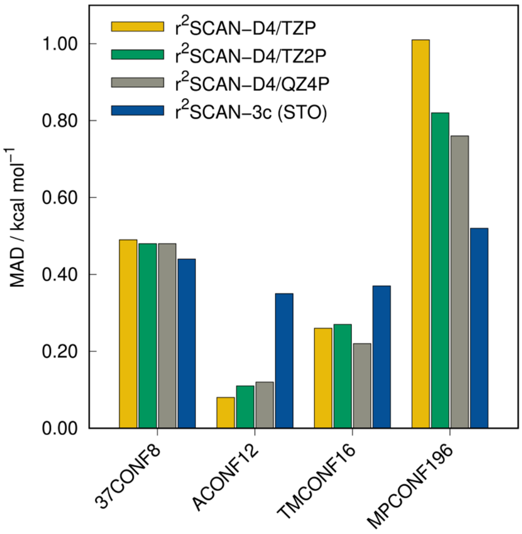


**Figure 12.** MADs of r$^2$SCAN-3c(STO) and r$^2$SCAN-D4 in combination with different STO basis sets for conformational energy benchmark sets.

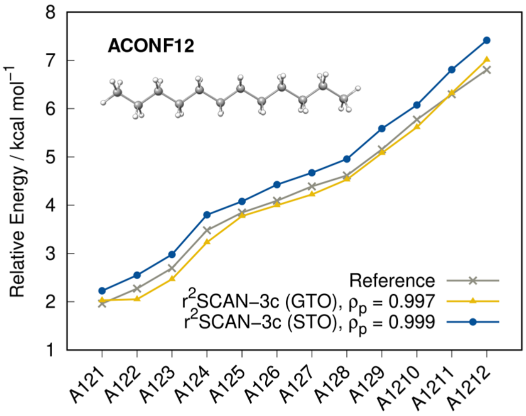


**Figure 13.** Conformational energies calculated with both versions of r$^2$SCAN-3c for the ACONF12 benchmark set as well as the Pearson coefficients $\rho_p$. The Spearman correlation coefficient is $\rho_s = 1$ for both variants of r$^2$SCAN-3c. The reference was calculated at DLPNO−CCSD(T1)/*VeryTightPNO*/CBS level of theory.[12]

Therefore, it can be crucial to consider a conformer ensemble that is routinely created with methods that apply semiempirical methods, for example, with the CREST algorithm.[94] They still require a subsequent higher-level energy ranking for which DFT is usually employed.[95] One of the remarkable features of the original r$^2$SCAN-3c implementation is the very good performance for conformational energies where it surpasses the accuracy of hybrid-DFT/QZ approaches at a considerably lower computational cost.[12] Thus, the implementation with STOs should ideally perform similarly.

In addition to the eight conformer tests sets of the GMTKN55 database, the ACONF12 set with long alkane chains, TMCONF16[96] with transition metal complexes, and MPCONF196,[97] as well as 37CONF8,[98] with large molecules are evaluated in this section. The TMCONF16 set is essentially the TMCONF5 benchmark set without the AYISEG system. The results are depicted in Figures 12 and 13.

For both large molecule test sets (37CONF8 and MPCONF196), r$^2$SCAN-3c(STO) yields slightly better results than r$^2$SCAN-D4, similar to the findings in the previous sections. However, it yields larger deviations for alkane chains and transition metal complexes. These deviations are also observed in the comparison between the results obtained with both implementations (Figure 9). Nevertheless, the error is still small and practically negligible. The conformational energies in the ACONF12 test set are depicted in Figure 13. While the MAD value of 0.14 kcal mol$^{-1}$) with the GTO approach is lower than with the STOs (MAD = 0.35 kcal mol$^{-1}$), the relative energy ranking is better described by r$^2$SCAN-3c(STO), which can be derived from the better Pearson correlation coefficient ($\rho_p$ = 0.999).

**Organometallic Thermochemistry.** As the GMTKN55 database does not include any transition metal complexes, additional test sets are evaluated in this section. Reaction energies of closed-shell complexes are considered with the MOR41 and WCCR10[99,100] benchmark sets, as well as open-shell systems in the ROST61[26] set. Transition metal barrier heights are tested on the TMBH[101−104] benchmark set, which contains 34 barrier heights and on the revised MOBH35 benchmark set, termed revMOBH35.[105] The original MOBH35[106,107] set is also included to provide comparability to prior works. Binding energies of metal-linked alkyl chains are assessed on the MLA24[108] benchmark set. The results are depicted in Figures 9 and 14.

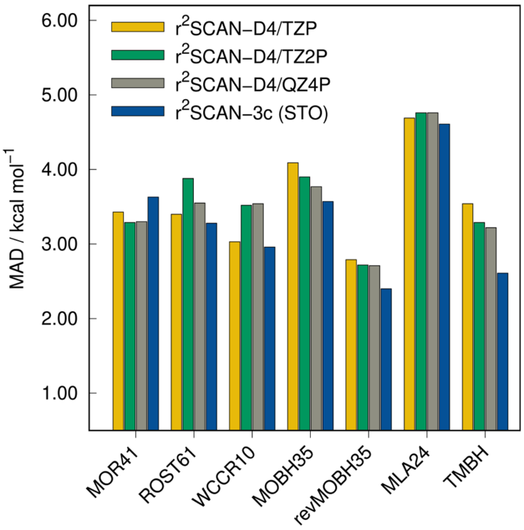


**Figure 14.** MADs of r$^2$SCAN-3c(STO) and r$^2$SCAN-D4 in combination with different STO basis sets for organometallic thermochemistry benchmark sets.

In all tests except MOR41, r$^2$SCAN-3c(STO) outperforms r$^2$SCAN-D4, independent of the applied STO basis set. However, the differences between the DFAs are rather small considering the range of energies included. The largest difference is found for reaction barrier heights in the TMBH set where the MAD of r$^2$SCAN-3c(STO) (MAD = 2.61 kcal mol$^{-1}$) is 19% lower than the MAD of r$^2$SCAN-D4/QZ4P (MAD = 2.91 kcal mol$^{-1}$). Also, the deviation between both implementations of r$^2$SCAN-3c is fairly small. While r$^2$SCAN-3c(GTO) yields lower errors for open-shell systems (ROST61), the STO-based method yields lower errors for closed-shell systems (MOR41). The analysis for the MOR41 benchmark set in Figure 15 reveals that the STO-based approach is more accurate in this test set due to a better description of systems with $\pi$-interactions, which is in line with the findings for the noncovalent interaction benchmark set IONPI19 (cf. Figure 11).

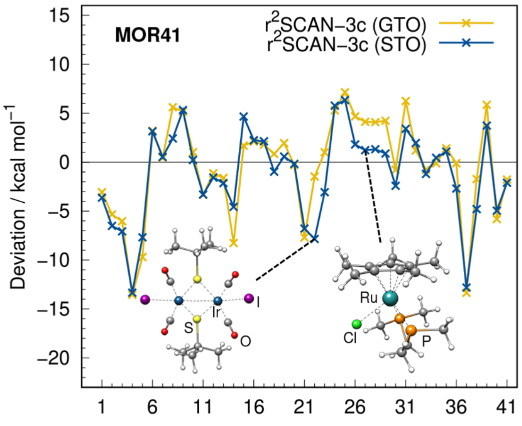


**Figure 15.** Deviations of the MOR41 benchmark set calculated with both versions of r$^2$SCAN-3c.

**Computation Time.** In this section, the timings for single-point energy calculations within the ADF code are assessed on a small test set that includes eight data points that were taken from the MOR41 (**13**, **40**), ROST61 (**R31**, **R33**), S30L (**9**, **19**), and L7 (**GGG**, **C2C2PD**) benchmark sets. The performance of all tested STO-based DFAs is depicted in Figure 16.

In this comparison, r$^2$SCAN-3c(STO) is the most efficient DFA. It is faster than PBE0/TZ2P by a factor of 2.51 and faster than PBE0/QZ4P by a factor of 10.65. Surprisingly, the timing difference between meta-GGA and hybrid DFAs is not as large as in common GTO-based codes such as ORCA[109] and TURBOMOLE,[56,57] where hybrid DFAs are by a factor of about 15−20 slower than meta-GGA DFAs. In the STO-based ADF code, r$^2$SCAN/TZ2P is only 1.3 times faster than PBE0/TZ2P, although the latter requires the additional computation of Fock exchange. This might be an effect of the Libxc implementation of r$^2$SCAN(STO), which may slow down the computation. For example, the GGA PBE is about twice as fast in the native implementation compared to the Libxc variant (cf. Supporting Information).

To compare the modified basis set of r$^2$SCAN-3c with the underlying TZP and TZ2P basis sets, we also tested the computation time on water clusters of different sizes. The results are depicted in Figure 17. Overall, the STO basis set applied in r$^2$SCAN-3c lead to similar computation times as the TZP basis set and the mTZ2P basis set is about twice as fast as the TZ2P basis set.

## ■ CONCLUSION

In this work, we presented the Slater-type atomic orbital basis set optimized variant of the composite electronic-structure method r$^2$SCAN-3c. It combines the meta-generalized-gradient-approximation density functional r$^2$SCAN with a tailored triple-$\zeta$ all-electron STO basis set and applies the readjusted semiclassical D4 and gCP corrections for London dispersion effects and basis set superposition errors, respectively. Instead of the originally applied effective core potentials in the GTO approach, relativistic effects are treated explicitly with the scalar-relativistic zeroth-order regular approximation (SR-ZORA), keeping the flexibility to also apply spin−orbit relativistic ZORA. For robust and accurate results, the *NumericalQuality* should be generally set to the *good* level in the ADF code.

In this comprehensive study, the performance of r$^2$SCAN-3c was assessed on a collection of 82 benchmark sets that cover geometries, thermochemistry, barrier heights, noncovalent interactions, and conformational energies of main-group as well as transition metal systems. In total, 621 data points for geometrical properties and 4405 data points for energies were evaluated for both implementations of r$^2$SCAN-3c. In the geometry study, r$^2$SCAN-3c(STO) has proven to be on par with or better than the hybrid M06-2X-D3(0)/TZP approach. In the energy study, r$^2$SCAN-3c was further compared to r$^2$SCAN-D4 in combination with different sizes of STO basis sets. It was shown that r$^2$SCAN-3c(STO), in most cases, provides more accurate results than r$^2$SCAN-D4/QZ4P at a 6-fold speed-up. The most significant improvement over the large basis set was found for noncovalent interactions of large systems (S30L) where the MADs are 1.59 and 3.63 kcal mol$^{-1}$, respectively. On average, r$^2$SCAN-3c(STO) yields similar

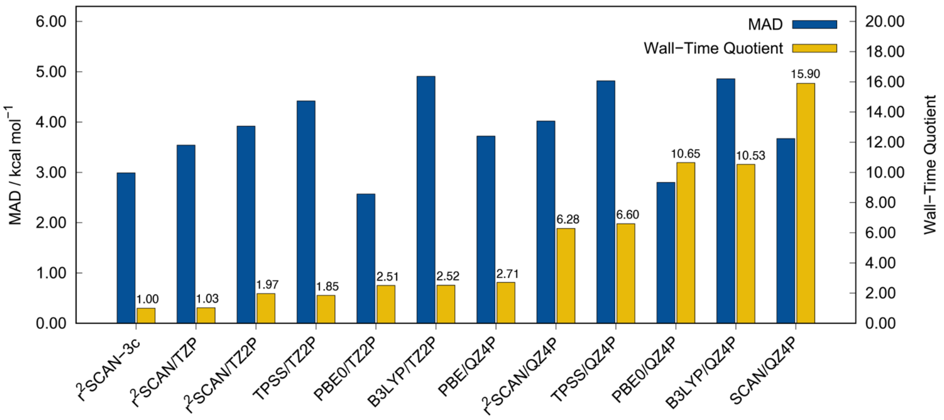


**Figure 16.** MADs and wall-time quotients relative to STO-type r$^2$SCAN-3c (1.0 = 38691 s) in a single-point energy calculation for eight structures from the S30L, MOR41, L7, and ROST61 benchmark sets. All calculations were carried out with ADF and include the D4 correction. Note that the wall-times for the semiclassical D4 and gCP corrections are negligible at 0.35 and 0.15 s, respectively. Computations were done on four Intel Xeon CPU E3-1270 v5@3.60 GHz cores.

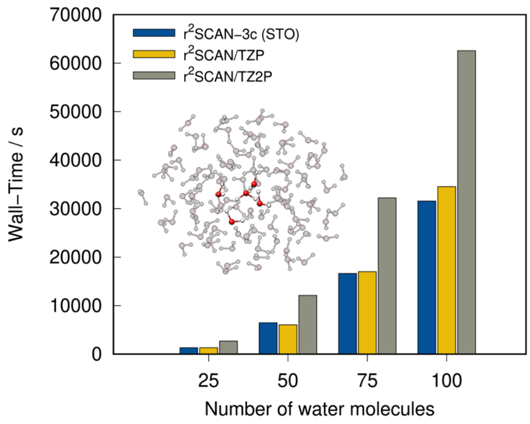


**Figure 17.** Wall-time for a single-point calculation of water clusters with different sizes applying r$^2$SCAN-3c(STO) as well as r$^2$SCAN-D4 in combination with TZP and TZ2P. All computations were done on four cores with the same CPU as in Figure 16.

results as the original GTO version which was also observed for geometrical properties. Reaction energies and intramolecular NCIs, such as conformational energies, are slightly better described with the GTO approach but basic properties and intermolecular NCIs, such as ion-$\pi$ interactions, are better described with the STO approach. This results in a lower WTMAD-2 for the GMTKN55 database with the STO version (WTMAD-2 = 7.15 kcal mol$^{-1}$) instead of the GTO version (WTMAD-2 = 7.50 kcal mol$^{-1}$). Overall, r$^2$SCAN-3c reaches the accuracy of hybrid DFAs, which apply quadruple-$\zeta$ AO basis sets at a significantly reduced computational cost.

The fast, robust, and accurate STO-based r$^2$SCAN-3c method can be applied safely for a broad range of quantum chemical problems and therefore represents an efficient choice in many chemical applications.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.2c02951.

D4 parameters for both variants of r$^2$SCAN-3c, statistical measures for the benchmark sets, and timings (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Stefan Grimme** − *Mulliken Center for Theoretical Chemistry, Universität Bonn, D-53115 Bonn, Germany;* orcid.org/0000-0002-5844-4371; Email: grimme@thch.uni-bonn.de

**Markus Bursch** − *Max-Planck-Institut für Kohlenforschung, D-45470 Mülheim an der Ruhr, Germany;* orcid.org/0000-0001-6711-5804; Email: bursch@kofo.mpg.de

### Authors

**Thomas Gasevic** − *Mulliken Center for Theoretical Chemistry, Universität Bonn, D-53115 Bonn, Germany;* orcid.org/0000-0003-4864-1758

**Julius B. Stückrath** − *Mulliken Center for Theoretical Chemistry, Universität Bonn, D-53115 Bonn, Germany;* orcid.org/0000-0001-7232-4897

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.2c02951

### Funding

Open access funded by Max Planck Society.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

S.G. and M.B. gratefully acknowledge financial support by the Max Planck Society through the Max Planck fellow program. J.B.S. is most grateful for financial support by the "Fonds der Chemischen Industrie (FCI)". We further thank Erik van Lenthe for sharing his expertise and steady support with the AMS code. Sebastian Ehlert and Hagen Neugebauer are acknowledged for fruitful discussions and support with the *dftd4* code.

## ■ REFERENCES

(1) Kohn, W. Nobel lecture: Electronic structure of matter - Wave functions and density functional. *Rev. Mod. Phys.* **1999**, *71*, 1253−1266.

(2) Burke, K. Perspective on density functional theory. *J. Chem. Phys.* **2012**, *136*, 150901.

(3) Becke, A. D. Perspective: Fifty years of density-functional theory in chemical physics. *J. Chem. Phys.* **2014**, *140*, 18A301.

(4) Jones, R. O. Density functional theory: Its origins, rise to prominence, and future. *Rev. Mod. Phys.* **2015**, *87*, 897−923.

(5) Jain, A.; Shin, Y.; Persson, K. A. Computational predictions of energy materials using density functional theory. *Nat. Rev. Mater.* **2016**, *1*, 15004.

(6) Hu, W.; Chen, M. Editorial: Advances in Density Functional Theory and Beyond for Computational Chemistry. *Front. Chem.* **2021**, *9*, 536.

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

(8) 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.

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

(10) 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.

(11) 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.

(12) Grimme, S.; Hansen, A.; Ehlert, S.; Mewes, J. M. r$^2$SCAN-3c: A "Swiss army knife" composite electronic-structure method. *J. Chem. Phys.* **2021**, *154*, 64103.

(13) Caldeweyher, E.; Bannwarth, C.; Grimme, S. Extension of the D3 dispersion coefficient model. *J. Chem. Phys.* **2017**, *147*, 34112.

(14) 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.

(15) 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*, 154101.

(16) Furness, J. W.; Kaplan, A. D.; Ning, J.; Perdew, J. P.; Sun, J. Accurate and Numerically Efficient r$^2$SCAN Meta-Generalized Gradient Approximation. *J. Phys. Chem. Lett.* **2020**, *11*, 8208−8215.

(17) Furness, J. W.; Kaplan, A. D.; Ning, J.; Perdew, J. P.; Sun, J. Correction to: Accurate and Numerically Efficient r$^2$SCAN Meta-Generalized Gradient Approximation. *J. Phys. Chem. Lett.* **2020**, *11*, 9248.

(18) Sun, J.; Ruzsinszky, A.; Perdew, J. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.* **2015**, *115*, 036402.

(19) Kato, T. On the eigenfunctions of many-particle systems in quantum mechanics. *Commun. Pure Appl. Math.* **1957**, *10*, 151−177.

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

(21) Baerends, E. J.; et al. ADF 2020, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, 2020, http://www.scm.com.

(22) Rüger, R.; Franchini, M.; Trnka, T.; Yakovlev, A.; van Lenthe, E.; Philipsen, P.; van Vuren, T.; Klumpers, B.; Soini, T. AMS 2021.1, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, http://www.scm.com.

(23) 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.

(24) Spicher, S.; Caldeweyher, E.; Hansen, A.; Grimme, S. Benchmarking London dispersion corrected density functional theory for noncovalent ion-$\pi$ interactions. *Phys. Chem. Chem. Phys.* **2021**, *23*, 11635−11648.

(25) 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.

(26) 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.

(27) Förster, A.; Visscher, L. Double hybrid DFT calculations with Slater type orbitals. *J. Comput. Chem.* **2020**, *41*, 1660−1684.

(28) Bartók, A. P.; Yates, J. R. Regularized SCAN functional. *J. Chem. Phys.* **2019**, *150*, 161101.

(29) Lehtola, S.; Steigemann, C.; Oliveira, M. J.; Marques, M. A. Recent developments in LIBXC − A comprehensive library of functionals for density functional theory. *SoftwareX* **2018**, *7*, 1−5.

(30) Chang, C.; Pelissier, M.; Durand, P. Regular two-component pauli-like effective hamiltonians in dirac theory. *Phys. Scr.* **1986**, *34*, 394−404.

(31) van Lenthe, E.; Baerends, E. J.; Snijders, J. G. Relativistic regular two-component Hamiltonians. *J. Chem. Phys.* **1993**, *99*, 4597−4610.

(32) Van Lenthe, E.; Baerends, E. J.; Snijders, J. G. Relativistic total energy using regular approximations. *J. Chem. Phys.* **1994**, *101*, 9783−9792.

(33) Bryantsev, V. S.; Diallo, M. S.; Van Duin, A. C.; 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.

(34) Jurečka, P.; Šponer, J.; Černý, 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.

(35) Řezáč, J.; Riley, K. E.; Hobza, P. Erratum to "S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2014**, *10*, 1359−1360.

(36) 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*, 2427−2438.

(37) Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets 2: Hydrogen Bonding in an Extended Chemical Space. *J. Chem. Theory Comput.* **2020**, *16*, 6305−6316.

(38) Grimme, S. Density functional theory with London dispersion corrections. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 211−228.

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

(40) Al-Hamdani, Y. S.; Nagy, P. R.; Zen, A.; Barton, D.; Kállay, M.; Brandenburg, J. G.; Tkatchenko, A. Interactions between large molecules pose a puzzle for reference quantum mechanical methods. *Nat. Commun.* **2021**, *12*, 1−12.

(41) Sedlak, R.; Janowski, T.; Pitoňák, M.; Řezáč, J.; Pulay, P.; Hobza, P. Accuracy of quantum chemical methods for large noncovalent complexes. *J. Chem. Theory Comput.* **2013**, *9*, 3364−3374.

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

(43) Bursch, M.; Neugebauer, H.; Ehlert, S.; Grimme, S. Dispersion Corrected r$^2$SCAN Based Global Hybrid Functionals: r$^2$SCANh, r$^2$SCAN0, and r$^2$SCAN50. *J. Chem. Phys.* **2022**, *156*, 10−12.

(44) Taylor, D. E.; et al. Blind test of density-functional-based methods on intermolecular interaction energies. *J. Chem. Phys.* **2016**, *145*, 124105.

(45) Pyykkö, P. Relativistic Effects in Structural Chemistry. *Chem. Rev.* **1988**, *88*, 563−594.

(46) Bursch, M.; Gasevic, T.; Stückrath, J. B.; Grimme, S. Comprehensive Benchmark Study on the Calculation of 29Si NMR Chemical Shifts. *Inorg. Chem.* **2021**, *60*, 272−285.

https://doi.org/10.1021/acs.jpca.2c02951
*J. Phys. Chem. A* 2022, 126, 3826−3838

(47) Stückrath, J. B.; Gasevic, T.; Bursch, M.; Grimme, S. Benchmark Study on the Calculation of $^{119}$Sn NMR Chemical Shifts. *Inorg. Chem.* **2022**, *61*, 3903−3917.
(48) Vı́cha, J.; Novotný, J.; Komorovsky, S.; Straka, M.; Kaupp, M.; Marek, R. Relativistic Heavy-Neighbor-Atom Effects on NMR Shifts: Concepts and Trends across the Periodic Table. *Chem. Rev.* **2020**, *120*, 7065−7103.
(49) Dohn, A.; Moller, K.; Sauer, S. Optimizing the Structure of Tetracyanoplatinate(II): A Comparison of Relativistic Density Functional Theory Methods. *Curr. Inorg. Chem.* **2014**, *3*, 213−219.
(50) van Lenthe, E.; Baerends, E. J.; Snijders, J. G. Relativistic regular two-component Hamiltonians. *J. Chem. Phys.* **1993**, *99*, 4597−4610.
(51) Shamov, G. A.; Schreckenbach, G. Density functional studies of actinyl aquo complexes studied using small-core effective core potentials and a scalar four-component relativistic method. *J. Phys. Chem. A* **2005**, *109*, 10961−10974.
(52) Shamov, G. A.; Schreckenbach, G. Density functional studies of actinyl aquo complexes studied using smallcore effective core potentials and a scalar four-component relativistic method. *J. Phys. Chem. A* **2006**, *110*, 12072.
(53) Berger, R. J. F.; Schoiber, J.; Monkowius, U. A Relativity Enhanced, Medium-Strong Au(I)···H−N Hydrogen Bond in a Protonated Phenylpyridine-Gold(I) Thiolate. *Inorg. Chem.* **2017**, *56*, 956−961.
(54) Bootsma, A. N.; Wheeler, S. Popular Integration Grids Can. Result in Large Errors in DFT-Computed Free Energies. *ChemRxiv* **2019**, na.
(55) Gruzman, D.; Karton, A.; Martin, J. M. 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.
(56) Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. Turbomole. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 91−100.
(57) Balasubramani, S. G.; et al. TURBOMOLE: Modular program suite for ab initio quantum-chemical and condensed-matter simulations. *J. Chem. Phys.* **2020**, *152*, 184107.
(58) 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.
(59) Zhao, Y.; Truhlar, D. G. Assessment of Density Functionals for $\pi$ Systems: Energy Differences between Cumulenes and Poly-ynes; Proton Affinities, Bond Length Alternation, and Torsional Potentials of Conjugated Polyenes; and Proton Affinities of Conjugated Shiff Bases. *J. Phys. Chem. A* **2006**, *110*, 10478−10486.
(60) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. Multi-coefficient extrapolated density functional theory for thermochemistry and thermochemical kinetics. *Phys. Chem. Chem. Phys.* **2005**, *7*, 43−52.
(61) Zhao, Y.; González-García, N.; Truhlar, D. G. Benchmark Database of Barrier Heights for Heavy Atom Transfer, Nucleophilic Substitution, Association, and Unimolecular Reactions and its Use to Test Theoretical Methods. *J. Phys. Chem. A* **2006**, *110*, 4942.
(62) Goerigk, L.; Grimme, S. A general database for main group thermochemistry, kinetics, and noncovalent interactions - Assessment of common and reparameterized (meta-)GGA density functionals. *J. Chem. Theory Comput.* **2010**, *6*, 107−126.
(63) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. Gaussian-2 theory for molecular energies of first- and second-row compounds. *J. Chem. Phys.* **1991**, *94*, 7221−7230.
(64) Parthiban, S.; Martin, J. M. Assessment of W1 and W2 theories for the computational of electron affinities, ionization potentials, heats of formation, and proton affinities. *J. Chem. Phys.* **2001**, *114*, 6014−6029.
(65) Karton, A.; Daon, S.; Martin, J. M. W4−11: A high-confidence benchmark dataset for computational thermochemistry derived from first-principles W4 data. *Chem. Phys. Lett.* **2011**, *510*, 165−178.
(66) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(67) 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.
(68) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158−6170.
(69) Lee, C.; Yang, E.; Parr, R. G. Development of the Colic-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785.
(70) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(71) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785−789.
(72) Van Lenthe, E.; Baerends, E. J. Optimized Slater-type basis sets for the elements 1−118. *J. Comput. Chem.* **2003**, *24*, 1142−1156.
(73) Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33*, 8822.
(74) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098.
(75) 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 function. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(76) 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.
(77) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(78) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.* **1995**, *240*, 283−290.
(79) Dunlap, B. I.; Connolly, J. W.; Sabin, J. R. On some approximations in applications of X$\alpha$ theory. *J. Chem. Phys.* **1979**, *71*, 3396.
(80) Baerends, E. J.; Ellis, D. E.; Ros, P. Self-consistent molecular Hartree−Fock−Slater calculations I. The computational procedure. *Chem. Phys.* **1973**, *2*, 41−51.
(81) 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.
(82) Bühl, M.; Kabrede, H. Geometries of transition-metal complexes from density-functional theory. *J. Chem. Theory Comput.* **2006**, *2*, 1282−1290.
(83) Piccardo, M.; Penocchio, E.; Puzzarini, C.; Biczysko, M.; Barone, V. Semi-Experimental Equilibrium Structure Determinations by Employing B3LYP/SNSD Anharmonic Force Fields: Validation and Application to Semirigid Organic Molecules. *J. Phys. Chem. A* **2015**, *119*, 2058−2082.
(84) Piccardo, M.; Penocchio, E.; Puzzarini, C.; Biczysko, M.; Barone, V. Correction to: Semi-Experimental Equilibrium Structure Determinations by Employing B3LYP/SNSD Anharmonic Force Fields: Validation and Application to Semirigid Organic Molecules. *J. Phys. Chem. A* **2016**, *120*, 3754.
(85) 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.
(86) 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.

(87) Brauer, B.; Kesharwani, M. K.; Kozuch, S.; Martin, J. M. The S66 × 8 benchmark for noncovalent interactions revisited: explicitly correlated ab initio methods and density functional theory. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20905−20925.

(88) Mehta, N.; Fellowes, T.; White, J. M.; Goerigk, L. CHAL336 Benchmark Set: How Well Do Quantum-Chemical Methods Describe Chalcogen-Bonding Interactions? *J. Chem. Theory Comput.* **2021**, *17*, 2783−2806.

(89) Kesharwani, M. K.; Manna, D.; Sylvetsky, N.; Martin, J. M. The X40 × 10 Halogen Bonding Benchmark Revisited: Surprising Importance of (n-1)d Subvalence Correlation. *J. Phys. Chem. A* **2018**, *122*, 2184−2197.

(90) Miriyala, V. M.; Řezáč, J. Correction to: Testing semiempirical QM methods on a data set of interaction energies mapping repulsive contacts in organic molecules. *J. Phys. Chem. A* **2018**, *122*, 9585−9586.

(91) 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.

(92) Bohle, F.; Seibert, J.; Grimme, S. Automated Quantum Chemistry-Based Calculation of Optical Rotation for Large Flexible Molecules. *J. Org. Chem.* **2021**, *86*, 15522−15531.

(93) Bohle, F.; Grimme, S. Hydrocarbon Macrocycle Conformer Ensembles and $^{13}$C-NMR Spectra. *Angew. Chem., Int. Ed.* **2022**, *61*, e202113905.

(94) 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.

(95) 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.

(96) Bursch, M.; Hansen, A.; Pracht, P.; Kohn, J. T.; Grimme, S. Theoretical study on conformational energies of transition metal complexes. *Phys. Chem. Chem. Phys.* **2021**, *23*, 287−299.

(97) Řezáč, J.; Bím, D.; Gutten, O.; Rulíšek, L. Toward Accurate Conformational Energies of Smaller Peptides and Medium-Sized Macrocycles: MPCONF196 Benchmark Energy Data Set. *J. Chem. Theory Comput.* **2018**, *14*, 1254−1266.

(98) Sharapa, D. I.; Genaev, A.; Cavallo, L.; Minenkov, Y. A Robust and Cost-Efficient Scheme for Accurate Conformational Energies of Organic Molecules. *ChemPhysChem* **2019**, *20*, 92−102.

(99) Husch, T.; Freitag, L.; Reiher, M. Correction to: Calculation of Ligand Dissociation Energies in Large Transition-Metal Complexes. *J. Chem. Theory Comput.* **2019**, *15*, 4295−4296.

(100) Husch, T.; Freitag, L.; Reiher, M. Calculation of Ligand Dissociation Energies in Large Transition-Metal Complexes. *J. Chem. Theory Comput.* **2018**, *14*, 2456−2468.

(101) Sun, Y.; Chen, H. Performance of density functionals for activation energies of Zr-mediated reactions. *J. Chem. Theory Comput.* **2013**, *9*, 4735−4743.

(102) Sun, Y.; Chen, H. Performance of density functionals for activation energies of re-catalyzed organic reactions. *J. Chem. Theory Comput.* **2014**, *10*, 579−588.

(103) Sun, Y.; Hu, L.; Chen, H. Comparative assessment of DFT performances in Ru- and Rh-promoted $\sigma$-bond activations. *J. Chem. Theory Comput.* **2015**, *11*, 1428−1438.

(104) Hu, L.; Chen, H. Assessment of DFT Methods for Computing Activation Energies of Mo/W-Mediated Reactions. *J. Chem. Theory Comput.* **2015**, *11*, 4601−4614.

(105) Semidalas, E.; Martin, J. M. The MOBH35 Metal−Organic Barrier Heights Reconsidered: Performance of Local-Orbital Coupled Cluster Approaches in Different Static Correlation Regimes. *J. Chem. Theory Comput.* **2022**, *18*, 883−898.

(106) Iron, M. A.; Janes, T. Evaluating Transition Metal Barrier Heights with the Latest Density Functional Theory Exchange-Correlation Functionals: The MOBH35 Benchmark Database. *J. Phys. Chem. A* **2019**, *123*, 3761−3781.

(107) Iron, M. A.; Janes, T. Correction to: Evaluating transition metal barrier heights with the latest density functional theory exchange-correlation functionals: The MOBH35 benchmark database. *J. Phys. Chem. A* **2019**, *123*, 6379−6380.

(108) Blaško, M.; Pašteka, L. F.; Urban, M. DFT Functionals for Modeling of Polyethylene Chains Cross-Linked by Metal Atoms. DLPNO-CCSD(T) Benchmark Calculations. *J. Phys. Chem. A* **2021**, *125*, 7382−7395.

(109) Neese, F. Software update: The ORCA program system−Version 5.0. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2022**, e1606.

## APPENDIX B

# Benchmark Study on the Calculation of $^{207}$Pb NMR Chemical Shifts

Thomas Gasevic,[†] Julius B. Kleine Büning, Stefan Grimme, and Markus Bursch

*Received: December 21, 2023*
*First published: March 6, 2024*

Reprinted (adapted) with permission from:
T. Gasevic, J. B. Kleine Büning, S. Grimme, and M. Bursch, *Benchmark Study on the Calculation of $^{207}$Pb NMR Chemical Shifts*, Inorg. Chem. **63**.11 (2024) 5052, DOI: `10.1021/acs.inorgchem.3c04539`

**Own contributions**

- Conceptualization
- Data Curation – Collecting experimental $^{207}$Pb NMR Chemical Shifts and Crystal Structures
- Formal Analysis & Investigation – Generating Conformer Ensembles
- Visualization
- Writing – Original Draft Preparation, Review & Editing

---

[†] Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany

[‡] Max-Planck-Institut für Kohlenforschung, 45470 Mülheim an der Ruhr, Germany

Inorganic Chemistry

pubs.acs.org/IC

Article

# Benchmark Study on the Calculation of $^{207}$Pb NMR Chemical Shifts

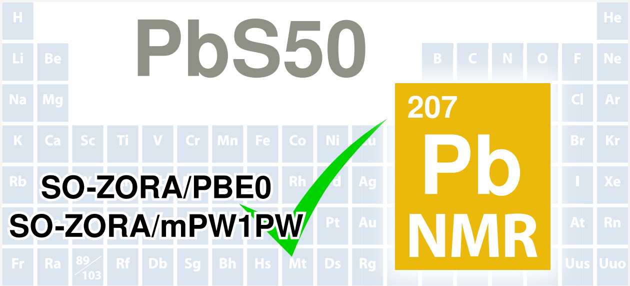
Thomas Gasevic, Julius B. Kleine Büning, Stefan Grimme,* and Markus Bursch*

Cite This: *Inorg. Chem.* 2024, 63, 5052−5064

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

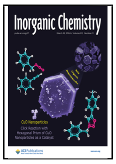
**ABSTRACT:** A benchmark set for the computation of $^{207}$Pb nuclear magnetic resonance (NMR) chemical shifts is presented. The *PbS50* set includes conformer ensembles of 50 lead-containing molecular compounds and their experimentally measured $^{207}$Pb NMR chemical shifts. Various bonding motifs at the Pb center with up to seven bonding partners are included. Six different solvents were used in the measurements. The respective shifts lie in the range between +10745 and −5030 ppm. Several calculation settings are assessed by evaluating computed $^{207}$Pb NMR shifts for the use with different density functional approximations (DFAs), relativistic approaches, treatment of the conformational space, and levels for geometry optimization. Relativistic effects were included explicitly with the zeroth order regular approximation (ZORA), for which only the spin−orbit variant was able to yield reliable results. In total, seven GGAs and three hybrid DFAs were tested. Hybrid DFAs significantly outperform GGAs. The most accurate DFAs are mPW1PW with a mean absolute deviation (MAD) of 429 ppm and PBE0 with an MAD of 446 ppm. Conformational influences are small as most compounds are rigid, but more flexible structures still benefit from Boltzmann averaging. Including explicit relativistic treatments such as SO-ZORA in the geometry optimization does not show any significant improvement over the use of effective core potentials (ECPs).



## ■ INTRODUCTION

One of the leading analytical tools for the elucidation of chemical structures is nuclear magnetic resonance (NMR) spectroscopy. Respective measurements are commonly applied to NMR-active $^{1}$H and $^{13}$C nuclei. Nevertheless, heteronuclei such as $^{207}$Pb can also reveal toxicity in organisms[1] and give a valuable insight into the chemical environment around the respective nucleus, e.g., the coordination sphere.[1,2] Recorded NMR spectra and the NMR measurement itself can already be extremely complicated for small but complex structures. Further, NMR signals for heavier nuclei can occur over a broad range of shifts, which requires time-consuming measurements in order to locate the actual signal. Therefore, theoretical prediction methods have become indispensable for the setup and evaluation of experimental NMR data. Density functional theory (DFT) in combination with gauge-including atomic orbitals (GIAO)[3−6] is routinely applied to compute NMR chemical shifts, especially because of the excellent compromise between computational cost and accuracy.[7−9] Although reliable results can be achieved with this approach, it can be difficult to find the right density functional approximation (DFA) among a plethora of DFAs. Therefore, benchmark studies are essential for the choice of DFA. Computations of $^{207}$Pb NMR shifts have already been carried out for, e.g., dirhodioplumbole,[10] organolead[11] and halogen lead complexes,[12] diarylplumbylenes,[13] plumbacyclopentadienylidenes,[14] and endohedral plumbaspherenes.[15] Further studies include the computation of solid-state $^{207}$Pb NMR chemical shifts that already reveal the importance of Fock exchange and spin−orbit (SO) effects.[16−20] Still, they all lack a thorough assessment of the underlying theoretical method based on a diverse set of lead-containing compounds that were measured in solution.[10−15,21] Due to the presence of unoccupied p-orbitals in group 14 elements, the $^{207}$Pb nucleus has a huge SO heavy atom effect on the shielding of neighboring light atoms (HALA effects), especially in Pb$^{II}$ compounds.[22] Heavy atoms in the vicinity of the Pb nucleus can also have a heavy atom effect on the heavy atom Pb (HAHA). Both phenomena require the explicit treatment of relativistic effects, which can be taken into account by the Douglas−Kroll−Hess,[23,24] the zeroth-order regular approximation (ZORA),[25,26] or the exact two-component (X2C) approaches.[27−31] All of them apply a scalar relativistic (SR) treatment but can optionally also include SO coupling, which is necessary to describe the HALA and HAHA effects. As a follow-up to our $^{29}$Si[32] and $^{119}$Sn NMR[33] chemical shift benchmark sets, we now introduce a set of compounds for the quantum chemical computation of $^{207}$Pb NMR chemical shifts. The *PbS50* set features 50 unique experimental $^{207}$Pb NMR

**Received:** December 21, 2023
**Revised:** February 8, 2024
**Accepted:** February 14, 2024
**Published:** March 6, 2024



ACS Publications

https://doi.org/10.1021/acs.inorgchem.3c04539
*Inorg. Chem.* 2024, 63, 5052−5064

chemical shifts of molecular compounds in solution. It includes diverse structures in different chemical environments to cover a broad range of available chemical shifts and represent a large part of the Pb-containing molecular chemical space. The main goal is to find recommendations for the computational setup for predictions of $^{207}$Pb NMR chemical shifts.

## ■ COMPUTATIONAL DETAILS

All quantum chemical computations have been carried out with the xtb 6.4.1,[34,35] TURBOMOLE 7.6,[36−38] and AMS 2023.103[39,40] program packages. Initial geometries were obtained from X-ray crystal structures. If no crystal structure was available, the compounds were manually constructed from similar crystal structures (compounds **39**, **41**, **42**, **44**, and **50**). All structures were preoptimized with the semiempirical quantum mechanical (SQM) GFN2-xTB[41] method in combination with the ALPB implicit solvation model[42] for the respective solvent that was used in the NMR experiment. Conformer ensembles (CE) were created with the conformer-rotamer ensemble sampling tool CREST 2.11.3[43,44] at the GFN2-xTB level of theory. Due to optimization issues, the CEs of compounds **1**, **10**, **20**, **21**, and **34** were computed at the generic force field GFN-FF[45] level of theory. The ensembles were further refined with the command-line energetic sorting tool CENSO 1.2.0,[46,47] and still identical conformers were manually removed by inspecting each ensemble. Within the CENSO algorithm, geometries were optimized with the efficient r$^2$SCAN-3c[48] composite DFT method and the COSMO[49] implicit solvation model in TURBOMOLE. A fine radial integration grid was applied (*radsize* = 10). Solvation free energies were computed with COSMO-RS[50,51] using the COSMOtherm 19.0[52] program package ($G_{\text{solv}}$ option, BP_TZVP_C30_1601 parameterization, $T$ = 298.15 K, $p$ = 1 atm). Thermostatistical contributions were obtained with the single-point Hessian method[53] within the modified rigid-rotor harmonic oscillator[54] approach using GFN2-xTB. All structures were verified as minima on the potential energy surface by the absence of significant imaginary frequencies below $-i\omega$ = 30 cm$^{-1}$ in the numerical harmonic frequency calculation. The resolution of identity approximation for Coulomb integrals (RI-J) was applied in all calculations carried out with matching auxiliary basis sets in TURBOMOLE.

All $^{207}$Pb NMR chemical shieldings were computed in ADF from the AMS program package with GIAOs and the COSMO solvation model[55] for the respective solvent applied in the experimental NMR measurement. Relativistic effects were treated by the SR or the SO variant of ZORA.[25,56] The density functionals listed in Table 1 were applied in combination with the ZORA/TZP[57] basis set. The *symmetry* option was turned off, and the *NumericalQuality* was set to good in all computations.

For a subset consisting of compounds **26**, **28−43**, and **48**, the SO contribution was found to be below 200 ppm. $^{207}$Pb NMR chemical shieldings for this subset were computed in the ORCA 5.0.4 program package[58,59] with the CPCM solvation model[60] for the respective solvent applied in the experimental NMR measurement. SR-ZORA was applied in all calculations together with the ZORA-def2-TZVP basis set[61] for all elements with $Z \leq 36$ and SARC-ZORA-TZVP[62,63] for all with $Z > 36$. The chain-of-spheres approach to RIJCOSX and the Dobson ansatz for $\tau$ in meta-GGAs were used throughout. The density functionals PBE, TPSS,[64] r$^2$SCAN,[65] B97M-V,[66] TPSSh,[67] r$^2$SCAN0,[68] $\omega$B97X-V,[69] and $\omega$B97M-V[70] were tested in this study.

**Table 1. List of All Tested DFAs and Relativistic Treatments (SR = Scalar Relativistic, SO = Spin−Orbit)**[a]

| | | ZORA | |
|---|---|---|---|
| class | DFA | SR | SO |
| GGA | PBE[71] | + | + |
| | revPBE[72] | + | + |
| | BLYP[73,74] | + | + |
| | OLYP[75] | + | + |
| | BP86[76,77] | + | + |
| | mPW[78] | + | + |
| | KT2[79] | + | + |
| hybrid | PBE0[80] | + | + |
| | B3LYP[81−83] | + | + |
| | mPW1PW[78] | + | + |

[a]Each DFA was applied with the ZORA/TZP basis set.

NMR chemical shifts $\delta$ were obtained by referencing the respective shielding constant of compound k to the reference compound PbMe$_4$ according to eq 1

$$\delta_{\text{k}}(^{207}\text{Pb}) = \sigma_{\text{PbMe}_4}(^{207}\text{Pb}) - \sigma_{\text{k}}(^{207}\text{Pb}) \tag{1}$$

To obtain thermally averaged NMR chemical shifts, Boltzmann weights $w_i$ of each conformer $i$ in the CE were calculated according to

$$w_i = \frac{\exp(-\Delta G_i/RT)}{\sum_j^{\text{CE}} \exp(-\Delta G_j/RT)} \tag{2}$$

with the difference in Gibbs free energy between each conformer and the lowest conformer of the ensemble $\Delta G_i = G_i - G_{\text{lowest}}$, the molar gas constant $R$, and the temperature $T$ = 298.15 K.

To investigate the influence of the applied method in the geometry optimization, the lowest-lying conformer of each ensemble was reoptimized with the SQM methods GFN1- and GFN2-xTB, the GFN-FF force field, as well as with the SR- and SO-ZORA variants of r$^2$SCAN-3c(STO),[84] as implemented in ADF. Solvation effects were treated with the ALPB and COSMO solvation models, respectively. The *NumericalQuality* was set to good in ADF.

If not stated otherwise, default settings were applied.

## ■ RESULTS AND DISCUSSION

**General Considerations.** Although quantum chemical calculations of NMR chemical shifts are routinely used, one faces various obstacles in the computational workflow. For example, the question of conformational influences arises when the system under investigation is flexible. Since the measured nucleus is differently shielded in each conformer, it is crucial to consider the whole CE and take the respective populations into account to predict NMR shifts more accurately. Further, the choice of the density functional, basis set, solvation model, and relativistic treatment has to be made for the geometry optimization and the subsequent NMR chemical shielding calculation. For a more general discussion of those effects, see ref 85.

Some of the mentioned problems were previously assessed. For the choice of the basis set, a triple-$\zeta$ quality is already sufficient as it is converged with respect to the NMR chemical

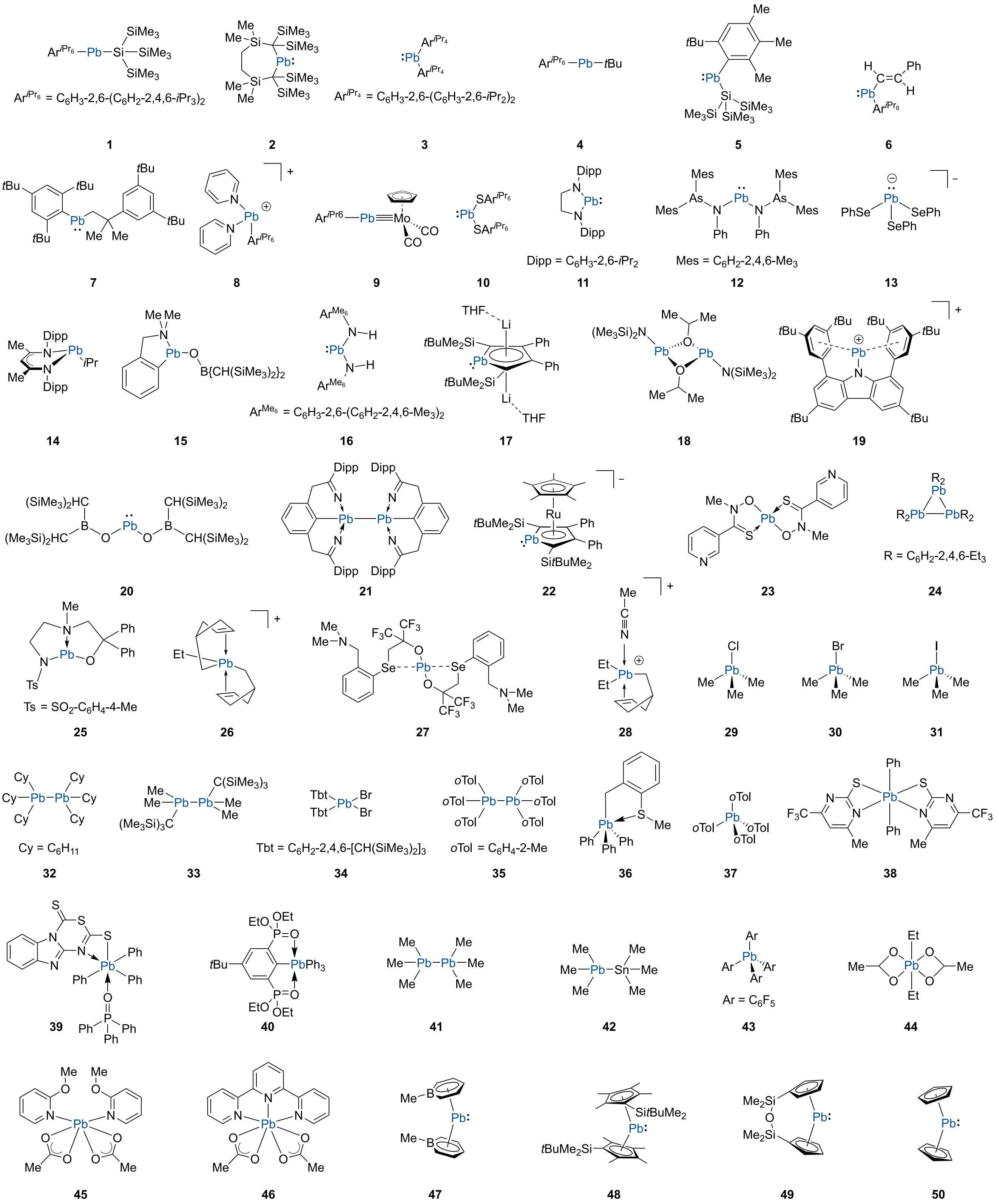


**Figure 1.** All compounds included in the *PbS50* benchmark study.

shift.[32] The influence of different implicit solvation models has also been proven to be small and is hence not tested here.[86] Still, explicit solvation effects can become crucial if a polar solvent, such as DMSO or THF, closely coordinates to the solute. This has already been reported, e.g., for $^{195}$Pt NMR shifts in the works by Autschbach et al.[20,87] of systems with water molecules explicitly coordinating to the platinum center. Due to the bulky ligands in the investigated compounds, the effect of explicit solvation is expected to be small for this test set. Studies on light-atom NMR shift calculations further

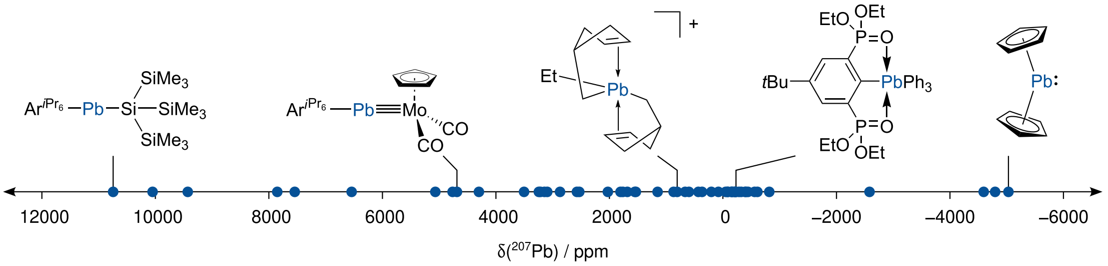


**Figure 2.** Overview of all experimentally determined $^{207}$Pb NMR chemical shifts included in this study.

investigated more density functionals[88] than presented in this work. The main reasons for this are the availability and limitations of quantum chemical codes that support relativistic effects beyond scalar. The AMS program package that was used in this work, e.g., does not support meta-GGAs for the computation of NMR shifts. Other program packages to be mentioned here are TURBOMOLE, which features a 2-component implementation at the X2C level[89] and DIRAC[90] as well as ReSpect[91] which both feature a 4-component implementation.

In this work, we study the performance of various density functionals applied for the $^{207}$Pb NMR chemical shift calculations as well as the influence of the conformational space and relativistic effects on that property. Further, we investigate different levels of theory for the geometry optimization. Finally, recommendations are given for the workflow of predicting $^{207}$Pb NMR chemical shifts with DFT by using a multilevel approach.

**Benchmark Set.** The newly compiled *PbS50* benchmark set includes CEs for 50 organolead compounds of various sizes ranging from 14 to 187 atoms, with experimental reference data measured in solution for 50 $^{207}$Pb NMR chemical shifts between +10745 and −5030 ppm. In total, 282 conformers are included. Numerous bonding motifs, such as formal single, double, triple, aromatic, or dative bonds, with up to seven bonding partners on one Pb atom are represented. The NMR shifts were recorded in six solvents: benzene, chloroform, dichloromethane, DMSO, THF, and toluene. All structures are depicted in Figure 1, and an overview of all experimental $^{207}$Pb NMR chemical shifts is shown in Figure 2. The experimental shifts and solvents that were used in the NMR experiment are listed in Table 2. Additional details can be found in the Supporting Information.

**Study of Density Functional Approximations and Relativistic Approaches for the Calculation of $^{207}$Pb NMR Chemical Shifts.** The performance of different DFAs is assessed in combination with the ZORA/TZP basis set on the *PbS50* benchmark set for the computation of the $^{207}$Pb NMR chemical shifts. If not stated otherwise, the calculated NMR results refer to the ensemble-averaged NMR chemical shifts based on the Boltzmann weights at room temperature. A comparison of all tested DFAs is given in Figure 3, and the respective statistical data are listed in Table 3.

Relativistic effects have an immense influence on the $^{207}$Pb NMR chemical shift, which was also observed for the computation of solid-state $^{207}$Pb NMR shifts.[19,132] The explicit treatment of only SR effects is insufficient to describe the chemical environment around the Pb nucleus (Figure 3a). The mean absolute deviations (MADs) range from 1999 to 2123 ppm, and the mean deviations (MDs) range from −1950 to −2083 ppm, indicating an underestimation for the majority of $^{207}$Pb NMR chemical shifts. The standard deviations are in a similar range as the MADs and therefore represent a large spread of errors. This results in small coefficients of determination, with the average over all methods being $R_{\text{av.}}^2 = 0.7095$. In contrast to our previous work on $^{29}$Si and $^{119}$Sn NMR chemical shift computations, the accuracy of $^{207}$Pb NMR chemical shifts does not differ significantly with different density functionals when SR-ZORA is applied. The largest errors were obtained with OLYP (MAD = 2123 ppm) and the lowest errors with BYLP (MAD = 1999 ppm), but the difference is negligible considering the range of investigated shifts. Especially plumbylenes and some systems with heavy-atom effects on heavy-atom shielding (HAHA), such as **9**, are described poorly.

By incorporating SO coupling into the calculation, the results become considerably more accurate, as former outlier values are improved significantly. Selected examples are shown in Figure 4. The MADs lie between 856 and 429 ppm and are hence reduced by a factor of almost five. Hybrid functionals systematically outperform GGAs and yield lower errors. Therefore, Fock exchange is required for more reliable results, which is in line with the "Jacob's Ladder" picture of DFAs and was also observed for the computation of solid-state $^{207}$Pb NMR shifts.[19,132] In comparison to our $^{29}$Si and $^{119}$Sn NMR benchmark sets, smaller coefficients of determination are observed for $^{207}$Pb NMR chemical shifts. This underlines the challenging computation of heavier elements. Outliers are mainly compounds with positively charged Pb atoms and plumbylenes, which have complex electronic structures that are difficult to describe. This is also in line with our previous work, where large errors were observed for silylenes and stannylenes. The overall best-performing DFAs are PBE0 (MAD = 446 ppm, $R^2 = 0.9828$) and mPW1PW (MAD = 429 ppm, $R^2 = 0.9831$).

Based on the best-performing DFA, we selected a subset of the *PbS50* set where the SO contributions are below 200 ppm. This allows for a screening of a wider variety of density functionals, such as meta-GGAs or range-separated hybrids (RSHs) (see Supporting Information for further details). No significant improvement over the remaining calculations was observed, but the RSHs ωB97X-V[69] and ωB97M-V[70] yield a slightly lower MAD than the tested global hybrids. Therefore, RSHs should be investigated in future studies.

**Linear Scaling Correction.** Errors inherent to the underlying theoretical method for the calculation of NMR chemical shifts are routinely corrected by an empirical linear scaling approach.[133−137] In the following, the correction is applied to the $^{207}$Pb NMR chemical shifts according to

$$\delta_{\text{calc,scaled}} = \frac{\delta_{\text{calc}} - \beta}{\alpha} \tag{3}$$

Table 2. Experimental $^{207}$Pb NMR Chemical Shifts for All Compounds in the *PbS50* Set, the Solvents Used during the Measurement, and the Respective Reference

| no | solvent | $\delta(^{207}Pb)$/ppm | refs |
|---|---|---|---|
| 1 | benzene | 10 745 | 92 |
| 2 | benzene | 10 050 | 93 |
| 3 | benzene | 9430 | 92 |
| 4 | benzene | 7853 | 92 |
| 5 | benzene | 7545 | 94 |
| 6 | benzene | 6543 | 95 |
| 7 | benzene | 5067 | 94 |
| 8 | benzene | 4764 | 96 |
| 9 | benzene | 4686 | 96 |
| 10 | benzene | 4299.2 | 97 |
| 11 | benzene | 3504 | 98 |
| 12 | benzene | 3244 | 99 |
| 13 | chloroform | 3214 | 100 |
| 14 | benzene | 3145 | 101 |
| 15 | benzene | 3095 | 102 |
| 16 | benzene | 2871 | 103 |
| 17 | benzene | 2572.5 | 104 |
| 18 | benzene | 2531 | 105 |
| 19 | dichloromethane | 2027 | 106 |
| 20 | benzene | 1805 | 102 |
| 21 | benzene | 1684 | 107 |
| 22 | benzene | 1552.5 | 108 |
| 23 | chloroform | 1534 | 109 |
| 24 | toluene | 1152 | 110 |
| 25 | DMSO | 870 | 111 |
| 26 | toluene | 806.7 | 112 |
| 27 | benzene | 665 | 113 |
| 28 | benzene | 598 | 114 |
| 29 | dichloromethane | 432 | 115 |
| 30 | dichloromethane | 367 | 115 |
| 31 | dichloromethane | 203.6 | 115 |
| 32 | benzene | 80.2 | 116 |
| 33 | benzene | −48.2 | 117 |
| 34 | chloroform | −83 | 118 |
| 35 | chloroform | −88.7 | 119 |
| 36 | chloroform | −146.2 | 120 |
| 37 | chloroform | −166.7 | 119 |
| 38 | chloroform | −206.7 | 121 |
| 39 | THF | −214.5 | 122 |
| 40 | benzene | −231 | 123 |
| 41 | benzene | −281 | 124 |
| 42 | benzene | −324 | 124 |
| 43 | chloroform | −391 | 125 |
| 44 | chloroform | −441 | 124 |
| 45 | chloroform | −558.9 | 126 |
| 46 | DMSO | −816.68 | 127 |
| 47 | dichloromethane | −2584 | 128 |
| 48 | benzene | −4595 | 129 |
| 49 | benzene | −4795 | 130 |
| 50 | benzene | −5030 | 131 |

with $\alpha$ being the slope and $\beta$ being the intercept of the linear regression done for the calculated $^{207}$Pb NMR chemical shift with respect to the experimental reference data. The two parameters for each DFA can be found in the Supporting Information.

By scaling the calculated shifts, the mean deviation vanishes, and the MAD is decreased (Figure 5). The MAD is lowered from 1999 to 1751 ppm in the SR-ZORA approach for BLYP. Nevertheless, the error spread for SR-ZORA is still too large, and therefore, the corrected results are not reliable. If the SO-ZORA results are scaled, however, the MAD is also decreased going from 856 to 582 ppm, thus almost reaching the accuracy of the computationally more expensive hybrid DFAs. When the correction is applied to the results computed with hybrid functionals, the error is even further reduced, leading to the overall best performance. For example, the MAD of mPW1PW is reduced from 429 to 338 ppm and the MAD of PBE0 from 446 to 341 ppm. These results should still be taken with caution as a linear scaling correction might hide intrinsic errors in the calculations.

**Conformational Analysis.** The impact of distinct conformers on the NMR chemical shift can be large in flexible structures.[138−140] The results in the previous sections were obtained by averaging the calculated $^{207}$Pb NMR shift of each compound based on their Boltzmann weights, but in the majority of cases, the most populated conformer dominates the NMR chemical shift. Table 3 lists the MADs for the results obtained only for the lowest energy conformer ($\mathrm{MAD}_{\mathrm{lowconf}}$). The $\mathrm{MAD}_{\mathrm{lowconf}}$ is always slightly larger than the Boltzmann-averaged MAD, but the difference is 5.5 ppm at most and is therefore negligible. Since most structures in the test set are considered to be rigid and primarily contain only one central Pb atom, the conformational influence was expected to be small. The largest influence is observed for compound **19**, with $\delta_{\mathrm{lowconf}}$ = 853 ppm and $\delta_{\mathrm{ensemble}}$ = 986 ppm when calculated with SO-ZORA-PBE0/TZP. The corresponding experimental shift is $\delta_{\mathrm{exp},19}$ = 2027 ppm. The reason for the difference is the different coordination of phenyl rings in the structure. The lowest conformer features an $\eta^6$ coordination of parallel-aligned phenyl rings to the central Pb atom. The less-populated conformers, however, exhibit differently coordinated phenyl rings that are not oriented parallel. Only the combination of both bonding motives leads to the most accurately computed $^{207}$Pb NMR chemical shift. This example illustrates the importance of conformational influences for a more accurate representation of the experimental NMR chemical shifts.

**Structure Dependence.** The accuracy of the NMR chemical shift calculation depends on the quality of the equilibrium structure obtained by geometry optimization. Hence, different levels of theory for the optimization are evaluated in this section (Figure 6). The treatment of relativistic effects is compared for effective core potentials (ECPs), SR-, and SO-ZORA. In addition, the SQM methods GFN1-xTB, GFN2-xTB, and the general force field GFN-FF are investigated. For this, the geometry of the lowest conformer in each ensemble was reoptimized at the respective level of theory, and the $^{207}$Pb NMR chemical shifts were computed at the SO-ZORA/PBE0/TZP level of theory.

The treatment of relativistic effects is essential for reliable geometries that contain heavy elements such as Pb.[141] These can be treated implicitly either by, e.g., ECPs that include relativistic reference data in their fit or by an all-electron treatment such as ZORA. The Stuttgart-Cologne def2-ECPs, which were applied for the geometry optimization of all conformers in the ensembles, yield very similar results as the computationally more demanding SO variant of ZORA in combination with the composite r$^2$SCAN-3c DFT method. The respective MADs in subsequent $^{207}$Pb NMR chemical shift calculations are 449 and 437 ppm. Also, the mean deviation is similar with −375 and −352 ppm, respectively. In the SR-

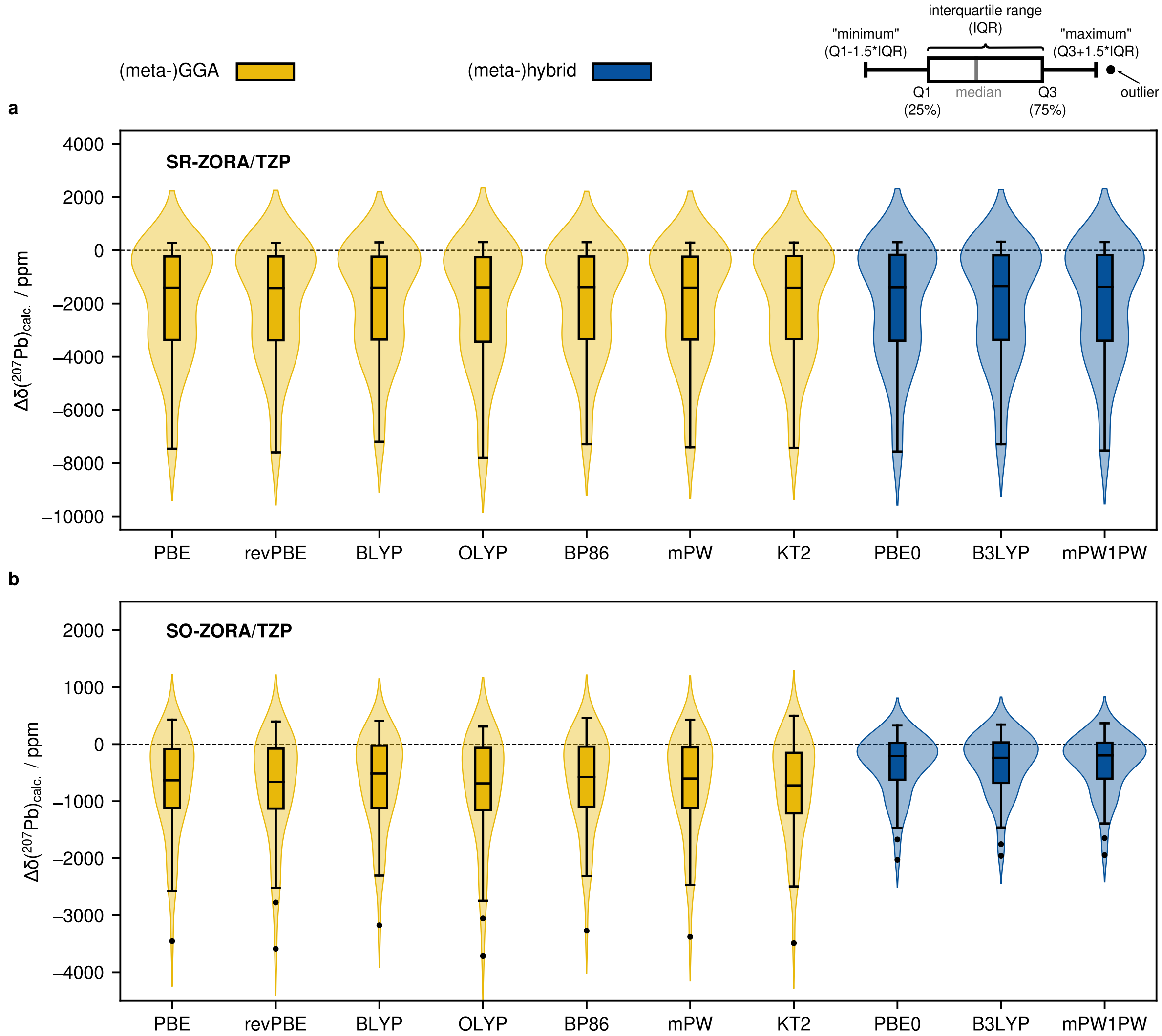


**Figure 3.** $^{207}$Pb NMR shift deviation between calculation (ensemble-averaged shifts) and experiment represented in violin plots ($\Delta\delta = \delta_{calc} - \delta_{exp}$) for all tested DFAs using the SR (a) and SO relativistic (b) ZORA approach. The central lines represent the median values, the boxes the range of 25−75% of the data, and the whiskers all points within 1.5 times the interquartile range. Outliers are depicted by black dots.

ZORA approach, where SO coupling is not considered, the errors are slightly larger. The MAD is 483 ppm, and the MD is −405 ppm. Since ECPs are more common and usually computed faster than all-electron approaches when several heavy nuclei are present, it is advised to apply them in the geometry optimization, for which many energy and gradient computations have to be performed.

Even more efficient geometry optimizations are possible with semiempirical or force field methods. However, some calculations with the GFN1-xTB and GFN-FF methods do not converge for all systems or yield qualitatively wrong geometries. If these compounds are removed from the test set, the workflow still yields reasonable results. The MADs of the reduced test set (MAD$_{red.}$) are 425 and 792 ppm for GFN1-xTB and GFN-FF, respectively. In comparison, the ECP-based r$^2$SCAN-3c method yield an MAD$_{red.}$ of 359 ppm.

The GFN2-xTB method has an MAD$_{red.}$ of 474 ppm, which is slightly worse than that of GFN1-xTB, but it is more robust as all geometries were optimized correctly. The results can further be improved by applying the aforementioned linear scaling approach (Figure 6b), which here mainly corrects for systematically incorrect bond lengths. After linearly scaling the NMR chemical shifts, an MAD$_{scaled}$ of 409 ppm is achieved, which is lower than the unscaled results obtained from r$^2$SCAN-3c/ECP geometries (MAD = 449 ppm). Therefore, geometries obtained with GFN2-xTB represent a viable alternative for the computation of $^{207}$Pb NMR chemical shifts. As the computations are much faster than DFT optimizations, this can, e.g., be applied in large-scale screening processes.

## ■ CONCLUSIONS

The computation of NMR chemical shifts can facilitate experimental measurements and support the evaluation of measured spectra. In this work, a comprehensive benchmark set for the computation of $^{207}$Pb NMR chemical shifts was introduced, which features 50 experimental $^{207}$Pb NMR chemical shifts measured in solution. It includes conformers of 50 compounds with various bonding patterns at the lead center. The set is aimed to be used for the evaluation of NMR prediction methods, but it can also be regarded as a database to find structures that exhibit $^{207}$Pb NMR signals in a similar range. Herein, it was used to evaluate the performance of DFT and the influence of conformational and relativistic effects in

Table 3. Mean Deviation (MD), Mean Absolute Deviation (MAD), Standard Deviation (SD), and Root Mean Square Deviation (RMSD) of the Ensemble-Based Computed $^{207}$Pb NMR Chemical Shifts in ppm as Well as Their Respective Coefficient of Determination ($R^2$)[a]

| DFA | MD | MAD | SD | RMSD | $R^2$ | MAD$_{scaled}$ | MAD$_{lowconf}$ |
|---|---|---|---|---|---|---|---|
| | | | SR-ZORA/ZORA/TZP | | | | |
| PBE | −1996 | 2038 | 2131 | 2904 | 0.7211 | 1807 | 2038 |
| revPBE | −2022 | 2062 | 2163 | 2945 | 0.7090 | 1858 | 2062 |
| BLYP | −1950 | 1999 | 2076 | 2833 | 0.7356 | 1751 | 1999 |
| OLYP | −2083 | 2123 | 2225 | 3032 | 0.6772 | 2002 | 2123 |
| BP86 | −1958 | 2005 | 2093 | 2850 | 0.7323 | 1761 | 2005 |
| mPW | −1985 | 2029 | 2118 | 2887 | 0.7239 | 1797 | 2029 |
| KT2 | −1992 | 2033 | 2117 | 2891 | 0.7308 | 1762 | 2034 |
| PBE0 | −2018 | 2065 | 2205 | 2972 | 0.6808 | 2018 | 2066 |
| B3LYP | −1970 | 2025 | 2140 | 2893 | 0.7022 | 1921 | 2025 |
| mPW1PW | −2011 | 2060 | 2197 | 2962 | 0.6822 | 2013 | 2061 |
| | | | SO-ZORA/ZORA/TZP | | | | |
| PBE | −731 | 801 | 863 | 1124 | 0.9605 | 519 | 805 |
| revPBE | −756 | 823 | 895 | 1165 | 0.9580 | 537 | 827 |
| BLYP | −668 | 748 | 807 | 1041 | 0.9632 | 510 | 751 |
| OLYP | −787 | 856 | 941 | 1219 | 0.9516 | 582 | 860 |
| BP86 | −679 | 759 | 823 | 1061 | 0.9626 | 506 | 763 |
| mPW | −707 | 780 | 844 | 1094 | 0.9615 | 513 | 784 |
| KT2 | −773 | 844 | 866 | 1154 | 0.9608 | 527 | 848 |
| PBE0 | −373 | 446 | 526 | 640 | 0.9828 | 342 | 449 |
| B3LYP | −360 | 451 | 530 | 636 | 0.9811 | 364 | 457 |
| mPW1PW | −350 | 429 | 510 | 614 | 0.9831 | 338 | 433 |

[a] Note that the mean deviation of the linear scaling approach is, by definition, always zero. Further, the MAD of the scaled (MAD$_{scaled}$) and of the lowest conformer in each ensemble (MAD$_{lowconf}$) are presented.

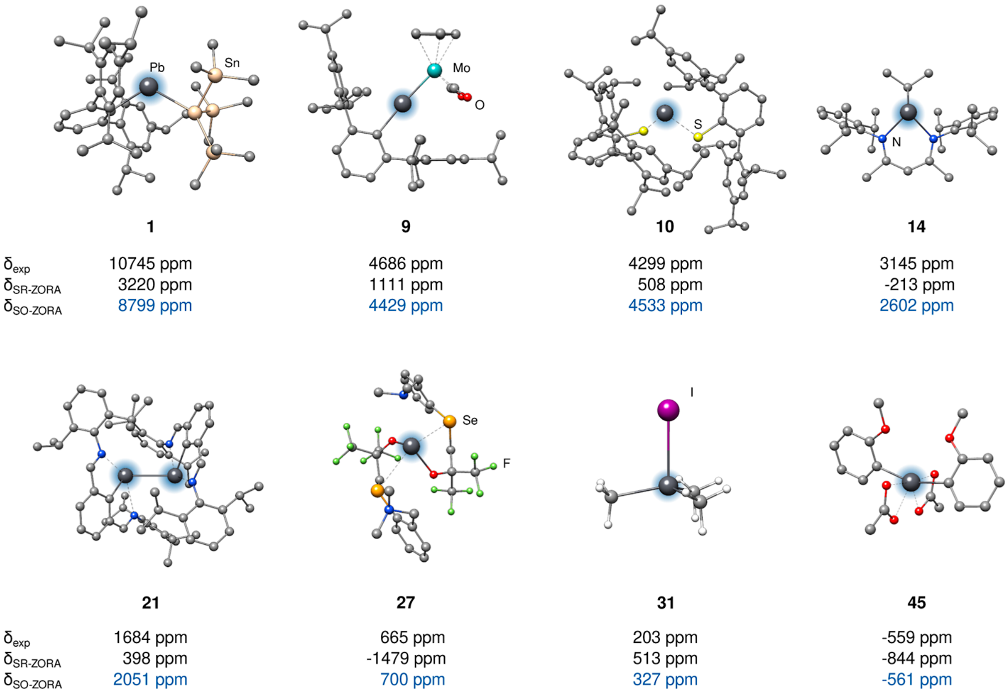


**Figure 4.** Selected systems of the *PbS50* set with their respective experimental and calculated $^{207}$Pb NMR chemical shifts computed at the mPW1PW/TZP level of theory. The most accurately calculated NMR shifts are highlighted in blue.

the computation of the $^{207}$Pb NMR chemical shifts. Further, the effect of the underlying level on geometry optimization was assessed.

Relativistic effects were explicitly treated by SR- and SO-ZORA in the calculations. When SR-ZORA is applied in the $^{207}$Pb NMR shift calculation, the shifts are on average underestimated, and the errors are immense, regardless of the used DFA. By applying SO-ZORA, the results are greatly improved, as expected, for the heavy element Pb. Inclusion of SO effects for the calculation of NMR chemical shifts is not very common in quantum chemical codes, yet. Therefore, efforts for additive corrections, e.g., by applying machine

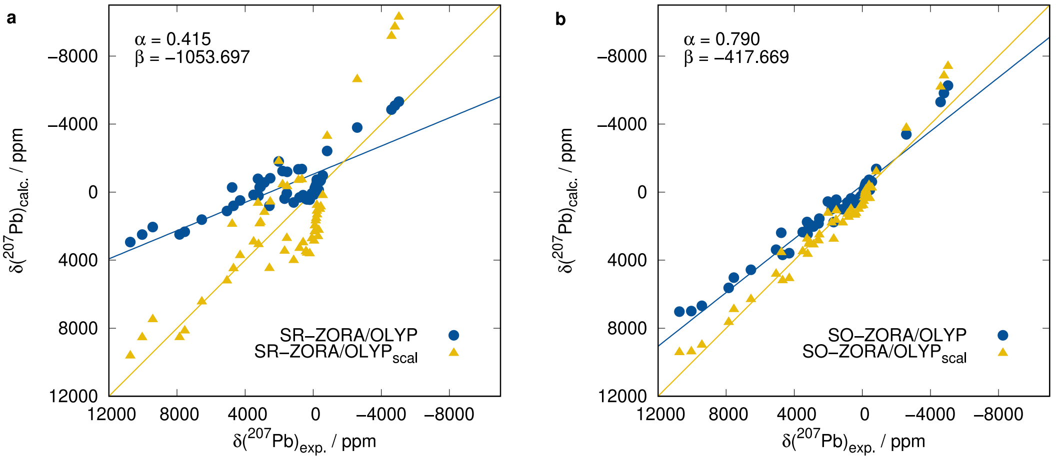


**Figure 5.** Correlation plots showing the calculated and experimental $^{207}$Pb NMR shifts for the worst performing DFA OLYP/TZP applying (a) SR-ZORA and (b) SO-ZORA. The plots also show a comparison between the unscaled and scaled (scal) results.

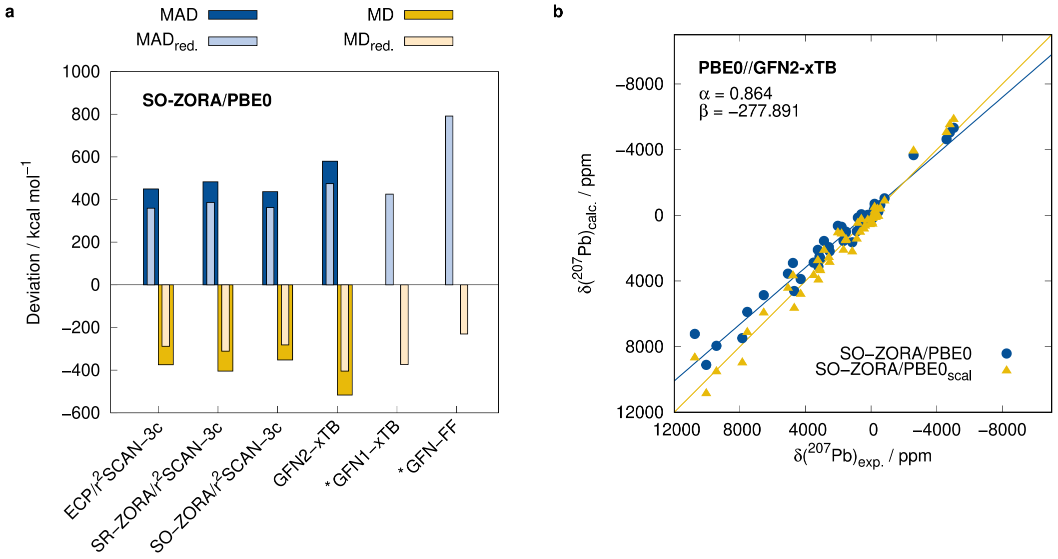


**Figure 6.** (a) MAD and MD of the $^{207}$Pb NMR shift calculated with SO-ZORA/PBE0/TZP for different levels of geometry optimization and (b) correlation plot between the calculated and experimental $^{207}$Pb NMR shift for GFN2-xTB geometries, including unscaled and scaled (scal) results. Plot (b) contains all 50 data points. *Compounds **1**, **2**, **3**, **5**, **9**, **19**, **21**, **33**, and **46** were removed from the test set to obtain the reduced mean (absolute) deviations MAD$_{red.}$ and MD$_{red.}$ as these systems did not converge or were poorly optimized by GFN1-xTB and GFN-FF.

learning approaches, have already been made.[142] Hybrid density functionals were found to generally outperform the GGA functionals.

As the majority of tested systems are rigid, conformational effects are small. For most cases, it suffices to compute the NMR chemical shift only for the most stable conformer. Nevertheless, special cases require the inclusion of some low-lying conformers in order to predict the correct NMR shifts. For example, the $^{207}$Pb NMR chemical shift of compound **19** is computed more accurately, and the error is reduced by over 130 ppm when the shift is ensemble-averaged.

Our test on the level of theory applied in the geometry optimization showed that the explicit treatment of relativistic effects is negligible, and ECPs should be used instead for substantial computational time savings. Faster geometry optimizations are possible with the semiempirical GFN2-xTB method. The errors in a subsequent $^{207}$Pb NMR chemical shift calculation are still reasonable if the shifts are scaled linearly. Therefore, this approach can be successfully used in screening processes.

The overall best-performing method combinations for the computation of $^{207}$Pb NMR chemical shifts evaluated on the *PbS50* set are PBE0/ZORA/TZP (MAD = 446 ppm) and mPW1PW/ZORA/TZP (MAD = 429 ppm) in combination with the SO variant of ZORA.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.inorgchem.3c04539.

Geometry files for each compound of the PbS50 set are in XYZ format (pbs50.zip) (ZIP)

Computed $^{207}$Pb NMR shielding and shift values (XLSX)

More detailed computational settings and supporting figures (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Stefan Grimme** − *Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany;* orcid.org/0000-0002-5844-4371; Email: grimme@thch.uni-bonn.de

**Markus Bursch** − *Max-Planck-Institut für Kohlenforschung, 45470 Mülheim an der Ruhr, Germany;* orcid.org/0000-0001-6711-5804; Email: bursch@kofo.mpg.de

### Authors

**Thomas Gasevic** − *Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany;* orcid.org/0000-0003-4864-1758

**Julius B. Kleine Büning** − *Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany;* orcid.org/0000-0001-7232-4897

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.inorgchem.3c04539

### Funding

Open access funded by Max Planck Society.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

S.G. and M.B. gratefully acknowledge financial support from the Max Planck Society through the Max Planck fellow program.

## ■ REFERENCES

(1) Neupane, K. P.; Pecoraro, V. L. Probing a Homoleptic PbS 3 Coordination Environment in a Designed Peptide Using 207 Pb NMR Spectroscopy: Implications for Understanding the Molecular Basis of Lead Toxicity. *Angew. Chem., Int. Ed.* **2010**, *49*, 8177−8180.

(2) Claudio, E. S.; ter Horst, M. A.; Forde, C. E.; Stern, C. L.; Zart, M. K.; Godwin, H. A. 207 Pb 1 H Two-Dimensional NMR Spectroscopy: A Useful New Tool for Probing Lead(II) Coordination Chemistry. *Inorg. Chem.* **2000**, *39*, 1391−1397.

(3) Ditchfield, R. Self-Consistent Perturbation Theory of Diamagnetism: I. A Gauge-Invariant LCAO Method for N.M.R. Chemical Shifts. *Mol. Phys.* **1974**, *27*, 789−807.

(4) Wolinski, K.; Hinton, J. F.; Pulay, P. Efficient Implementation of the Gauge-Independent Atomic Orbital Method for NMR Chemical Shift Calculations. *J. Am. Chem. Soc.* **1990**, *112*, 8251−8260.

(5) Schreckenbach, G.; Ziegler, T. Calculation of NMR shielding tensors using gauge-including atomic orbitals and modern density functional theory. *J. Phys. Chem.* **1995**, *99*, 606−611.

(6) Pankratyev, E. Y.; Tulyabaev, A. R.; Khalilov, L. M. How reliable are GIAO calculations of $^{1}$H and $^{13}$C NMR chemical shifts? A statistical analysis and empirical corrections at DFT (PBE/3z) level. *J. Comput. Chem.* **2011**, *32*, 1993−1997.

(7) Safi, Z. S.; Wazzan, N. DFT calculations of 1H- and 13C-NMR chemical shifts of 3-methyl-1-phenyl-4-(phenyldiazenyl)-1H-pyrazol-5-amine in solution. *Sci. Rep.* **2022**, *12*, 17798.

(8) Lauro, G.; Das, P.; Riccio, R.; Reddy, D. S.; Bifulco, G. DFT/NMR Approach for the Configuration Assignment of Groups of Stereoisomers by the Combination and Comparison of Experimental and Predicted Sets of Data. *J. Org. Chem.* **2020**, *85*, 3297−3306.

(9) Yesiltepe, Y.; Nuñez, J. R.; Colby, S. M.; Thomas, D. G.; Borkum, M. I.; Reardon, P. N.; Washton, N. M.; Metz, T. O.; Teeguarden, J. G.; Govind, N.; Renslow, R. S. An automated framework for NMR chemical shift calculations of small organic molecules. *J. Cheminf.* **2018**, *10*, 52.

(10) Narayanan, R.; Nakada, M.; Abe, M.; Saito, M.; Hada, M. 13C and 207Pb NMR Chemical Shifts of Dirhodio- and Dilithioplumbole Complexes: A Quantum Chemical Assessment. *Inorg. Chem.* **2019**, *58*, 14708−14719.

(11) Taylor, M. J.; Coakley, E. J.; Coles, M. P.; Cox, H.; Fulton, J. R. *β*-Diketiminato Organolead Complexes: Structures, $^{207}$Pb NMR, and Hammett Correlations. *Organometallics* **2015**, *34*, 2515−2521.

(12) Rodriguez-Fortea, A.; Alemany, P.; Ziegler, T. Density Functional Calculations of NMR Chemical Shifts with the Inclusion of Spin-Orbit Coupling in Tungsten and Lead Compounds. *J. Phys. Chem. A* **1999**, *103*, 8288−8294.

(13) McCrea-Hendrick, M. L.; Bursch, M.; Gullett, K. L.; Maurer, L. R.; Fettinger, J. C.; Grimme, S.; Power, P. P. Counterintuitive Interligand Angles in the Diaryls E{C$_6$H$_3$-2,6-(C$_6$H$_2$-2,4,6-$^{\text{<i>i</i>}}$Pr$_3$)$_2$}$_2$ (E = Ge, Sn, or Pb) and Related Species: The Role of London Dispersion Forces. *Organometallics* **2018**, *37*, 2075−2085.

(14) Kawamura, T.; Abe, M.; Saito, M.; Hada, M. Quantum-chemical analyses of aromaticity, UV spectra, and NMR chemical shifts in plumbacyclopentadienylidenes stabilized by Lewis bases. *J. Comput. Chem.* **2014**, *35*, 847−853.

(15) Li, A.; Wang, Y.; Downing, D. O.; Chen, F.; Zavalij, P.; Muñoz-Castro, A.; Eichhorn, B. W. Endohedral Plumbaspherenes of the Group 9 Metals: Synthesis, Structure and Properties of the [M@Pb 12 ] 3- (M = Co, Rh, Ir) Ions. *Chem.—Eur. J.* **2020**, *26*, 5824−5833.

(16) Dybowski, C.; Gabuda, S.; Kozlova, S.; Neue, G.; Perry, D.; Terskikh, V. Correlation and Relativistic Effects in *β*-PbO and Other Lead (II) Oxides: A Quantum ab Initio Explanation of 207Pb NMR and XANES Spectra. *J. Solid State Chem.* **2001**, *157*, 220−224.

(17) Dmitrenko, O.; Bai, S.; Dybowski, C. Prediction of 207Pb NMR parameters for the solid ionic lead(II) halides using the relativistic ZORA-DFT formalism: Comparison with the lead-containing molecular systems. *Solid State Nucl. Magn. Reson.* **2008**, *34*, 186−190.

(18) Dmitrenko, O.; Bai, S.; Beckmann, P. A.; van Bramer, S.; Vega, A. J.; Dybowski, C. The Relationship between 207 Pb NMR Chemical Shift and Solid-State Structure in Pb(II) Compounds. *J. Phys. Chem. A* **2008**, *112*, 3046−3052.

(19) Alkan, F.; Dybowski, C. Chemical-shift tensors of heavy nuclei in network solids: a DFT/ZORA investigation of 207 Pb chemical-shift tensors using the bond-valence method. *Phys. Chem. Chem. Phys.* **2015**, *17*, 25014−25026.

(20) Batista, P. R.; Ducati, L. C.; Autschbach, J. Solvent effect on the 195 Pt NMR properties in pyridonate-bridged Pt III dinuclear complex derivatives investigated by ab initio molecular dynamics and localized orbital analysis. *Phys. Chem. Chem. Phys.* **2021**, *23*, 12864−12880.

(21) Adrjan, B.; Makulski, W.; Jackowski, K.; Demissie, T. B.; Ruud, K.; Antušek, A.; Jaszuński, M. NMR absolute shielding scale and nuclear magnetic dipole moment of 207 Pb. *Phys. Chem. Chem. Phys.* **2016**, *18*, 16483−16490.

(22) Vícha, J.; Novotný, J.; Komorovsky, S.; Straka, M.; Kaupp, M.; Marek, R. Relativistic Heavy-Neighbor-Atom Effects on NMR Shifts: Concepts and Trends across the Periodic Table. *Chem. Rev.* **2020**, *120*, 7065−7103.

(23) Hess, B. A. Relativistic electronic-structure calculations employing a two-component no-pair formalism with external-field projection operators. *Phys. Rev. A: At., Mol., Opt. Phys.* **1986**, *33*, 3742−3748.

(24) Jansen, G.; Hess, B. A. Revision of the Douglas-Kroll transformation. *Phys. Rev. A: At., Mol., Opt. Phys.* **1989**, *39*, 6016−6017.

(25) Lenthe, E. v.; Baerends, E. J.; Snijders, J. G. Relativistic Regular Two-component Hamiltonians. *J. Chem. Phys.* **1993**, *99*, 4597−4610.

(26) Wolff, S. K.; Ziegler, T.; van Lenthe, E.; Baerends, E. J. Density Functional Calculations of Nuclear Magnetic Shieldings Using the Zeroth-Order Regular Approximation (ZORA) for Relativistic

Effects: ZORA Nuclear Magnetic Resonance. *J. Chem. Phys.* **1999**, *110*, 7689−7698.

(27) Kutzelnigg, W.; Liu, W. Quasirelativistic theory equivalent to fully relativistic theory. *J. Chem. Phys.* **2005**, *123*, 241102.

(28) Liu, W.; Peng, D. Infinite-order quasirelativistic density functional method based on the exact matrix quasirelativistic theory. *J. Chem. Phys.* **2006**, *125*, 044102.

(29) Iliaš, M.; Saue, T. An infinite-order two-component relativistic Hamiltonian by a simple one-step transformation. *J. Chem. Phys.* **2007**, *126*, 064102.

(30) Peng, D.; Liu, W.; Xiao, Y.; Cheng, L. Making four- and two-component relativistic density functional methods fully equivalent based on the idea of "from atoms to molecule. *J. Chem. Phys.* **2007**, *127*, 104106.

(31) Peng, D.; Middendorf, N.; Weigend, F.; Reiher, M. An efficient implementation of two-component relativistic exact-decoupling methods for large molecules. *J. Chem. Phys.* **2013**, *138*, 184105.

(32) Bursch, M.; Gasevic, T.; Stückrath, J. B.; Grimme, S. Comprehensive Benchmark Study on the Calculation of $^{29}$Si NMR Chemical Shifts. *Inorg. Chem.* **2021**, *60*, 272−285.

(33) Stückrath, J. B.; Gasevic, T.; Bursch, M.; Grimme, S. Benchmark Study on the Calculation of $^{119}$Sn NMR Chemical Shifts. *Inorg. Chem.* **2022**, *61*, 3903−3917.

(34) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding quantum chemistry methods. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2020**, *11*, No. e01493.

(35) *Semiempirical Extended Tight-Binding Program Package xtb*, Version 6.4.1; Universität Bonn Mulliken Center for Theoretical Chemistry: Bonn, Germany, 2020, https://github.com/grimme-lab/xtb/releasewebs.

(36) *TURBOMOLE 7.6*; Universität Karlsruhe & Forschungszentrum Karlsruhe GmbH: Karlsruhe, Germany, 2020, https://www.turbomole.org/.

(37) 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.

(38) Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. Turbomole. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 91−100.

(39) *AMS 2023.103, SCM, Theoretical Chemistry*; Vrije Universiteit: Amsterdam, The Netherlands, 2020, http://www.scm.com/.

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

(41) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB − An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652−1671.

(42) Ehlert, S.; Stahn, M.; Spicher, S.; Grimme, S. Robust and Efficient Implicit Solvation Model for Fast Semiempirical Methods. *J. Chem. Theory Comput.* **2021**, *17*, 4250−4261.

(43) 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.

(44) *Conformer-Rotamer Ensemble Sampling Tool CREST*, Version 2.11; Universität Bonn, Mulliken Center for Theoretical Chemistry: Bonn, Germany, 2021, https://github.com/grimme-lab/crest/releasewebs.

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

(46) *Commandline Energetic Sorting of Conformer-Rotamer Ensembles CENSO*, Version 1.2.0; Universität Bonn; Mulliken Center for Theoretical Chemistry: Bonn, Germany, 2021, https://github.com/grimme-lab/censo/releasewebs.

(47) 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.

(48) Grimme, S.; Hansen, A.; Ehlert, S.; Mewes, J.-M. r$^2$SCAN-3c: A "Swiss army knife" composite electronic-structure method. *J. Chem. Phys.* **2021**, *154*, 064103.

(49) Klamt, A.; Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans.* **1993**, *2*, 799−805.

(50) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99*, 2224−2235.

(51) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. Refinement and Parametrization of COSMO-RS. *J. Phys. Chem. A* **1998**, *102*, 5074−5085.

(52) Eckert, F.; Klamt, A. Fast solvent screening via quantum chemistry: COSMO-RS approach. *AIChE J.* **2002**, *48*, 369−385.

(53) Spicher, S.; Grimme, S. Efficient Computation of Free Energy Contributions for Association Reactions of Large Molecules. *J. Phys. Chem. Lett.* **2020**, *11*, 6606−6611.

(54) Grimme, S. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chem.—Eur. J.* **2012**, *18*, 9955−9964.

(55) Pye, C. C.; Ziegler, T. An implementation of the conductor-like screening model of solvation within the Amsterdam density functional package. *Theor. Chem. Acc.* **1999**, *101*, 396−408.

(56) van Lenthe, E.; Snijders, J. G.; Baerends, E. J. The Zero-order Regular Approximation for Relativistic Effects: The Effect of Spin−Orbit Coupling in Closed Shell Molecules. *J. Chem. Phys.* **1996**, *105*, 6505−6516.

(57) Van Lenthe, E.; Baerends, E. J. Optimized Slater-type basis sets for the elements 1−118. *J. Comput. Chem.* **2003**, *24*, 1142−1156.

(58) Neese, F. Software update: The ORCA program system—Version 5.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2022**, *12*, No. e1606.

(59) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA quantum chemistry program package. *J. Chem. Phys.* **2020**, *152*, 224108.

(60) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995−2001.

(61) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.

(62) Pantazis, D. A.; Chen, X.-Y.; Landis, C. R.; Neese, F. All-Electron Scalar Relativistic Basis Sets for Third-Row Transition Metal Atoms. *J. Chem. Theory Comput.* **2008**, *4*, 908−919.

(63) Pantazis, D. A.; Neese, F. All-electron scalar relativistic basis sets for the 6p elements. *Theor. Chem. Acc.* **2012**, *131*, 1292.

(64) 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.

(65) Furness, J. W.; Kaplan, A. D.; Ning, J.; Perdew, J. P.; Sun, J. Accurate and Numerically Efficient r$^2$SCAN Meta-Generalized Gradient Approximation. *J. Phys. Chem. Lett.* **2020**, *11*, 8208−8215.

(66) 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*, 142.

(67) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J. Chem. Phys.* **2003**, *119*, 12129−12137.

(68) Bursch, M.; Neugebauer, H.; Ehlert, S.; Grimme, S. Dispersion corrected r$^2$SCAN based global hybrid functionals: r$^2$SCANh, r$^2$SCAN0, and r$^2$SCAN50. *J. Chem. Phys.* **2022**, *156*, 134105.

(69) 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.

(70) Mardirossian, N.; Head-Gordon, M. ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **2016**, *144*, 214110.

(71) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.

(72) Zhang, Y.; Yang, W. Comment on "Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1998**, *80*, 890.

(73) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098−3100.

(74) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.

(75) Handy, N. C.; Cohen, A. J. Left-right correlation energy. *Mol. Phys.* **2001**, *99*, 403−412.

(76) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098−3100.

(77) Perdew, J. P. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 8822−8824.

(78) Adamo, C.; Barone, V. Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models. *J. Chem. Phys.* **1998**, *108*, 664−675.

(79) Keal, T. W.; Tozer, D. J. The Exchange-Correlation Potential in Kohn−Sham Nuclear Magnetic Resonance Shielding Calculations. *J. Chem. Phys.* **2003**, *119*, 3015−3024.

(80) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110*, 6158−6170.

(81) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.

(82) Becke, A. D. Density-functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(83) 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. Chem.* **1994**, *98*, 11623−11627.

(84) Gasevic, T.; Stückrath, J. B.; Grimme, S.; Bursch, M. Optimization of the r$^2$SCAN-3c Composite Electronic-Structure Method for Use with Slater-Type Orbital Basis Sets. *J. Phys. Chem. A* **2022**, *126*, 3826−3838.

(85) Bursch, M.; Mewes, J. M.; Hansen, A.; Grimme, S. Best-Practice DFT Protocols for Basic Molecular Computational Chemistry. *Angew. Chem.* **2022**, *134*, No. e202205735.

(86) Gao, P.; Wang, X.; Huang, Z.; Yu, H. 11 B NMR Chemical Shift Predictions via Density Functional Theory and Gauge-Including Atomic Orbital Approach: Applications to Structural Elucidations of Boron-Containing Molecules. *ACS Omega* **2019**, *4*, 12385−12392.

(87) Truflandier, L. A.; Autschbach, J. Probing the Solvent Shell with 195 Pt Chemical Shifts: Density Functional Theory Molecular Dynamics Study of Pt II and Pt IV Anionic Complexes in Aqueous Solution. *J. Am. Chem. Soc.* **2010**, *132*, 3472−3483.

(88) Schattenberg, C. J.; Lehmann, M.; Bühl, M.; Kaupp, M. Systematic Evaluation of Modern Density Functional Methods for the Computation of NMR Shifts of 3d Transition-Metal Nuclei. *J. Chem. Theory Comput.* **2022**, *18*, 273−292.

(89) Franzke, Y. J.; Holzer, C. Exact two-component theory becoming an efficient tool for NMR shieldings and shifts with spin−orbit coupling. *J. Chem. Phys.* **2023**, *159*, 159.

(90) Bast, R.; Gomes, A. S. P.; Saue, T.; Visscher, L.; Jensen, H. J. A., Aucar, I. A.; Bakken, V.; Chibueze, C.; Creutzberg, J.; Dyall, K. G.; Dubillard, S.; Ekström, U.; Eliav, E.; Enevoldsen, T.; Faßhauer, E.; Fleig, T.; Fossgaard, O.; Halbert, L.; Hedegård, E. D.; Helgaker, T.; Helmich−Paris, B.; Henriksson, J.; van Horn, M.; Iliaš, M.; Jacob, C. R.; Knecht, S.; Komorovský, S.; Kullie, O.; Lærdahl, J. K.; Larsen, C. V.; Lee, Y. S.; List, N. H.; Nataraj, H. S.; Nayak, M. K.; Norman, P.; Nyvang, A.; Olejniczak, G.; Olsen, J.; Olsen, J. M. N.; Papadopoulos, A.; Park, Y. C.; Pedersen, J. K.; Pernpointner, M.; Pototschnig, J. V.; di Remigio, R.; Repisky, M.; Ruud, K.; Sałek, P.; Schimmelpfennig, B.; Senjean, B.; Shee, A.; Sikkema, J.; Sunaga, A.; Thorvaldsen, A. J.; Thyssen, J.; van Stralen, J.; Vidal, M. L.; Villaume, S.; Visser, O.; Winther, T.; Yamamoto, S.; Yuan, X. *DIRAC, a Relativistic Ab Initio Electronic Structure Program, Release DIRAC23*, 2023. 10.5281/zenodo.7670749, https://www.diracprogram.org.

(91) Repisky, M.; Komorovsky, S.; Kadek, M.; Konecny, L.; Ekström, U.; Malkin, E.; Kaupp, M.; Ruud, K.; Malkina, O. L.; Malkin, V. G. ReSpect: Relativistic spectroscopy DFT program package. *J. Chem. Phys.* **2020**, *152*, 152.

(92) Hino, S.; Olmstead, M.; Phillips, A. D.; Wright, R. J.; Power, P. P. Terphenyl Ligand Stabilized Lead(II) Derivatives: Steric Effects and Lead-Lead Bonding in Diplumbenes. *Inorg. Chem.* **2004**, *43*, 7346−7352.

(93) Eaborn, C.; Ganicz, T.; Hitchcock, P. B.; Smith, J. D.; Sözerli, S. E. A Novel Organolead(II) Species, the Plumbacycloalkane Derivative $[CH_2SiMe_2C(SiMe_3)_2PbC(SiMe_3)_2SiMe_2CH_2]$. *Organometallics* **1997**, *16*, 5621−5622.

(94) Stürmann, M.; Weidenbruch, M.; Klinkhammer, K. W.; Lissner, F.; Marsmann, H. New Plumbylenes and a Plumbylene Dimer with a Short Lead-Lead Separation. *Organometallics* **1998**, *17*, 4425−4428.

(95) Weiß, S.; Schubert, H.; Wesemann, L. Low valent lead hydride chemistry: hydroplumbylation of phenylacetylene and 1,1-dimethylallene. *Chem. Commun.* **2019**, *55*, 10238−10240.

(96) Hino, S.; Brynda, M.; Phillips, A. D.; Power, P. P. Synthesis and Characterization of a Quasi-One-Coordinate Lead Cation. *Angew. Chem., Int. Ed.* **2004**, *43*, 2655−2658.

(97) Rekken, B. D.; Brown, T. M.; Fettinger, J. C.; Lips, F.; Tuononen, H. M.; Herber, R. H.; Power, P. P. Dispersion Forces and Counterintuitive Steric Effects in Main Group Molecules: Heavier Group 14 (Si−Pb) Dichalcogenolate Carbene Analogues with Sub-90° Interligand Bond Angles. *J. Am. Chem. Soc.* **2013**, *135*, 10134−10148.

(98) Charmant, J. P. H.; Haddow, M. F.; Hahn, F. E.; Heitmann, D.; Fröhlich, R.; Mansell, S. M.; Russell, C. A.; Wass, D. F. Syntheses and molecular structures of some saturated N-heterocyclic plumbylenes. *Dalton Trans.* **2008**, 6055.

(99) Chen, X.; Gamer, M. T.; Roesky, P. W. Synthesis and structural characterization of arsinoamides − early transition metal (Zr and Hf) and main group metal (Al, In, Sn, and Pb) complexes. *Dalton Trans.* **2019**, *48*, 15207−15211.

(100) Dean, P. A. W.; Vittal, J. J.; Payne, N. C. Discrete trigonal-pyramidal lead(II) complexes: syntheses and x-ray structure analyses of $[(C_6H_5)_4As]$ $[Pb(EC_6H_5)_3]$ (E = S, Se). *Inorg. Chem.* **1984**, *23*, 4232−4236.

(101) Taylor, M. J.; Coakley, E. J.; Coles, M. P.; Cox, H.; Fulton, J. R. β-Diketiminato Organolead Complexes: Structures, $^{207}$Pb NMR, and Hammett Correlations. *Organometallics* **2015**, *34*, 2515−2521.

(102) Someşan, A. A.; Le Coz, E.; Roisnel, T.; Silvestru, C.; Sarazin, Y. Stable lead(II) boroxides. *Chem. Commun.* **2018**, *54*, 5299−5302.

(103) Merrill, W. A.; Wright, R. J.; Stanciu, C. S.; Olmstead, M. M.; Fettinger, J. C.; Power, P. P. Synthesis and Structural Characterization of a Series of Dimeric Metal(II) Imido Complexes $\{M(\mu\text{-NAr}^{\#})\}_2$ [M = Ge, Sn, Pb; Ar$^{\#}$ = $C_6H_3$-2,6-$(C_6H_2\text{-}2,4,6\text{-}Me_3)_2$] and the Related Monomeric Primary Amido Derivatives MN(H)Ar$^{\#}$2 (M = Ge, Sn, Pb): Spectroscopic Manifestations of Secondary Metal-Ligand Interactions. *Inorg. Chem.* **2010**, *49*, 7097−7105.

(104) Saito, M.; Nakada, M.; Kuwabara, T.; Minoura, M. A Reversible Two-Electron Redox System Involving a Divalent Lead Species. *Chem. Commun.* **2015**, *51*, 4674−4676.

(105) Wang, L.; Fadlallah, S.; Bellini, C.; Orione, C.; Dorcet, V.; Carpentier, J.-F.; Sarazin, Y. Structurally Characterized Lead(II) Alkoxides as Potent Ring-Opening Polymerization Catalysts. *Organometallics* **2015**, *34*, 1321−1327.

(106) Hinz, A. Pseudo-One-Coordinate Tetrylenium Salts Bearing a Bulky Carbazolyl Substituent. *Chem.—Eur. J.* **2019**, *25*, 3267−3271.

(107) Chia, S.-P.; Xi, H.-W.; Li, Y.; Lim, K. H.; So, C.-W. A Base-Stabilized Lead(I) Dimer and an Aromatic Plumbylidenide Anion. *Angew. Chem., Int. Ed.* **2013**, *52*, 6298−6301.

(108) Nakada, M.; Kuwabara, T.; Furukawa, S.; Hada, M.; Minoura, M.; Saito, M. Synthesis and reactivity of a ruthenocene-type complex bearing an aromatic $\pi$-ligand with the heaviest group 14 element. *Chem. Sci.* **2017**, *8*, 3092−3097.

(109) Rupprecht, S.; Franklin, S. J.; Raymond, K. N. Synthesis of monothiohydroxamic ligands and their lead complexes. Structures of N-methyl-3-pyridothiohydroxamic acid, bis(N-methyl-3-pyridothiohydroxamato) lead(II) and bis(N-cyclohexyl-phenylacetothiohydroxamato) lead(II). *Inorg. Chim. Acta* **1995**, *235*, 185−194.

(110) Stabenow, F.; Saak, W.; Marsmann, H.; Weidenbruch, M. Hexaarylcyclotriplumbane: A Molecule with a Homonuclear Ring System of Lead. *J. Am. Chem. Soc.* **2003**, *125*, 10172−10173.

(111) Zaitsev, K. V.; Cherepakhin, V. S.; Churakov, A. V.; Peregudov, A. S.; Tarasevich, B. N.; Egorov, M. P.; Zaitseva, G. S.; Karlov, S. S. Extending the family of stable heavier carbenes: New tetrylenes based on N, N, O-ligands. *Inorg. Chim. Acta* **2016**, *443*, 91−100.

(112) Müller, T.; Bauch, C.; Bolte, M.; Auner, N. Unique Coordination of Two C = C Double Bonds to an Electron-Deficient Lead Center. *Chem.—Eur. J.* **2003**, *9*, 1746−1749.

(113) Pop, A.; Wang, L.; Dorcet, V.; Roisnel, T.; Carpentier, J.-F.; Silvestru, A.; Sarazin, Y. On the coordination chemistry of organochalcogenolates $R^{NMe2}$ ceE- and $R^{NMe2}EO^-$ (E = S, Se) onto lead(ii) and lighter divalent tetrel elements. *Dalton Trans.* **2014**, *43*, 16459−16474.

(114) Müller, T.; Bauch, C.; Ostermeier, M.; Bolte, M.; Auner, N. Norbornyl Cations of Group 14 Elements. *J. Am. Chem. Soc.* **2003**, *125*, 2158−2168.

(115) Kennedy, J. D.; McFarlane, W.; Pyne, G. S. Lead-207 shielding in organolead compounds. *J. Chem. Soc. Dalton Trans.* **1977**, 2332.

(116) Kleiner, N.; Drager, M. Über gemischte Bindungen in der IV. Hauptgruppe, II [1]. Hexacyclohexylethan-Analoga $Pb_2$(c− Hex)$_6$, (c− Hex)$_3$Pb−Sn(c− Hex)$_3$ und $Sn_2$(c− Hex)$_6$. *Z. Naturforsch. B* **1985**, *40*, 477−483.

(117) Whittaker, S. M.; Cervantes-Lee, F.; Pannell, K. H. Synthesis, Single-Crystal X-ray Structural Characterization, and Reactivity of a Hexaalkyldiplumbane, $TsiMe_2Pb$-$PbMe_2Tsi$, Tsi = $(Me_3Si)_3C$. *Inorg. Chem.* **1994**, *33*, 6406−6408.

(118) Kano, N.; Tokitoh, N.; Okazaki, R. Synthesis and X-ray Crystal Structure of Bis{2,4,6- tris[bis(trimethylsilyl)methyl]phenyl}-dibromoplumbane: The First Monomeric Diorganodihaloplumbane in the Crystalline State. *Organometallics* **1997**, *16*, 2748−2750.

(119) Schneider-Koglin, C.; Behrends, K.; Dräger, M. Über gemischte Gruppe 14-Gruppe 14-Bindungen: VI. Hexa-o-tolylethan-analoga o-$Tol_6Sn_2$, o-$Tol_6PbSn$ und o-$Tol_6Pb_2$: ein Vergleich von Bindungsstärke und Polarität in der Reihung Sn-Sn, Pb-Sn, Pb-Pb. *J. Organomet. Chem.* **1993**, *448*, 29−38.

(120) Munguia, T.; Pavel, I. S.; Kapoor, R. N.; Cervantes-Lee, F.; Párkányi, L.; Pannell, K. H. Lewis acidity of group 14 elements toward intramolecular sulfur in ortho-aryl-thioanisoles. *Can. J. Chem.* **2003**, *81*, 1388−1397.

(121) Rodríguez, A.; Sousa-Pedrares, A.; García-Vázquez, J. A.; Romero, J.; Sousa, A.; Russo, U. Synthesis and Structural Characterisation of Diorganotin(IV) and Diphenyllead(IV) Complexes of Pyrimidine-2-thionate Derivatives. *Eur. J. Inorg. Chem.* **2007**, *2007*, 1444−1456.

(122) Peña-Hueso, A.; Esparza-Ruiz, A.; Ramos-García, I.; Flores-Parra, A.; Contreras, R. Triphenyl lead, tin and germanium coordination compounds derived from 9H-3-thia-1,4a,9-triaza-fluorene-2,4-dithione. *J. Organomet. Chem.* **2008**, *693*, 492−504.

(123) Peveling, K.; Schürmann, M.; Jurkschat, K. The First [4 + 2]-Coordinated Tetraorganolead Compound: Synthesis, Structure and Conversion into a Triorganolead Cation, a Benzoxaphosphaplumbole, and Diorganolead Dihalides. *Z. Anorg. Allg. Chem.* **2002**, *628*, 2435−2442.

(124) Pettinari, C. NMR Spectroscopy, Heteronuclei, Ge, Sn, Pb. *Encyclopedia of Spectroscopy and Spectrometry*; Elsevier, 2017; pp 330−341.

(125) Klapötke, T. M.; Krumm, B.; Niemitz, M.; Polborn, K.; Rienäcker, C. M. Tetrakis(pentafluorophenyl)lead(IV): crystal structure, semiempirical PM3 calculations and NMR studies. *J. Fluorine Chem.* **2000**, *104*, 129−133.

(126) Buston, J. E.; Compton, R. G.; Leech, M. A.; Moloney, M. G. On the structure and reaction with pyridine of o-methoxyphenyllead acetates. *J. Organomet. Chem.* **1999**, *585*, 326−330.

(127) Morsali, A. Syntheses and Characterization of Two New Lead(II) Acetate Complexes, Pb(L)(CH 3 COO) 2, L = 2,2′:6′,2″-Terpyridine (tpy) and 2,4,6-Tris(2-pyridyl)- 1,3,5-Triazine (trz), Crystal Structure of Pb(tpy)(CH3COO)2. *Z. Naturforsch. B* **2004**, *59*, 1039−1044.

(128) Herberich, G. E.; Zheng, X.; Rosenplänter, J.; Englert, U. Borabenzene Derivatives. 30. Bis(1-methylboratabenzene) Compounds of Germanium, Tin, and Lead. First Structural Characterization of Facial Bonding of a Boratabenzene to a p-Element and the Structures of $Pb(C_5H_5BMe)_2$ and Its 2,2‘-Bipyridine Adduct. *Organometallics* **1999**, *18*, 4747−4752.

(129) Constantine, S. P.; Cox, H.; Hitchcock, P. B.; Lawless, G. A. Parallel Metallocenes of Germanium, Tin, and Lead. *Organometallics* **2000**, *19*, 317−326.

(130) Wirtz, L.; Jourdain, M.; Huch, V.; Zimmer, M.; Schäfer, A. Synthesis, Structure, and Reactivity of Disiloxa[3]tetrelocenophanes. *ACS Omega* **2019**, *4*, 18355−18360.

(131) Janiak, C.; Schumann, H.; Stader, C.; Wrackmeyer, B.; Zuckerman, J. J. Decaphenylgermanocen, -stannocen und -plumbocen sowie Pentaphenylstannocen: Synthese, Eigenschaften und CPMAS-Metall-NMR-Messungen. *Chem. Ber.* **1988**, *121*, 1745−1751.

(132) Alkan, F.; Dybowski, C. Effect of Co-Ordination Chemistry and Oxidation State on the 207 Pb Magnetic-Shielding Tensor: A DFT/ZORA Investigation. *J. Phys. Chem. A* **2016**, *120*, 161−168.

(133) Latypov, S. K.; Polyancev, F. M.; Yakhvarov, D. G.; Sinyashin, O. G. Quantum Chemical Calculations of $^{31}$P NMR Chemical Shifts: Scopes and Limitations. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6976−6987.

(134) Rablen, P. R.; Pearlman, S. A.; Finkbiner, J. A Comparison of Density Functional Methods for the Estimation of Proton Chemical Shifts with Chemical Accuracy. *J. Phys. Chem. A* **1999**, *103*, 7357−7363.

(135) Jain, R.; Bally, T.; Rablen, P. R. Calculating Accurate Proton Chemical Shifts of Organic Molecules with Density Functional Methods and Modest Basis Sets. *J. Org. Chem.* **2009**, *74*, 4017−4023.

(136) Aliev, A. E.; Courtier-Murias, D.; Zhou, S. Scaling Factors for Carbon NMR Chemical Shifts Obtained from DFT B3LYP Calculations. *J. Mol. Struct.* **2009**, *893*, 1−5.

(137) Konstantinov, I. A.; Broadbelt, L. J. Regression Formulas for Density Functional Theory Calculated $^1$H and $^{13}$C NMR Chemical Shifts in Toluene-$D_8$. *J. Phys. Chem. A* **2011**, *115*, 12364−12372.

(138) Ando, I. Some aspects of the NMR chemical shift/structure correlation in the structural characterization of polymers and biopolymers. *Polym. J.* **2012**, *44*, 734−747.

(139) 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.

(140) Willoughby, P. H.; Jansma, M. J.; Hoye, T. R. A guide to small-molecule structure assignment through computation of (1H and 13C) NMR chemical shifts. *Nat. Protoc.* **2014**, *9*, 643−660.

(141) Pyykko, P. Relativistic effects in structural chemistry. *Chem. Rev.* **1988**, *88*, 563−594.

(142) Kleine Büning, J. B.; Grimme, S.; Bursch, M. Machine learning-based correction for spin−orbit coupling effects in NMR chemical shift calculations. *Phys. Chem. Chem. Phys.* **2024**, *26*, 4870−4884.

# APPENDIX C

# The $p$-block challenge: assessing quantum chemistry methods for inorganic heterocycle dimerizations

Thomas Gasevic,[†] Markus Bursch,[‡§] Qianli Ma,[¶] Stefan Grimme[†], Hans-Joachim Werner,[¶] and Andreas Hansen[†]

*Received: December 21, 2023*
*First published: April 11, 2024*

Reprinted (adapted) with permission from:
T. Gasevic, M. Bursch, Q. Ma, S. Grimme, H.-J. Werner, and A. Hansen, *The p-block challenge: assessing quantum chemistry methods for inorganic heterocycle dimerizations*, Phys. Chem. Chem. Phys. **26**.18 (2024) 13884, DOI: `10.1039/d3cp06217a`
T. Gasevic, M. Bursch, Q. Ma, S. Grimme, H.-J. Werner, and A. Hansen, *Correction: The p-block challenge: assessing quantum chemistry methods for inorganic heterocycle dimerizations*, Phys. Chem. Chem. Phys. **27**.16 (2025) 8572, DOI: `10.1039/d5cp90062j`

**Own contributions**

- Data Curation & Investigation – Performing DFT calculations
- Formal Analysis
- Software – Writing Python-based code for the generation of dimer structures
- Writing & Visualization

---

[†] Mulliken Center for Theoretical Chemistry, Universität Bonn, Beringstr. 4, D-53115 Bonn, Germany
[‡] Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, D-45470 Mülheim an der Ruhr, Germany
[§] FACCTs GmbH, D-50677, Köln, Germany
[¶] Institut für Theoretische Chemie, Universität Stuttgart, Pfaffenwaldring 55, D-70569 Stuttgart, Germany

PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2024, **26**, 13884

# The p-block challenge: assessing quantum chemistry methods for inorganic heterocycle dimerizations†

Thomas Gasevic, ‡[c] Markus Bursch, ‡*[ad] Qianli Ma, [b] Stefan Grimme, [c] Hans-Joachim Werner ‡*[b] and Andreas Hansen ‡*[c]

Received 21st December 2023, Accepted 11th April 2024

DOI: 10.1039/d3cp06217a

rsc.li/pccp

The elements of the p-block of the periodic table are of high interest in various chemical and technical applications like frustrated Lewis-pairs (FLP) or opto-electronics. However, high-quality benchmark data to assess approximate density functional theory (DFT) for their theoretical description are sparse. In this work, we present a benchmark set of 604 dimerization energies of 302 "inorganic benzenes" composed of all non-carbon p-block elements of main groups III to VI up to polonium. This so-called *IHD302* test set comprises two classes of structures formed by covalent bonding and by weaker donor–acceptor (WDA) interactions, respectively. Generating reliable reference data with *ab initio* methods is challenging due to large electron correlation contributions, core–valence correlation effects, and especially the slow basis set convergence. To compute reference values for these dimerization reactions, after thorough testing, we applied a computational protocol using state-of-the-art explicitly correlated local coupled cluster theory termed PNO-LCCSD(T)-F12/cc-VTZ-PP-F12(corr.). It includes a basis set correction at the PNO-LMP2-F12/aug-cc-pwCVTZ level. Based on these reference data, we assess 26 DFT methods in combination with three different dispersion corrections and the def2-QZVPP basis set, five composite DFT approaches, and five semi-empirical quantum mechanical methods. For the covalent dimerizations, the $r^2$SCAN-D4 *meta*-GGA, the $r^2$SCAN0-D4 and $\omega$B97M-V hybrids, and the revDSD-PBEP86-D4 double-hybrid functional are found to be the best-performing methods among the evaluated functionals of the respective class. However, since def2 basis sets for the 4th period are not associated to relativistic pseudo-potentials, we obtained significant errors in the covalent dimerization energies (up to 6 kcal mol$^{-1}$) for molecules containing p-block elements of the 4th period. Significant improvements were achieved for systems containing 4th row elements by using ECP10MDF pseudopotentials along with re-contracted aug-cc-pVQZ-PP-KS basis sets introduced in this work with the contraction coefficients taken from atomic DFT (PBE0) calculations. Overall, the *IHD302* set represents a challenge to contemporary quantum chemical methods. This is due to a large number of spatially close p-element bonds which are underrepresented in other benchmark sets, and the partial covalent bonding character for the WDA interactions. The *IHD302* set may be helpful to develop more robust and transferable approximate quantum chemical methods in the future.

## 1 Introduction

In the last decades, cutting-edge chemical synthesis involving inorganic main group compounds covering large parts of the periodic table experienced a renaissance.[1] Compared to their organic counterparts, inorganic elements bring various challenges such as a plethora of possible bonding motifs and often more difficult electronic structures.[2] A very prominent example

---

[a] *Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany. E-mail: bursch@kofo.mpg.de*
[b] *Institut für Theoretische Chemie, Universität Stuttgart, Pfaffenwaldring 55, D-70569 Stuttgart, Germany. E-mail: werner@theochem.uni-stuttgart.de*
[c] *Mulliken Center for Theoretical Chemistry, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstr. 4, 53115 Bonn, Germany. E-mail: hansen@thch.uni-bonn.de*
[d] *FACCTs GmbH, 50677, Koeln, Germany*

† Electronic supplementary information (ESI) available: Supplementary definitions of statistical measures, further computational details and supplementary figures and tables: esi.pdf; Cartesian coordinates of all compounds: *IHD302*.tar.xz; additional data for the correlated calculations: data_correlation_methods.xlsx; Detailed DFT and SQM results: data_dft_sqm.xlsx; The aug-cc-pVQZ-PP-KS basis sets: avqz-pp-ks.basis; Molpro example input files: pno_input_examples_molpro2024.1.zip. See DOI: **https://doi.org/10.1039/d3cp06217a**

‡ These authors contributed equally.

of important and actively researched p-block chemistry is that of (frustrated) Lewis pairs (FLPs).[3,4] Moreover, rings and clusters of p-block elements are of high interest in opto-electronics[5–8] and as promising precursor materials for film composition.[9–11] p-Block elements are further discussed to be incorporated into polymers for property optimization.[12] In most of such compounds, the direct covalent bonds or the donor- and acceptor interactions of the p-elements are of key importance for their chemical properties and reactivity. Nevertheless, systems with heavier p-block elements are typically underrepresented in comprehensive thermochemistry databases like GMTKN55[13] or LP14,[14] even though more recent benchmark sets such as CHAL336[15] or the supramolecular HS13L[16] extend the evaluated chemical space in this respect. However, large organic substituents are often used in such sets to saturate the p-block elements, thus somewhat limiting the explicit insight into the respective interactions of the donor–acceptor pairs.

Chemically diverse benchmark sets are extremely important for evaluation and cross-checking of contemporary theoretical (approximate first principle quantum mechanical or atomistic) models that have become a valuable tool for understanding a wide range of chemically diverse systems. However, suitable reference data for heavier p-elements are rare with consequences on the development and performance of these methods. For example, density functional theory (DFT) methods, the workhorse of modern quantum chemistry, are typically designed to be mostly accurate for (bio)organic chemistry with less focus on inorganic compounds. Further, most semi-empirical quantum mechanical (SQM) methods are limited in their parameterization space and accuracy regarding such elements.[17–19] The problem of under-representation of chemical space becomes even more evident for the fast growing field of machine-learning (ML) techniques, which require huge amounts of carefully selected training data as well as thorough cross-validation.

To address this lack of reliable reference data for the interaction of (heavier) p-block elements, and inspired by the work of Frenking *et al.* on the dimerization of $[\mathrm{BAlGaNPAs}]\mathrm{H}_6$[20] and comparable works,[21,22] we compiled a new benchmark set of dimerization reactions termed *IHD302*, consisting of planar six-membered heterocyclic monomer structures composed purely of p-block elements from boron to polonium (excluding carbon). The so-called *Inorganic Heterocycle Dimerizations 302* (*IHD302*) set is divided into two subsets, covalently bound and weaker donor–acceptor interacting dimers. The latter structures can be best characterized as strongly bound van der Waals complexes, *i.e.*, non-equilibrium structures on a path to covalent bonding. This poses a particular challenge for mean-field electronic structure methods due to a strong interplay of covalent (short-range) electron correlation and London dispersion interactions, which are usually specially treated in DFT and SQM.

The generation of reliable reference data of high accuracy is indispensable, yet a challenge on its own due to the significant (core–valence) electron correlation effects in these molecules, which we discuss in detail in this work. This newly introduced benchmark set can be considered a particularly hard test for existing quantum chemical methods as well as a basis for their further development and improvement. Specifically, approximate SQM methods, such as PMx[17,23,24] and GFN*n*-xTB[19,25,26] or composite DFT schemes[27] may be improved based on the data provided by the *IHD302* benchmark set. In the course of this work, we assess the performance of 26 DFT functionals together with three different London dispersion corrections, five composite DFT approaches, and five SQM methods.

## 2 The *IHD302* benchmark set

The *IHD302* benchmark is composed of 302 neutral six-membered heterocycles and their respective non-covalently interacting and covalently bound dimers (*cf.* Fig. 1) in their singlet ground state. The monomers can be categorized into three main group element (E) combinations, $[\mathrm{E}_3^{\mathrm{III}}\mathrm{E}_3^{\mathrm{VI}}]\mathrm{H}_3$, $[\mathrm{E}_3^{\mathrm{III}}\mathrm{E}_3^{\mathrm{V}}]\mathrm{H}_6$, and $[\mathrm{E}_3^{\mathrm{IV}}\mathrm{E}_3^{\mathrm{V}}]\mathrm{H}_3$. These combinations were chosen inspired by

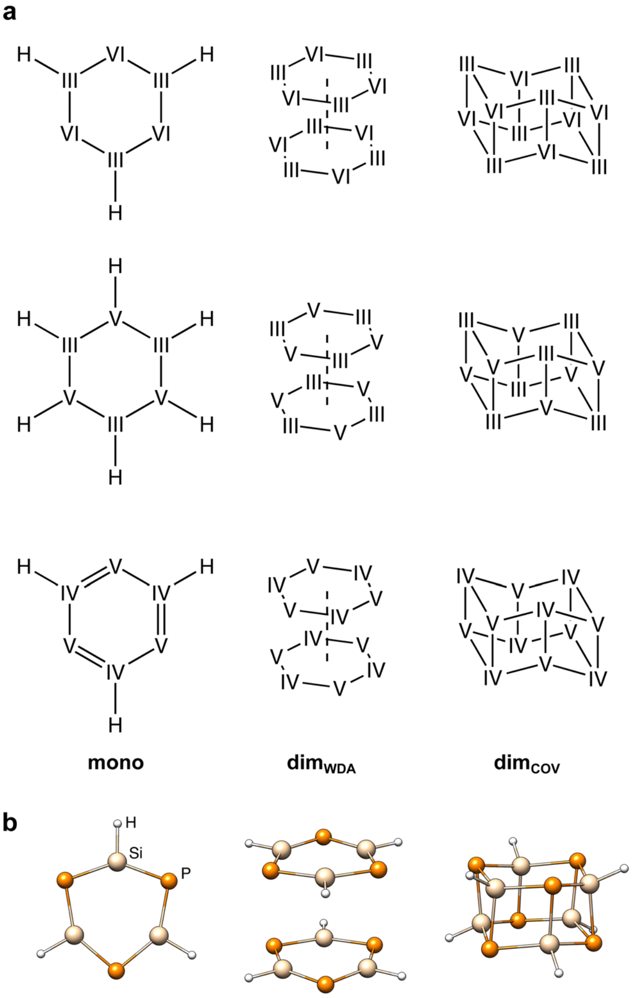


Fig. 1 (a) Simplified Lewis formulae of the investigated heterocycles and their respective dimers (hydrogen atoms omitted for clarity). Roman numbers depict the main group of the respective elements. III = B, Al, Ga, In, Tl; IV = Si, Ge, Sn, Pb; V = N, P, As, Sb, Bi, VI = O, S, Se, Te, Po; (b) exemplary depiction of the respective structures for $\overline{\mathrm{Si{-}P{-}Si{-}P{-}Si{-}P}}$.

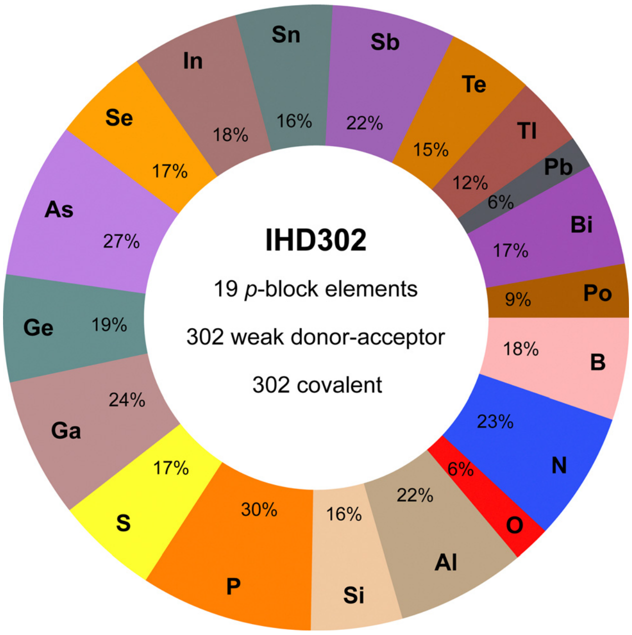


Fig. 2 Percent contribution of elements to the systems of the *IHD302* and key features of the data set (*e.g.*, 23% of the *IHD302* systems contain at least one nitrogen atom).

experimentally accessible parent "inorganic benzenes".[8] Other combinations involving, *e.g.*, V and VI would mainly result in non-planar environments and benchmark sets for non-covalent interactions of these element combinations such as the CHAL336[15] by Goerigk and co-workers already exist. All main group III, IV, V, and VI elements from B ($Z = 5$) to Po ($Z = 84$) except carbon were taken into account with an average of 53 compounds per element. Carbon was excluded as a typically saturated organic element with less pronounced donor–acceptor chemistry. The element composition of the benchmark set is illustrated in Fig. 2. Some elements (*e.g.* lead) are less represented as the respective elements strongly tend to leave the planar monomer structure. To keep comparability throughout the benchmark set, these monomer structures were discarded. The nomenclature of the model compounds $\overline{\text{A–B–C–D–E–F}}$ gives the ring atoms in clockwise order, while hydrogen atoms are omitted for clarity as their positions are clearly defined according to Fig. 1 (examples: $\overline{\text{Ga(H)–Te–In(H)–Te–Ga(H)–Se}}$ $\hat{=}$ $\overline{\text{Ga–Te–In–Te–Ga–Se}}$ $\hat{=}$ $[\text{GaTeInTeGaSe}]\text{H}_3$). Covalent and weak donor–acceptor dimers are further indicated by the subscripts "COV" and "WDA", respectively. To maximize the interaction surface in the non-covalently interacting dimers, only planar local minima after optimization of the monomers were considered. If necessary, these rings were then saturated with three or six hydrogen atoms depending on the involved p-block elements to avoid open-shell or highly ionic systems.

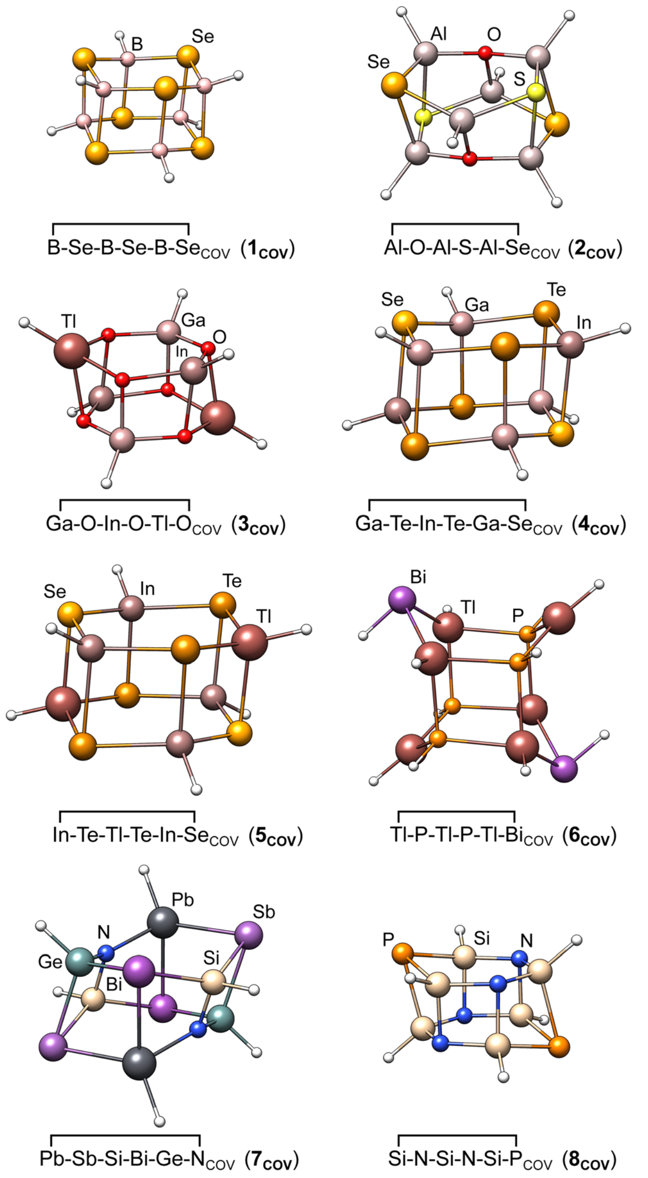
The monomeric heterocycles consist of formal single- or double-bonded atoms, depending on their atomic properties. Especially, combinations of light p-block elements, yield partly aromatic monomers as in the borazine case ($\overline{\text{B–N–B–N–B–N}}$). These structures are also expected to yield positive dimerization energies due to the stabilization of the monomers. All other monomers are best described by combinations of single-bonds and lone-pairs, as specifically for the heavier elements, the tendency to form multiple bonds decreases quickly descending the periodic table. As the dimerization of most heterocyclic combinations involving heavier elements strongly favor covalent dimerization, weakly donor–acceptor interacting dimers were generated from the planar monomer structures without further optimization. Here, the dimers were generated by 180° rotation (simplest approach to obtain motifs that are as symmetrical as possible and therefore do not form homoatomic artificial bonds) and a displacement of the ring center by twice the van der Waals radii[28,29] of the heaviest element involved. The covalent heterocyclic dimers were obtained by subsequent geometry optimization employing $r^2$SCAN-3c[30] as implemented in the ORCA program package.[31,32] $r^2$SCAN-3c was previously found to produce excellent structures.[30] Further, it yields very good energetic agreement with the high-level reference data as discussed in Section 4.1, supporting its suitability to reproduce the respective potential energy surface.

During optimization, most covalent dimers formed the archetypal crown-shaped dimer with six connecting bonds. Nevertheless, the three systems $\overline{\text{Tl–P–Tl–P–Tl–Bi}}_{\text{COV}}$, $\overline{\text{Pb–N–Pb–N–Pb–Sb}}_{\text{COV}}$, and $\overline{\text{Sn–P–Pb–P–Sn–As}}_{\text{COV}}$ optimized into a minimum with only four formed connecting bonds (*cf.* Fig. 3). The correspondingly large number of formed bonds and close interatomic contacts results in challenging chemical situations and allows a clear focus on the p-block element interactions. Although large correlation contributions are expected in these systems, this is dynamic correlation, which can be treated correctly with single-reference methods. To exclude multi-reference cases, we looked for spin symmetry breaking for all structures of the *IHD302* set following ref. 33 with unrestricted PBE0-D4, but found no hint for significant static correlation. In addition, we applied fractional occupation number weighted density (FOD)-based static correlation diagnostics[34–38] for eight test systems (*vide infra*), which also showed no evidence for multi-reference cases (see the ESI,† for details).

## 3 p-Block challenge 1: correlated wavefunction calculations

To provide reliable reference data, we used state-of-the-art explicitly correlated local coupled cluster theory, specifically PNO-LCCSD(T)-F12[39–45] with special consideration of core–valence correlation. To this end, also new features were implemented (details are given in the following subsections). All correlated wavefunction theory calculations reported in this work were carried out with the Molpro package of *ab initio* programs (release 2023.2).[46–48]

In principle, the wavefunction ansatz and the local approximations in the PNO-LCCSD(T)-F12 are rather similar as in the DLPNO-CCSD(T)$_{\overline{\text{F12}}}$ method of the Neese group.[49,50] A detailed



$\overline{\text{B-Se-B-Se-B-Se}}_{\text{cov}}$ (**1**$_{\textbf{cov}}$)

$\overline{\text{Al-O-Al-S-Al-Se}}_{\text{cov}}$ (**2**$_{\textbf{cov}}$)

$\overline{\text{Ga-O-In-O-Tl-O}}_{\text{cov}}$ (**3**$_{\textbf{cov}}$)

$\overline{\text{Ga-Te-In-Te-Ga-Se}}_{\text{cov}}$ (**4**$_{\textbf{cov}}$)

$\overline{\text{In-Te-Tl-Te-In-Se}}_{\text{cov}}$ (**5**$_{\textbf{cov}}$)

$\overline{\text{Tl-P-Tl-P-Tl-Bi}}_{\text{cov}}$ (**6**$_{\textbf{cov}}$)

$\overline{\text{Pb-Sb-Si-Bi-Ge-N}}_{\text{cov}}$ (**7**$_{\textbf{cov}}$)

$\overline{\text{Si-N-Si-N-Si-P}}_{\text{cov}}$ (**8**$_{\textbf{cov}}$)

Fig. 3 Molecular structures of the eight covalently bound heterocyclic dimers chosen as test systems for detailed method evaluation in Section 3.4.

comparison of both methods can be found in ref. 51. In particular, for F12 calculations, the PNO-LCCSD(T)-F12 program was found to be more efficient, robust, and accurate than the currently available DLPNO program, and therefore it has been used in the presented work. Alternative local correlation methods are, for example, the local natural orbital coupled cluster (LNO-CCSD(T)) methods of Nagy and Kallay,[52–55] or geminal-based electronic structure methods of Tecmer and Boguslawski,[56] but these methods have not been considered in the current work.

### 3.1 Special localization schemes

Proper orbital localization is essential for reliable local correlation methods. In the following, we distinguish inner-core, outer-core, and valence molecular orbitals (MOs). The electrons in the outer-core and the valence orbitals are correlated in the post-Hartree–Fock calculations discussed in this section. In most calculations of the current work, the outer-core orbitals only comprise the $(n-1)$d shells (with principle quantum number $n$) of p-block elements with $Z \geq 31$. In some test calculations, the $(n-1)$s,p orbitals were also correlated (*cf.* Section 3.4).

Valence orbitals are localized using the intrinsic bond orbital (IBO) method.[57] Core and outer core orbitals are localized using IBO(AO). This means that in the underlying Pipek–Mezey localization,[58] contributions of individual intrinsic atomic orbitals (IAOs) rather than all those belonging to an atom are maximized. This minimizes mixing of nearly degenerate core orbitals (*e.g.*, the different components of a p- or d-shell).

Before localization, the canonical orbitals are automatically resorted, so that inner-core orbitals come first, followed by the outer-core orbitals, and then the valence orbitals. This is achieved by computing the overlaps of the molecular orbitals with pure atomic core orbitals (stored in the basis set library) and ordering the orbitals such that the core part of the overlap matrix becomes closest to diagonal. This sort is necessary since the orbital energies of the outer-core orbitals are not always below those of all valence orbitals. The inner-core orbitals are then localized separately in order to avoid any mixing with correlated orbitals. In order to separate outer-core and valence orbitals as much as possible, we tested two different approaches, which are implemented in Molpro.

LOC_COREORB=SEP: in this case, outer-core and valence orbitals are localized separately, starting from the sorted canonical orbitals. The Fock matrix then has a block-diagonal structure, without couplings between the three groups. Thus, in LMP2-F12, the core–core, core–valence, and valence–valence pairs are uncoupled. This method works well if outer-core and valence orbitals are energetically well separated.

LOC_COREORB=MIX: first, outer-core and valence orbitals are localized together (using IBO localization). The intention of this step is to demix the outer-core and valence localized molecular orbitals (LMOs) as much as possible. Second, the LMOs are sorted such that the outer-core orbitals come first. Finally, the outer-core orbitals are re-localized among themselves using the IBO(AO) approach. This is necessary in order to demix (nearly) degenerate core orbitals and to obtain a unique set of LMOs.

The LOC_COREORB=MIX procedure is recommended for cases in which outer-core and valence orbitals mix significantly. This happened for some of the test systems (*e.g.* for $\overline{\text{In–Te–Tl–Te–In–Se}}$, *vide infra*). For consistency, it has been applied for all studied systems in this work. Using this approach, the partial charges of individual intrinsic atomic orbitals (IAOs) in the outer-core orbitals are very close to 2.0, *i.e.*, they are almost perfectly separated from the valence orbitals.

For calculations including also $(n-1)$s,p correlation, we used exactly the same orbitals for the respective calculations with $(n-1)$d outer-core orbitals, so that any effects originating from different orbital mixing can be excluded.

### 3.2 State-of-the-art explicitly correlated local MP2 and CCSD(T)

We applied the explicitly correlated pair natural orbital local second-order perturbation theory (PNO-LMP2-F12) and coupled cluster (PNO-LCCSD(T)-F12) methods as described in detail in previous work.[39–45] Reviews of this method can be found in ref. 59 and 60. We employed tight domain options [DOMOPT=TIGHT] as summarized in ref. 45, 51 and 60. The PNO occupation number threshold for pairs involving outer-core orbitals is reduced by a factor of 100 (this happens by default, but was found to have a very small effect in the current calculations). Default values were used for all other options unless otherwise noted. The triples (T) amplitude equations were solved iteratively in a basis of triples natural orbitals (TNOs) as described in ref. 43.

The explicitly correlated PNO-LMP2-F12 method uses F12 approximation 3*A[61,62] along with the fixed amplitude approximation.[63,64] This approach is simplest and most efficient but known to slightly overestimate the F12 correction compared to the formally more accurate 3C approximation,[65] especially for small basis sets. However, experience has shown that approximation 3*A combined with local approximations yields very accurate results and converges to the complete basis set (CBS) limit at least as fast as approximation 3C.[40,42,45] This is further corroborated by the tests carried out in the present work (*cf.* Section 3.4.4). For quadruple-$\zeta$ basis sets, the differences between results obtained with the two approximations usually become negligible.[45,51,60] In the LCCSD-F12 part, the F12b approximation[66,67] is used (shortly denoted F12 in the following), which was recently successfully applied to generate highly accurate reference isomerization and conformational energies.[68]

F12 methods are particularly well-suited for treating core and core–valence correlation effects in heavy main group elements,[69] which are very slowly convergent with basis set size. However, since the core orbitals are spatially more compact than the valence orbitals, the optimum exponents $\gamma$ of the F12 geminals $F_{12}(r_{12}) = -1/\gamma\exp(-\gamma r_{12})$ are larger than for valence orbitals. It has therefore been proposed to use pair-specific geminals with different exponents for valence–valence (vv), core–valence (cv), and core–core (cc) orbital pairs.[70] This option has also been implemented into the PNO-LCCSD(T)-F12 program and is used here for the first time (*cf.* Section 3.4.1).

In the course of this work, we observed significant outliers of the LMP2-F12 energy contributions for some systems and basis sets. It turned out that these were due to the last term in the (F12 strong orthogonality projector)[71–73]

$$\hat{Q}_{12}^{ij} = 1 - \sum_{m\in[ij]_{\mathrm{LMO}}}\sum_{\alpha\in[ij]_{\mathrm{RI}}}(|m\alpha\rangle\langle m\alpha| + |\alpha m\rangle\langle\alpha m|) + \sum_{m,n\in[ij]_{\mathrm{LMO}}}|mn\rangle\langle mn| - \sum_{a,b\in[ij]_{\mathrm{PNO}}}|ab\rangle\langle ab|. \quad (1)$$

The last term in eqn (1) projects out conventional double excitations into the domain $[ij]_{\mathrm{PNO}}$. The F12 contributions can approximately account for excitations into the remaining virtual MOs, which leads to a reduction of the domain error.[71,72,74] However, as has been noticed earlier,[75] in some cases PNOs with very small occupation numbers outside the domains $[ij]_{\mathrm{PNO}}$ give significant contributions to the matrix elements arising from this term, leading to an overestimation of the domain correction. This problem most likely occurs when strongly local LMOs are present, *e.g.*, lone pairs or core orbitals. It can be avoided by replacing the sum over the PNO pair domain $[ij]_{\mathrm{PNO}}$ by a sum over the PAO pair domain $[ij]_{\mathrm{PAO}}$ (which corresponds to including all PNOs spanned by the respective PAO domain). This means that the F12 terms cannot correct for the domain error that arises from excluding orbitals outside the PNO domain, but inside the PAO domain. In order to approximately compensate for this, a semi-canonical PAO–PNO domain correction is applied. In the current calculations, this correction is mostly slightly too small. The remaining domain error can be minimized by using a very tight PNO threshold in the LMP2-F12 calculation (*e.g.* option THRPNO_LMP2=1d-10).

In the following, we will denote the methods with the PAO and PNO projectors as PNO-LMP2-F12(PAO) and PNO-LMP2-F12(PNO), respectively. The choice of the projector has hardly any effect on the LCCSD-F12 contribution beyond the additive LMP2-F12 one. Using the PAO projector in the LCCSD-F12 part is also possible in Molpro but leads to exceedingly large CPU and memory requirements. Using the PAO and PNO projectors in the PNO-LMP2-F12 and LCCSD-F12 parts, respectively, is possible with the program option PROJECTOR=MIXED.

### 3.3 Computational settings

In the PNO-LCCSD(T)-F12 calculations we employed the cc-pVnZ-PP-F12 (n = T or Q) basis sets of Hill and Peterson[76] (cc-pVnZ-F12[77] for atoms with $Z \leq 18$) along with the associated OPTRI and MP2FIT auxiliary basis sets and the small-core effective core potentials (ECP10MDF, ECP28MDF, ECP60MDF)[78–80] for elements with $Z \geq 31$ (Ga) (simply denoted as vnz-f12 in the following). These basis sets include functions that are necessary for treating (outer) core–valence correlation effects, as well as diffuse functions as needed for non-covalent (long-range) interactions. For comparison, we also carried out a number of calculations using the larger aug-cc-pwCVnZ-PP[81] (n = T, Q) basis sets (aug-cc-pVnZ for atoms with $2 \leq Z \leq 18$ and cc-pVnZ for H). In the following, these sets are denoted shortly awcvnz. In the test calculations for which also the $(n-1)$s,p orbitals were correlated, the aug-cc-pwCVnZ-PP (aug-cc-pwCVnZ for $2 \leq Z \leq 18$) basis sets were exclusively used, since the vnZ-f12 basis sets do not contain tight polarization functions, which are essential in this case.

Relativistic spin–orbit effects were estimated using the zeroth-order regular approximation (ZORA)[82] and the exact two-component (X2C)[83–90] approach. The spin–orbit (SO) contribution at the X2C-PBE0-D4/x2c-QZVPPall-2c[91] level of theory for the dimerization energy of the heaviest system in the *IHD302* set ($\overline{\text{Tl}-\text{Po}-\text{Tl}-\text{Po}-\text{Tl}-\text{Po}}$) is rather large (for a closed-shell molecule), 18.4 kcal mol$^{-1}$ for the covalent and 5.9 kcal mol$^{-1}$ for the WDA dimer, respectively (see ESI,† Table S7). Similarly large

SOC contributions were calculated for heavy p-block dimers by Höfener *et al.*[92] Nonetheless, X2C results without SO contribution agree well with the respective dimerization energies obtained with effective core potentials (ECPs). Hence, for all other calculations, we employed ECPs to account approximately for scalar relativistic effects, which become important for heavier elements. The same ECPs (*vide supra*) have also been used in the DFT and SQM calculations so that the comparison of wave function and approximate methods is on equal footing unless ECPs are missing for certain elements, as for Ga, Ge, As, and Se in the def2-QZVPP basis set (*vide infa*).

Density fitting for the Fock matrix employed the def2-QZVPP/JKFIT[93] or aug-cc-pVnZ/JKFIT for light atoms, n = T or Q; the latter sets, which are available in the Molpro basis set library, have been derived from the cc-pVnZ/JKFIT basis sets of Weigend[94] by adding for each angular momentum a shell of diffuse functions in an even-tempered manner. For the RI-approximations in the PNO-LMP2-F12 and PNO-LCCSD(T)-F12 calculations, the so-called CABS basis sets were employed, which comprise the union of the vnz-f12 orbital basis and the associated VnZ-PP-F12/OPTRI basis.[95] For the CABS-singles corrections the complementing auxiliary (CA) orbital space was explicitly constructed by orthogonalizing the CABS basis set on the corresponding orbital basis.

The same CABS and DF basis sets were also employed in the PNO-LCCSD(T)-F12/awcvnz calculations, but in this case, the unions of the orbital and JKFIT basis sets were used to construct the CA space for the CABS-singles corrections (since there are no awcvnz/OPTRI basis sets available yet). In order to test the sensitivity of the results to the choice of the auxiliary basis sets, we also employed the quadruple-$\zeta$ DF and CABS basis sets (as well as few other choices for the RI basis, *cf.* Table S1 in the ESI†) in some PNO-LCCSD(T)-F12/awcvtz calculations. It should be noted that using the CABS basis, which corresponds to the respective orbital basis, is exactly equivalent to using the CABS approach[96] in the F12 treatment. However, when the current orbital basis differs from the one used in the CABS basis this is no longer exactly the case. However, as long as all occupied orbitals can be well represented by the CABS basis this causes only minor errors. This is also corroborated by the tests carried out in the present work (*cf.* ESI,† Table S1).

The CABS singles correction is included in all results of this paper (also for LMP2 without F12). It was applied to all occupied orbitals (option CORE_SINGLES=1). Note that by default Molpro excludes uncorrelated core orbitals from the CABS correction, which can lead to significant errors for the systems under consideration (up to 0.4 kcal mol$^{-1}$ for $\overline{\text{Pb}-\text{Sb}-\text{Si}-\text{Bi}-\text{Ge}-\text{N}}_{\text{COV}}$ ($7_{\text{COV}}$), see ESI† (data correlation methods)), unless either the option CORE_SINGLES=1 or an appropriate CORE directive is given in the CABS singles calculation.

Finally, we note that for an accurate description of the core–core correlation effects it was necessary to tighten the integral screening parameters in the PNO program (Molpro option BB_THRESH=10$^{-8}$). Likely, this is due to near linear dependencies of the RI basis. When using tight screening thresholds, the different choices of the RI basis had only a negligible effect on the results and all results were numerically stable. However, with the default screening threshold (BB_THRESH=10$^{-6}$), in some cases very large errors occurred with the CABS basis sets. These errors can be avoided by removing near linear dependencies of the RI basis using singular value decomposition (SVD). This yields results very close to those obtained with the tight screening threshold. However, SVD is rather expensive, since it requires diagonalizing the RI overlap matrix for each pair (because the RI domains are pair-dependent). SVD has therefore not been used in the final calculations.

### 3.4 Finding a suitable reference protocol

In order to check the dependence of the PNO-LCCSD(T)-F12 results on basis sets, thresholds, and further settings as well as to estimate the accuracy of the reference values, we carried out extensive test calculations for eight systems covering the majority of elements and all main group combinations (*i.e.*, III–VI, III–V, IV–V in this order) included in the whole *IHD302* benchmark set. The structures of the covalently bound dimers of these systems are shown in Fig. 3. Unless otherwise noted, the ($n$ − 1)s,p orbitals were not correlated in these calculations.

Initially, we tested whether the counter-poise (CP) correction should be applied to the dimerization energies to reduce basis set superposition effects (BSSE). Table S2 in the ESI,† shows that the CP corrections are significant for PNO–LMP2 (up to 7% of the respective dimerization energy), but negligibly small for PNO–LMP2–F12. We, therefore, decided not to apply CP corrections in further correlation calculations conducted in this work, which were mostly carried out with F12.

Fig. 4 summarizes the computed dimerization energies for different methods tested using the awcvqz basis (PROJECTOR=-MIXED, THRPNO_LMP2=1d-10, *cf.* Section 3.2). The total electron correlation contributions to the respective binding energies are rather large. At the HF level, the weak donor–acceptor complexes are mostly unbound. For the covalently bound systems, most systems are bound at the HF level and the total correlation effect on the dimerization energies is huge (more than 70 kcal mol$^{-1}$ for $\overline{\text{B}-\text{Se}-\text{B}-\text{Se}-\text{B}-\text{Se}}_{\text{COV}}$ ($\mathbf{1}_{\mathbf{COV}}$)). This makes it extremely difficult to obtain results with sub-kcal mol$^{-1}$ accuracy.

All dimerization energies are clearly overestimated at the LMP2-F12 level. In contrast, the LCCSD-F12 values are mostly too small, and the triples contributions to the binding energies are rather large and negative (−10.9 kcal mol$^{-1}$ for $\mathbf{1}_{\mathbf{COV}}$ with awcvqz). Only for the test systems **7** and **8**, which belong to the IV–V class, they are positive. This may be related to the different electronic structures of these systems, where formal double bonds are converted to single bonds (*cf.* Fig. 1).

In the following sections, we investigate how explicit correlation, local approximations, core correlation, and the choice of the basis set affect the accuracy of the dimerization energies. Our goal is to obtain reference values with a relative accuracy of 1–2% for the covalently bound systems and $<5\%$ for the weak-donor acceptor structures, which means an absolute accuracy of 1–2 kcal mol$^{-1}$ and 0.5 kcal mol$^{-1}$, respectively. In exceptional cases, where the absolute values of the dimerization

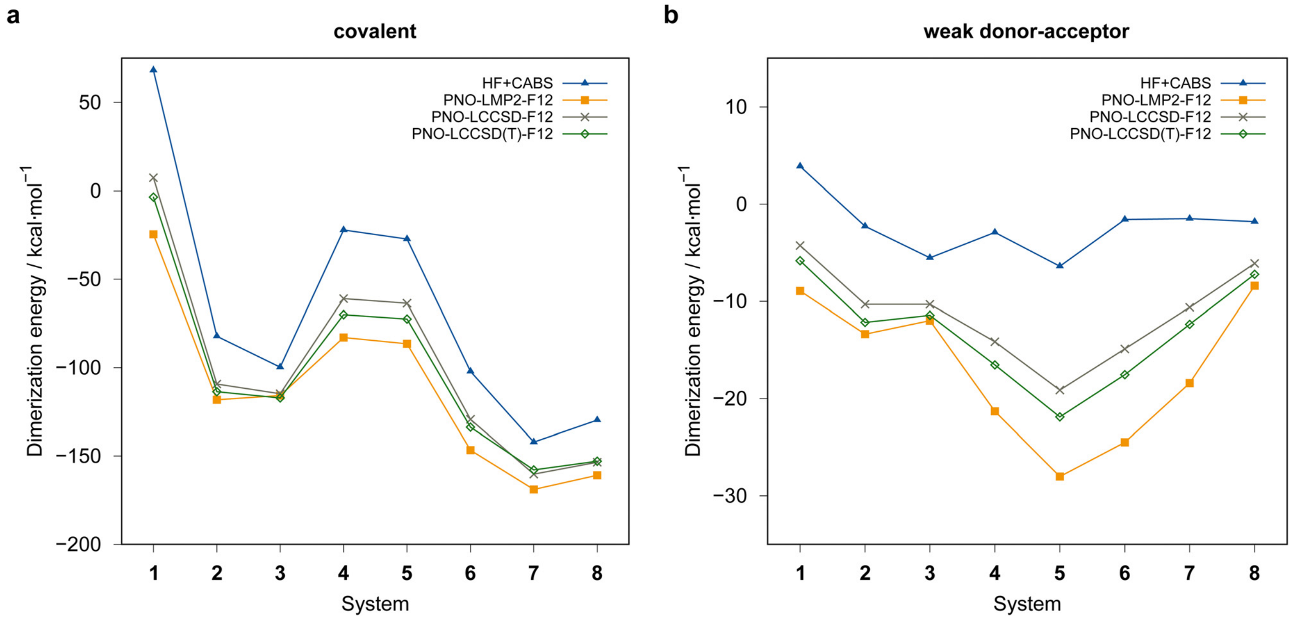


Fig. 4 Comparison of computed dimerization energies using different methods and the awcvqz basis (see text) for the (a) covalent and (b) weak donor–acceptor bound systems depicted in Fig. 3.

energies are very small relative to the total correlation contribution (*e.g.* for $\overline{\mathrm{B{-}Se{-}B{-}Se{-}B{-}Se}}_{\mathrm{COV}}$ ($\mathbf{1}_{\mathbf{COV}}$)), slightly larger errors of our reference values protocol (*vide infa*) may occur.

Unfortunately, checking the intrinsic accuracy of the CCSD(T) method for these applications by including higher connected excitations in the coupled cluster expansion is computationally impossible due to the tremendous cost. Since neither the FOD plots[34] for the eight test systems (see Fig. S9 in the ESI†) nor the $r_{\mathrm{ND}}$ static correlation diagnostic[37] for the complete test set show signs of significant static correlation (see ESI,† for details), we are convinced that CCSD(T) is able to correctly describe the electronic structure of the systems in *IHD302*. Hence, based on previous experience and given that the electronic structures of systems in question are of single reference type, it can be assumed that it is in the same range as the above error bounds.

**3.4.1 Choice of F12 geminal exponents.** For calculations with pair-specific geminals, the F12 exponents $\gamma$ for cc, cv, and vv correlation were optimized by minimizing PNO-LMP2-F12 energies (in steps of $0.1{a_0}^{-1}$) using the vtz-f12 and vqz-f12 basis sets. Optimization of the F12 exponents by minimizing the total energies leads for all three structures (monomer, covalent, and weak-donor acceptor) to the same values (within the accuracy of the optimization). The optimized exponents (abbreviated as optv and opt3 for vv and vv+cv+cc correlation, respectively) are listed in Table S3 of the ESI.† These values were then applied in the respective PNO-LCCSD(T)-F12 calculations, since, in contrast to PNO-LMP2-F12 (upper bound for the exact MP2-F12 energy by using the Hylleraas functional), there is no variational energy functional.

The exponents for cc and cv correlation decrease with increasing row number in the periodic table and increase from left to right in the periodic table. This is reflected in the optimized values, which represent compromises of the optimized values one would obtain for the individual atoms contained in the individual systems. Another possibility would be to optimize exponents for each of the 19 elements contained in the *IHD302* set separately and then apply weighted averages of these values for the molecules. However, in view of the rather low sensitivity of the results to the choice of exponents, we did not attempt this. PNO-LCCSD(T)-F12 results for different choices of the geminal exponents and employing the vtz-f12 basis are summarized in Table S4 of the ESI.† Results for a subset of these test calculations are presented in Table 1. Given the rather small effect (compared to, *e.g.*, the inclusion of cv and cc correlation, *vide infra*) of varying the geminal exponents and to avoid the geminal exponent optimization for all systems comprised in the *IHD302* set, we decided to use just two fixed exponents for all systems, $\gamma_{\mathrm{vv}} = 1.0{a_0}^{-1}$ for valence correlation, and $\gamma_{\mathrm{cc}} = \gamma_{\mathrm{cv}} = 1.4{a_0}^{-1}$ for cc+cv correlation. As shown in Table 1, this has only a small effect on the PNO-LCCSD(T)-F12 relative energies. In most cases, the dimerization energies are slightly larger than those obtained with the three optimized exponents thus bringing them closer to our best awcvqz results, especially for the covalent structures (see ESI†). For the 13 systems of the *IHD302* set containing exclusively elements of the first three rows of the periodic table, only the valence electrons were correlated using the default value $\gamma = 1.0{a_0}^{-1}$.

**3.4.2 Convergence with thresholds for local approximations.** Table 2 shows a comparison of results obtained with canonical and local MP2-F12 calculations for three out of the eight test cases with especially large correlation contributions. In the canonical F12 calculations, approximation 3*A is compared with the formally more accurate approximation 3C. For $\overline{\mathrm{In{-}Te{-}Tl{-}Te{-}In{-}Se}}$ ($\mathbf{5}_{\mathbf{COV}}$ and $\mathbf{5}_{\mathbf{WDA}}$) calculated with the

**Table 1** HF (including CABS singles correction) and PNO-LCCSD(T)-F12 dimerization energies in kcal mol$^{-1}$ (basis: vtz-f12, CABS RI basis, THRPNO_LMP2=1d-8, projector=PNO) (see text)

| | | vv | | cc+cv+vv | | |
|---|---|---|---|---|---|---|
| System | HF+CABS | $\gamma$ (1.0) | $\gamma$ (optv) | $\gamma$ (1.0) | $\gamma$ = (1.0, 1.4) | $\gamma$ (opt3) |
| Covalent: | | | | | | |
| $\overline{\text{B–Se–B–Se–B–Se}}$ ($\mathbf{1}_{\mathbf{COV}}$) | 68.3 | −2.9 | −2.9 | −2.3 | −2.4 | −2.3 |
| $\overline{\text{Al–O–Al–S–Al–Se}}$ ($\mathbf{2}_{\mathbf{COV}}$) | −82.2 | −112.2 | −112.9 | −113.3 | −113.2 | −112.9 |
| $\overline{\text{Ga–O–In–O–Tl–O}}$ ($\mathbf{3}_{\mathbf{COV}}$) | −99.8 | −111.7 | −111.7 | −115.0 | −115.5 | −115.4 |
| $\overline{\text{Ga–Te–In–Te–Ga–Se}}$ ($\mathbf{4}_{\mathbf{COV}}$) | −22.1 | −66.6 | −66.8 | −68.3 | −68.9 | −68.9 |
| $\overline{\text{In–Te–Tl–Te–In–Se}}$ ($\mathbf{5}_{\mathbf{COV}}$) | −27.3 | −66.0 | −66.2 | −70.0 | −70.7 | −70.0 |
| $\overline{\text{Tl–P–Tl–P–Tl–Bi}}$ ($\mathbf{6}_{\mathbf{COV}}$) | −102.2 | −124.5 | −124.9 | −132.3 | −132.8 | −131.8 |
| $\overline{\text{Pb–Sb–Si–Bi–Ge–N}}$ ($\mathbf{7}_{\mathbf{COV}}$) | −142.2 | −154.8 | −155.1 | −155.0 | −155.3 | −154.8 |
| $\overline{\text{Si–N–Si–N–Si–P}}$ ($\mathbf{8}_{\mathbf{COV}}$) | −129.8 | −151.8 | −152.3 | | | |
| Weak donor–acceptor: | | | | | | |
| $\overline{\text{B–Se–B–Se–B–Se}}$ ($\mathbf{1}_{\mathbf{WDA}}$) | 3.9 | −5.6 | −5.7 | −5.4 | −5.5 | −5.6 |
| $\overline{\text{Al–O–Al–S–Al–Se}}$ ($\mathbf{2}_{\mathbf{WDA}}$) | −2.3 | −12.0 | −12.0 | −11.9 | −12.0 | −12.0 |
| $\overline{\text{Ga–O–In–O–Tl–O}}$ ($\mathbf{3}_{\mathbf{WDA}}$) | −5.6 | −9.8 | −9.8 | −11.2 | −11.2 | −11.2 |
| $\overline{\text{Ga–Te–In–Te–Ga–Se}}$ ($\mathbf{4}_{\mathbf{WDA}}$) | −2.9 | −15.5 | −15.5 | −16.0 | −16.1 | −16.2 |
| $\overline{\text{In–Te–Tl–Te–In–Se}}$ ($\mathbf{5}_{\mathbf{WDA}}$) | −6.4 | −19.8 | −19.8 | −21.0 | −21.3 | −21.1 |
| $\overline{\text{Tl–P–Tl–P–Tl–Bi}}$ ($\mathbf{6}_{\mathbf{WDA}}$) | −1.6 | −14.9 | −14.9 | −16.9 | −17.1 | −16.8 |
| $\overline{\text{Pb–Sb–Si–Bi–Ge–N}}$ ($\mathbf{7}_{\mathbf{WDA}}$) | −1.5 | −11.4 | −11.3 | −11.7 | −11.8 | −11.8 |
| $\overline{\text{Si–N–Si–N–Si–P}}$ ($\mathbf{8}_{\mathbf{WDA}}$) | −1.9 | −7.1 | −7.1 | | | |

**Table 2** Comparison of canonical and local MP2-F12 dimerization energies in kcal mol$^{-1}$. All F12 calculations with carried out with $\gamma$ = [1.0,1.4]

| | | $\overline{\text{B–Se–B–Se–B–Se}}$ | | $\overline{\text{In–Te–Tl–Te–In–Se}}$ | | $\overline{\text{Ga–Te–In–Te–Ga–Se}}$ | |
|---|---|---|---|---|---|---|---|
| Basis | Method | WDA | Covalent | WDA | Covalent | WDA | Covalent |
| vtz-f12 | HF | 3.8 | 68.3 | −6.5 | −27.8 | −3.0 | −22.3 |
| | HF+CABS | 3.9 | 68.3 | −6.4 | −27.3 | −2.9 | −22.1 |
| | MP2 | −9.4 | −25.0 | −33.0 | −92.9 | −24.8 | −88.8 |
| | MP2-F12/3C | −9.1 | −25.2 | −29.1 | −87.9 | −22.2 | −84.7 |
| | MP2-F12/3*A | −8.8 | −24.0 | −27.7 | −85.4 | −21.5 | −83.0 |
| | LMP2-F12/3*A[a] | −8.8 | −24.1 | −27.7 | −85.3 | −21.2 | −82.6 |
| | LMP2-F12/3*A[b] | −8.8 | −23.7 | −27.6 | −85.1 | −21.1 | −82.4 |
| awcvtz | HF | 3.8 | 68.2 | −6.5 | −27.4 | −3.0 | −22.2 |
| | HF+CABS | 3.9 | 68.3 | −6.4 | −27.3 | −2.9 | −22.1 |
| | MP2 | −9.8 | −24.1 | −29.0 | −85.3 | −22.2 | −81.5 |
| | MP2-F12/3C | −9.1 | −24.9 | −29.4 | −86.2 | −21.5 | −83.0 |
| | MP2-F12/3*A | −9.0 | −24.7 | −28.1 | −85.6 | −21.4 | −82.5 |
| | LMP2-F12/3*A[a] | −8.9 | −24.8 | −27.9 | −86.0 | −21.3 | −82.5 |
| | LMP2-F12/3*A[b] | −8.9 | −24.4 | −27.9 | −85.8 | −21.2 | −82.3 |
| awcvqz | LMP2-F12/3*A[a] | −8.9 | −24.6 | −28.0 | −86.5 | −21.3 | −82.9 |

[a] Projector=PAO,THRPNO_LMP2=1d-10. [b] Projector=PAO,THRPNO_LMP2=1d-8.

vtz-f12 basis, the observed behavior is as expected: approximation 3C underestimates the (positive) F12 contribution, yielding for both structures (covalent and weak donor–acceptor) results that are too negative as compared to the best awcvqz values. On the other hand, approximation 3*A overestimates the F12 contribution, thus somewhat underestimating the dimerization energy (too "positive" $\Delta E$). For the covalently bound dimers, the differences relative to the awcvqz results amount to −1.4 and +1.0 kcal mol$^{-1}$ for 3C and 3*A, respectively, while for the weak donor–acceptor ones, the corresponding differences are −1.0 and +0.4 kcal mol$^{-1}$. A rather similar behavior is observed for $\overline{\text{B–Se–B–Se–B–Se}}_{\text{COV}}$ ($\mathbf{1}_{\mathbf{COV}}$), although with the awcvtz basis, 3C seems to slightly overestimate the F12 corrections. Currently, we do not have an explanation for these subtle effects, but the results demonstrate that there is no disadvantage of using the simple and more efficient approximation 3*A.

The canonical and local results are in close agreement. The small differences seen with $T_{\text{PNO}}^{\text{LMP2}}$ = $10^{-10}$ (program option THRPNO_LMP2) may originate from the PAO domain approximation and the RI and LMO domain approximations in the strong orthogonality projector (eqn (1)). If the default threshold THRPNO_LMP2=1d-8 is used, an additional error of up to 0.4 kcal mol$^{-1}$ in case of $\overline{\text{B–Se–B–Se–B–Se}}_{\text{COV}}$ ($\mathbf{1}_{\mathbf{COV}}$) can arise. Given a total correlation contribution of over 90 kcal mol$^{-1}$ (LMP2-F12, COV) this is certainly acceptable.

For some test systems, we investigated the dependence of the results on the PAO and PNO thresholds (see ESI,† for details). The PAO domain sizes were varied by increasing the parameter REXT, which had a negligible effect. The PNO domains used in the PNO-LCCSD-F12 were enlarged by increasing the parameter THREN_CC from 0.90 to 0.997; this means that for each pair at least 99.7% of the semi-canonical PAO-LMP2 correlation energy is recovered with the PNO domains. Again, this affected the dimerization energies by less than 0.1 kcal mol$^{-1}$. Also, the differences in the dimerization energies obtained with DEFAULT and TIGHT domain options, which affect PAO, PNO, and TNO domain sizes, are rather small. The same holds for the pair approximations. Therefore, the values generated with TIGHT domain options should be well converged with respect to the local approximations. This is also supported by comparisons of canonical and local MP2-F12 results (*cf.* Section 3.4.4).

However, as already mentioned in Section 3.2, some unexpected outliers were observed for the calculations with the awcvtz basis. It turned out that these were due to the choice of the local approximations in the F12 strong orthogonality projector (*cf.* Section 3.2). Fig. 5 shows the convergence of the PNO-LMP2-F12 dimerization energies as a function of the PNO threshold $T_{\text{PNO}}^{\text{LMP2}}$ for $\overline{\text{B–Se–B–Se–B–Se}}_{\text{COV}}$ (**1$_{\text{COV}}$**) and $\overline{\text{Ga–Te–In–Te–Ga–Se}}_{\text{COV}}$ (**4$_{\text{COV}}$**), where the effect of the projector is particularly strong. With the PAO projector, the convergence is smooth, but the domain correction underestimates the PNO domain error by about 0.2 kcal mol$^{-1}$, if the default value $T_{\text{PNO}}^{\text{LMP2}} = 10^{-8}$ is used. This error can be reduced to a completely negligible amount by using a tighter PNO threshold, *e.g.* $T_{\text{PNO}}^{\text{LMP2}} = 10^{-10}$. In the current work, such problems were only observed for calculations with the awcvnz basis sets but not for vtz-f12. Hence, we applied $T_{\text{PNO}}^{\text{LMP2}} = 10^{-10}$ for all further local calculations carried out with the awcvnz basis sets.

For all systems comprised in the *IHD302* test set, the difference of the LMP2-F12(PAO)/vtz-f12 and respective LMP2-F12(PNO)/vtz-f12 dimerization energies of the covalent structures amounts on average to 0.21 kcal mol$^{-1}$ (almost always $>0$) with a maximum deviation of only 0.33 kcal mol$^{-1}$ (see ESI,† for details). For the weak donor–acceptor systems, the difference of the LMP2-F12 dimerization energies computed with PAO or PNO projector, respectively, is negligibly small (0.01 kcal mol$^{-1}$ on average). Thus, with the vtz-f12 basis, the dependence of the results on the choice of the F12 projector is much less pronounced than for awcvnz basis sets.

**3.4.3 Core correlation effects.** In Table 1, PNO-LCCSD(T)-F12 results with and without outer-core correlation are compared. The magnitudes of the $(n-1)$d cc+cv effects are strongly system dependent and, as expected, largest for the main group III elements (increasing with the nuclear charge $Z$). In general, cc+cv correlation leads to more negative dimerization energies. For $\overline{\text{Tl–P–Tl–P–Tl–Bi}}_{\text{WDA}}$ (**6$_{\text{WDA}}$**), the cc+cv correlation effect amounts to $\approx$2 kcal mol$^{-1}$, and for the respective covalent structure structure (**6$_{\text{COV}}$**) to about 7–8 kcal mol$^{-1}$, depending on the choice of the geminal exponents. The corresponding values for $\overline{\text{In–Te–Tl–Te–In–Se}}$ (**5$_{\text{WDA}}$** and **5$_{\text{COV}}$**) amount to $\approx$1 kcal mol$^{-1}$ and about 4 kcal mol$^{-1}$, respectively. These large effects demonstrate that including cc+cv correlation effects in these calculations is essential for obtaining accurate results.

In some calculations with the awcvtz basis, we tested also the effect of including the $(n-1)$s and p orbitals in the correlation treatment. The effect on the weak donor–acceptor dimerization energies is rather small (see ESI,† for details). However, in some cases, the PNO-LMP2-F12 dimerization energies of the covalent structures containing several main group III elements were significantly reduced (up to 1.75 kcal mol$^{-1}$ for $\overline{\text{In–Te–Tl–Te–In–Se}}_{\text{COV}}$ (**5$_{\text{COV}}$**), *i.e.*, here the correlation effect is positive). Note, however, that the cc+cv correlation is generally rather strongly over-estimated at the PNO-LMP2-F12 level. Core correlation effects are much smaller at the PNO-LCCSD(T)-F12 level (*e.g.*, 1.15 kcal mol$^{-1}$ for $\overline{\text{In–Te–Tl–Te–In–Se}}_{\text{COV}}$ (**5$_{\text{COV}}$**)).

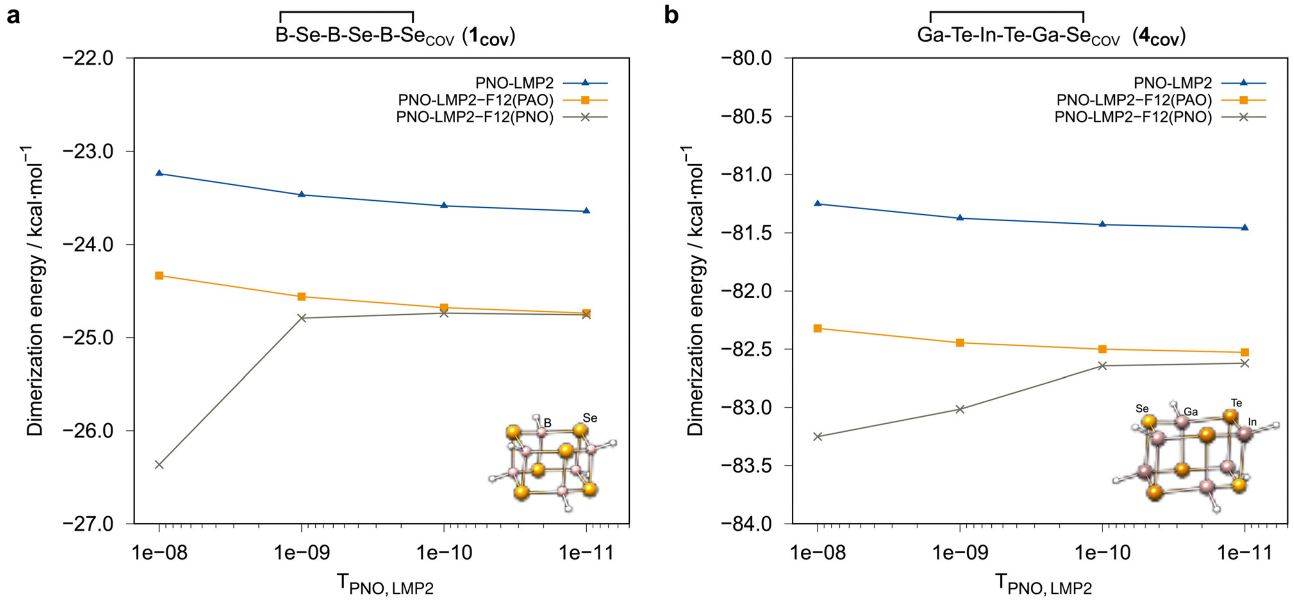


Fig. 5 Convergence of PNO-LMP2-F12/awcvtz dimerization energies of (a) $\overline{\text{B–Se–B–Se–B–Se}}$ (**1$_{\text{COV}}$**) and (b) $\overline{\text{Ga–Te–In–Te–Ga–Se}}$ (**4$_{\text{COV}}$**) with the respect to the threshold $T_{\text{PNO}}^{\text{LMP2}}$ employing the PAO and PNO projectors, respectively (see text).

In all other tested cases, the effect of correlating the ($n$ − 1)s,p orbitals was well below 1 kcal mol$^{-1}$. Moreover, due to the employed pseudopotential approximation, estimating the accuracy of the ($n$ − 1)s,p correlation effect is somewhat problematic. For the post-d elements, these orbitals are the lowest ones that are explicitly treated and have no radial nodes, which may lead to a significant overestimation of the respective correlation contributions. A more detailed study of these effects would require very expensive relativistic all-electron calculations, which is beyond the scope of this paper. We therefore decided not to include these orbitals in the final PNO-LCCSD(T)-F12 calculations for all molecules. Additionally, since the effects of correlating the ($n$ − 1)d and ($n$ − 1)s,p shells mostly have opposite signs, this may lead to some beneficial error compensations of remaining basis set deficiencies.

**3.4.4 Basis set dependence.** Fig. 6 shows the basis set errors of vtz-f12, vqz-f12, and awcvtz PNO-LCCSD(T)-F12 dimerization energies with respect to the PNO-LCCSD(T)-F12/awcvqz results, whereby the latter should be converged within a few tenths of a kcal mol$^{-1}$ with respect to the true complete basis set limits. In order to be able to carry out an unbiased analysis of the basis set errors, all calculations for this purpose were performed with PROJECTOR=MIXED and THRPNO_LMP2=1d-10.

Generally, the vtz-f12 dimerization energies of the eight test systems (*cf.* Fig. 3) are significantly underestimated (*i.e.*, the basis set errors are >0), up to about 0.5 kcal mol$^{-1}$ for the weak-doner–acceptor structures and 2 kcal mol$^{-1}$ for the covalently bound dimers, respectively. In view of the large and negative correlation contributions this is not unexpected, although the errors are much larger than those seen in previous PNO-LCCSD(T)-F12 benchmark calculations for reactions of organic molecules.[68] Using the vqz-f12 basis, the deviation from the awcvqz dimerization energies for the tested weak donor–acceptor structures are all below 0.16 kcal mol$^{-1}$. However, the errors for the covalently bound dimers are still rather large, up to 1.3 kcal mol$^{-1}$ for $\overline{\text{B–Se–B–Se–B–Se}}_{\text{COV}}$ ($\mathbf{1}_{\mathbf{COV}}$).

The rather large basis set errors obtained with the vtz-f12 and vqz-f12 basis sets are somewhat unexpected and unusual. They are related to the fact that in these sets only one set of additional d and f functions has been added to the underlying aug-cc-pV($n$+1)Z-PP valence basis sets to describe the cc+cv effects,[76] assuming that the F12 corrections would cover the rest. Without F12, this leads to strong intramolecular BSSE and much too large binding energies. The F12 contributions are therefore positive (up to almost 10 kcal mol$^{-1}$) (*cf.* Fig. 7). Surprisingly, for the covalent structures, they are even larger with vqz-f12 than with vtz-f12. One positive aspect, however, is that F12 can reduce the errors in these cases from up to 10 kcal mol$^{-1}$ to around 1 kcal mol$^{-1}$ or even less.

These findings are exemplified in Table 3 for $\overline{\text{In–Te–Tl–Te–In–Se}}$ ($\mathbf{5}_{\mathbf{COV}}$ and $\mathbf{5}_{\mathbf{WDA}}$), one of our test systems with the largest remaining errors (with respect to the corresponding awcvqz dimerization energies). First, it can be seen that due to the intramolecular BSSE, the PNO-LMP2/vnz-f12 dimerization energies are much over-estimated (too negative). This also holds for the triples (T) contributions, which are not directly affected by the F12 terms. Except for the triples, this problem is largely cured by the F12 treatment.

With the awcvtz and awcvqz basis sets, which have been specifically optimized to describe cc+cv correlation effects without F12, the F12 contributions to the binding energies are distinctively smaller than for the vnz-f12 basis, especially for the covalent structures. Moreover, they are mostly negative, indicating that the above-mentioned intramolecular BSSE

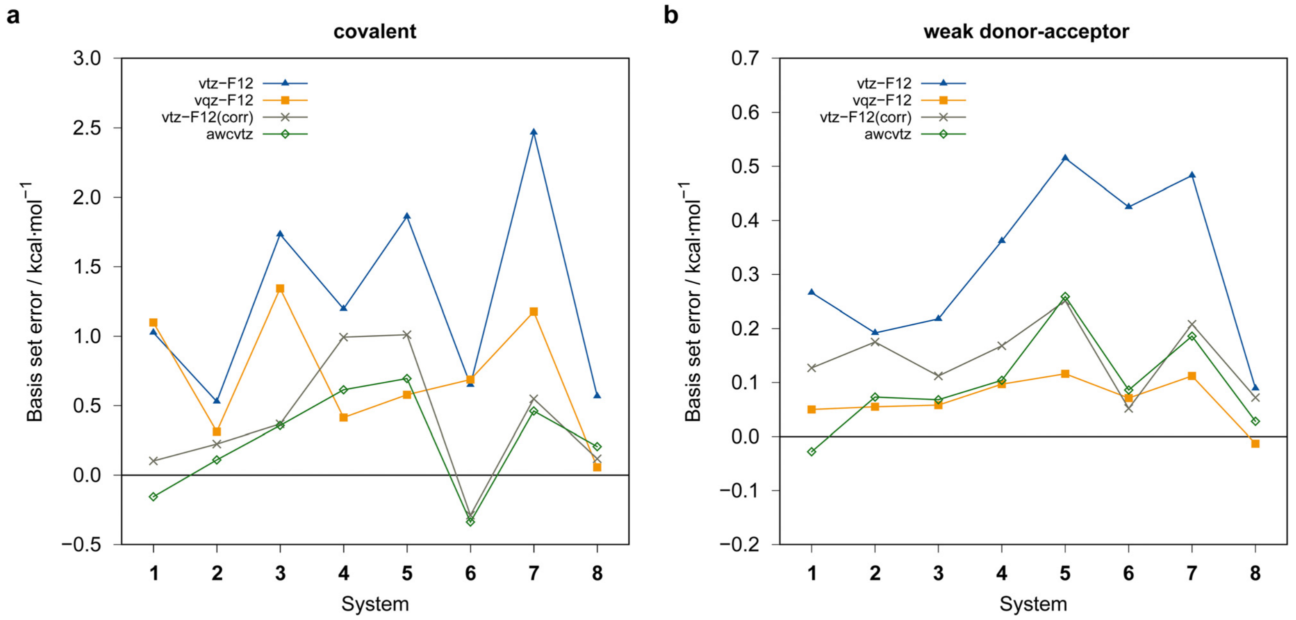


Fig. 6 PNO-LCCSD(T)-F12 basis set errors (in kcal mol$^{-1}$) relative to the PNO-LCCSD(T)-F12/awcvqz reference values (using PROJECTOR=MIXED and THRPNO_LMP2=1d-10 except for vtz-f12(corr.), see text). (a) Covalent dimers, (b) weak donor–acceptor dimers. vqz-f12 values computed with $\gamma$ (opt3), otherwise $\gamma$ = [1.0,1.4] was employed.

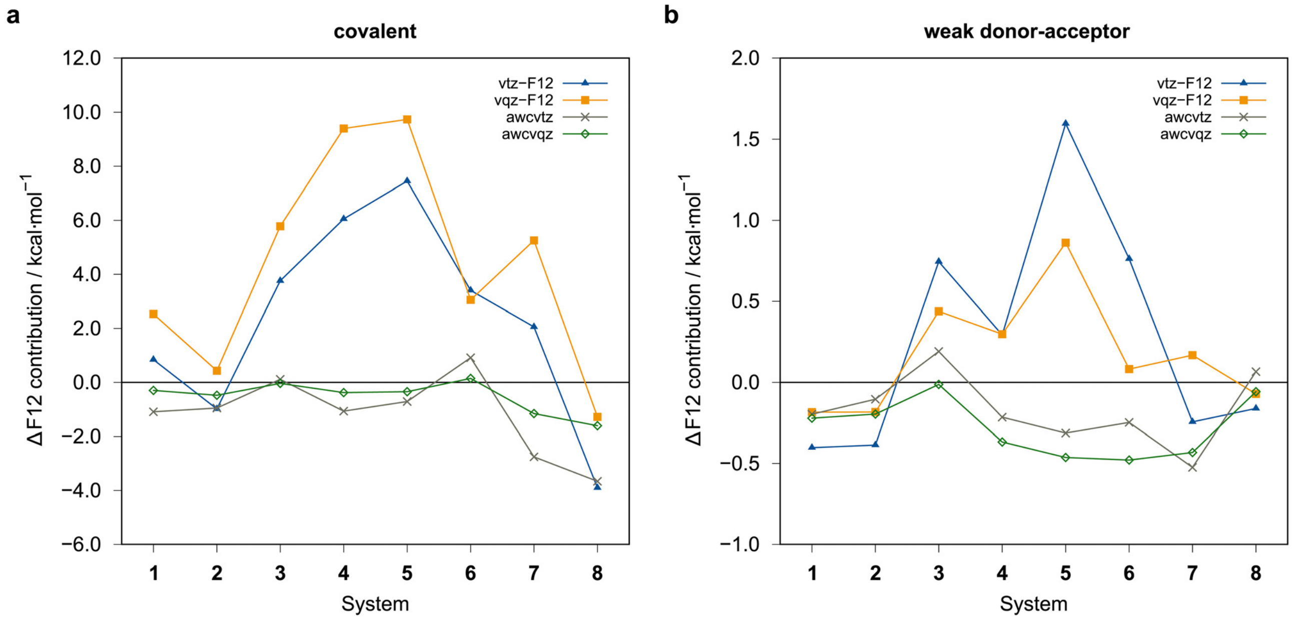


Fig. 7 $\Delta F_{12} = \Delta E_{\text{LMP2-F12(PAO)}} - \Delta E_{\text{LMP2}}$ contributions in to the binding energies (THRPNO_LMP2=1d-10) in kcal mol$^{-1}$. (a) Covalent dimers, (b) weak donor–acceptor dimers. vqz-f12 values computed with $\gamma$ (opt3), otherwise $\gamma$ = [1.0,1.4] was employed.

Table 3 Dimerization energies in kcal mol$^{-1}$ for $\overline{\text{In–Te–Tl–Te–In–Se}}$ (**5**$_{\text{WDA}}$ and **5**$_{\text{COV}}$) calculated with different PNO methods and basis sets

| Basis | Projector | $T_{\text{PNO}}^{\text{LMP2}}$ | $\gamma$ [vv,cv,cc] | LMP2 | LMP2-F12 | LCCSD-F12 | LCCSD(T)-F12 | $\Delta(T)$ |
|---|---|---|---|---|---|---|---|---|
| Weak donor–acceptor: | | | | | | | | |
| vtz-f12 | PNO | 1d-08 | [1.0,1.4,1.4] | −29.2 | −27.6 | −18.6 | −21.3 | −2.7 |
| vqz-f12 | PNO | 1d-08 | [0.8,1.3,1.4] | −28.7 | −27.9 | −18.9 | −21.7 | −2.9 |
| avcvtz | MIX | 1d-08 | [1.0,1.4,1.4] | −27.6 | −28.0 | −19.0 | −21.7 | −2.6 |
| awcvtz | MIX | 1d-10 | [1.0,1.4,1.4] | −27.6 | −27.9 | −19.1 | −21.7 | −2.6 |
| awcvqz | MIX | 1d-10 | [1.0,1.4,1.4] | −27.5 | −28.0 | −19.1 | −21.9 | −2.8 |
| Covalent: | | | | | | | | |
| vtz-f12 | PNO | 1d-08 | [1.0,1.4,1.4] | −93.0 | −85.7 | −62.2 | −71.1 | −8.9 |
| vqz-f12 | PNO | 1d-08 | [0.8,1.3,1.4] | −96.1 | −86.6 | −62.7 | −72.2 | −9.5 |
| awcvtz | MIX | 1d-08 | [1.0,1.4,1.4] | −85.3 | −86.0 | −63.3 | −71.8 | −8.5 |
| awcvtz | MIX | 1d-10 | [1.0,1.4,1.4] | −85.4 | −86.1 | −63.4 | −71.9 | −8.5 |
| awcvqz | MIX | 1d-10 | [1.0,1.4,1.4] | −86.0 | −86.5 | −63.6 | −72.5 | −9.0 |

effects (without F12) are strongly reduced. Employing the awcvtz basis, the errors for the covalent systems are all below 0.7 kcal mol$^{-1}$ relative to the awcvqz results but about 4–5 times faster (somewhat system-dependent).

Unfortunately, the awcvtz basis is significantly larger than the vtz-f12 one, and due to the many additional high-angular momentum basis functions, the calculations with the awcvtz basis are at least twice as expensive compared to the respective vtz-f12 ones. Almost the same improvement as with the awcvtz basis can be obtained by adding a PNO-LMP2-F12(PAO) basis set correction to the respective PNO-LCCSD(T)-F12(PNO)/vtz-f12 energies

$$\begin{aligned} E_{\text{LCCSD(T)-F12/vtz-f12(corr.)}} &= E_{\text{LCCSD(T)-F12(PNO)/vtz-f12}} \\ &\quad - E_{\text{LMP2-F12(PAO)/vtz-f12}} \\ &\quad + E_{\text{LMP2-F12(PAO)/awcvtz}}. \end{aligned}$$

This is in the spirit of composite methods as frequently used in CCSD(T)(-F12) calculations to approximate the basis set limit.[97,98] We note that in the second term, one could also use $E_{\text{LMP2-F12/(PNO)}}$ (rather than PAO projector). Since the F12(PAO) energy contributions are somewhat smaller than the F12(PNO) ones (*cf.* Fig. 5), the corrections are slightly more negative with the PAO projector thus yielding better agreement with the target awcvqz dimerization energies.

The amount of the basis set correction is mostly small (on average about −0.7 kcal mol$^{-1}$ and −0.2 kcal mol$^{-1}$ for the covalent and weak donor–acceptor complexes, respectively). Only for a few systems the correction is significantly larger (up to about −2.8 kcal mol$^{-1}$ and −0.8 kcal mol$^{-1}$, respectively; see ESI,† for details).

To further verify the effectiveness of the basis set correction, we also calculated the PNO-LCCSD(T)-F12/awcvnz (n = T, Q) dimerization energies for the eight systems with the largest correction values (see ESI,† for details). Even in these cases the agreement of the vtz-f12(corr) results with awcvqz ones is very satisfactory. The RMSD relative to the awcvqz results for all 16 test systems amounts to 0.17 kcal mol$^{-1}$ and 0.52 kcal mol$^{-1}$ (maximum errors: 0.29 and 1.01 kcal mol$^{-1}$) for the WDA and COV systems, respectively. It appears that a significant part of the remaining errors is due to the basis set error of the relatively large triples contributions (up to −13.6 kcal mol$^{-1}$ for $\overline{\text{B}-\text{Po}-\text{B}-\text{Po}-\text{B}-\text{Po}}_{\text{COV}}$ with vtz-f12), which is often about 0.2–0.5 kcal mol$^{-1}$ too small when computed with the vtz-f12 or awcvtz basis sets.

### 3.5 Generation of reference dimerization energies

The conclusion from the tests in the previous sections is that likely the basis set problem will cause the largest uncertainties of the reference values for the *IHD302* benchmark set. Therefore, it would in principle be desirable to use the awcvqz basis for the reference calculations. Unfortunately, the computational effort for PNO-LCCSD(T)-F12 calculations with this basis is clearly too large for the complete benchmark set. For example, a PNO-LCCSD(T)-F12/awcvqz calculation for $\overline{\text{In}-\text{Te}-\text{Tl}-\text{Te}-\text{In}-\text{Se}}_{\text{COV}}$ ($\mathbf{5}_{\mathbf{COV}}$) (2202 basis functions, TIGHT domain settings) took about 66 hours on 4 (rather old) Intel® Xeon® CPU E5-2650 v4 @ 2.20 GHz nodes with a total of 74 processing cores.

The large computational cost for these rather small molecules (up to 24 atoms) may be surprising at first sight. Partly, it is due to the large number of orbitals per atom that have to be correlated (7–8 LMOs per atom as compared to typically 2–3 LMOs for organic molecules); without pair or triples approximations, the computational effort scales cubic with the number of LMOs per atom. Furthermore, in these compact systems, there are no distant pairs, and screening as well as weak pair approximations have a much smaller impact than for extended organic molecules. And last but not least, the many high-angular momentum functions in the basis sets needed for post-d elements strongly increase the cost of the integral evaluations.

Due to its beneficial accuracy-cost ratio and because also awcvtz (PROJECTOR=MIXED,THRPNO_LMP2=1d-10) calculations for the whole *IHD302* set would take several weeks of computation time, we decided to use the PNO-LCCSD(T)-F12/TIGHT/vtz-f12(corr.) (*cf.* (2) with $\gamma_{\text{vv}} = 1.0a_0{}^{-1}$ and $\gamma_{\text{cc}} = \gamma_{\text{cv}} = 1.4a_0{}^{-1}$; abbreviated as PNO-LCCSD(T)-F12/vtz-f12(corr.) in the following) for computing the dimerization energies for the whole benchmark set (see ESI,† for details).

According to all test calculations carried out in this work, most of the resulting values should be well within the target accuracy with a relative accuracy of 1–2% for the covalently bound systems and <5% for the weak-donor acceptor structures (*i.e.*, an absolute accuracy of 1–2 kcal mol$^{-1}$ and 0.5 kcal mol$^{-1}$, respectively). Nonetheless, larger uncertainties for a few systems cannot be fully excluded, particularly for cases where the correlation contributions are much larger than the final absolute values. These residual errors are mainly basis errors. Compared to the awcvqz values, the triples contributions are typically too positive by about 0.5 kcal mol$^{-1}$, *i.e.*, the dimerization energies are slightly underestimated. Since most approximate methods tested (*vide infra*) give an MD < 0 with MD ≈ MAD (see ESI,† for details), these deviations would tend to be slightly smaller with real CBS reference values. Since the deviations of even the best DFT methods from the reference values are usually significantly larger, the residual uncertainty in the reference values should not lead to a significant increase in the scatter of methods evaluated on the *IHD302*. However, the statistic discriminability (*i.e.*, the square root of the number of reference values times their estimated uncertainty and then divided by the number of systems) is about 0.1 kcal mol$^{-1}$. This allows for a unambiguous ranking of even the best DFT methods.

### 3.6 Trends throughout the periodic table

Before employing these reference values for the complete *IHD302* in the next section to assess more approximate methods, we will use them to discuss trends in dimerization energies as a function of mean atomic number of the involved heavy atoms $\bar{Z}_{\text{heavy}}$ (see Fig. 8).

The average $\Delta E = -95.5$ kcal mol$^{-1}$ of the covalent structures (Fig. 8a) is about four to five times larger than the respective value of the WDA structures (Fig. 8b) and also the range of the covalent dimerization energies is significantly larger (up to −165.1 kcal mol$^{-1}$ for $\overline{\text{Si}-\text{N}-\text{Si}-\text{N}-\text{Si}-\text{N}}_{\text{COV}}$). Nitrogen is the strongest Lewis base and especially in combination with lighter elements, this leads to large dimerization energies. In combination with heavier elements, which are less effective donors due to diffuse lone pairs, the $\Delta E$ becomes smaller. There are also a few systems with rather small negative or even positive $\Delta E$s (up to 9.3 kcal mol$^{-1}$ for $\overline{\text{B}-\text{P}-\text{B}-\text{P}-\text{B}-\text{N}}_{\text{COV}}$). These are exclusively systems with boron. Here, the planar monomers are presumably partially aromatic (comparable to borazine), resulting in little or no energetic benefit from dimerization. This effect becomes weaker in combination with heavier elements due to their lower hybridization degree, which leads to larger dimerization energies. There is, however, no clear trend of $\Delta E$ with increasing $\bar{Z}_{\text{heavy}}$. However, $\Delta E$s of the IV–V systems are generally larger on average than those of the III–VI systems (the $\Delta E$s of the III–V systems are typically in between).

The mean $\Delta E$ for the WDA systems is −12.4 kcal mol$^{-1}$, with all dimers bound ($\Delta E < 0$), although especially the heavier III–VI systems show rather large dimerization energies for non-covalently bound neutral systems (up to −29.2 kcal mol$^{-1}$ for $\overline{\text{Tl}-\text{Po}-\text{Tl}-\text{Po}-\text{Tl}-\text{Po}}_{\text{WDA}}$). This indicates that the WDA structures represent bonding situations between non-covalent and covalent. Only the III–VI systems show a rough trend towards a linear increase of dimerization energy with $\bar{Z}_{\text{heavy}}$, which is not recognizable for the III–V and IV–V systems.

Additionally, the PNO-LCCSD(T)-F12/vtz-f12(corr.) correlation and triples contributions to the dimerization energies as well as the basis set correction, cv+cc correlation, and F12 contributions at the PNO-LMP2-F12 level of theory are analyzed for trends with respect to $\bar{Z}_{\text{heavy}}$ in the ESI† (Fig. S1–S4).

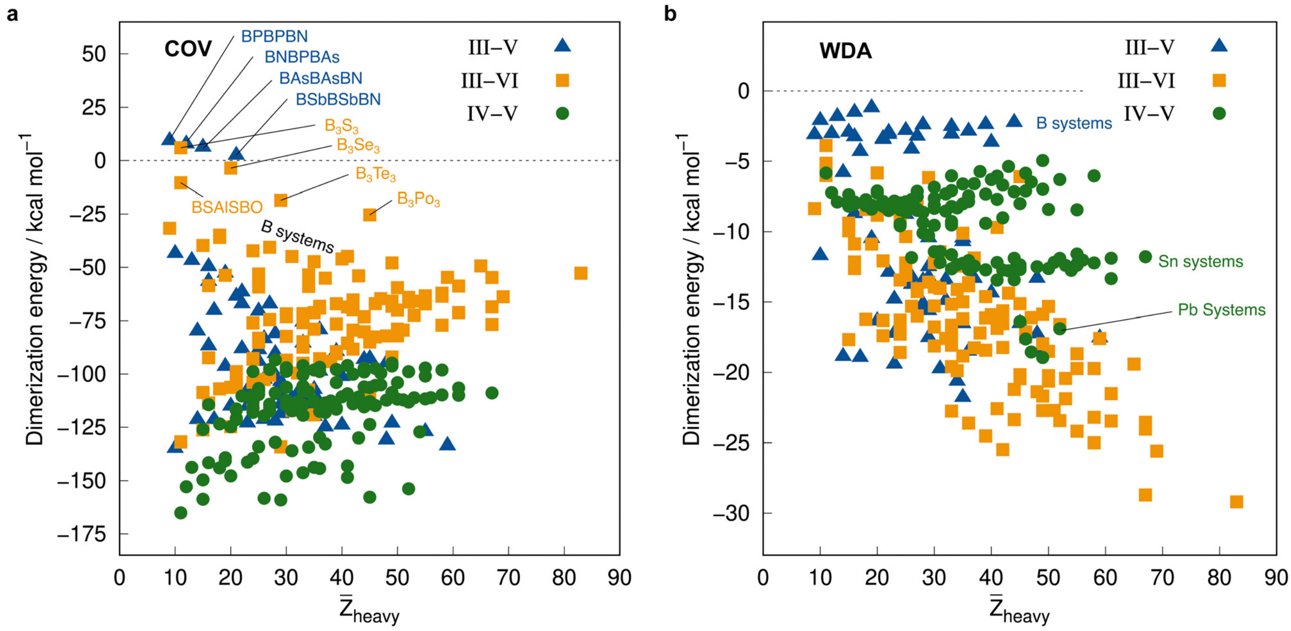


Fig. 8 Correlation plot of the PNO-LCCSD(T)-F12/vtz-f12(corr.) reference dimerization energies in kcal mol$^{-1}$ and the mean atomic number of the involved heavy atoms.

## 4 p-Block challenge 2: DFT and SQM calculations

Based on the previously discussed PNO-LCCSD(T)/vtz-f12(corr.) reference dimerization energies, various contemporary and commonly used density functional theory (DFT) and semi-empirical quantum mechanical (SQM) methods have been assessed for their performance on the *IHD302* benchmark set (Table 4). Seven (*meta*-)GGA, ten hybrid, and nine double-hybrid functionals (26 in total) were tested together with three different dispersion correction schemes (D3 with Becke–Johnson damping[135,136] or zero-damping for the Minnesota functionals, D4,[137–139] and the non-local DFT-NL[140,141] variant of Vydrov and van Voorhis' VV10[142] model). Furthermore, five composite DFT approaches and five SQM methods were covered. Because most force field methods are based on a separation of covalent and non-covalent interaction terms, their application for the here-considered situation in between both makes little sense although machine-learned potentials may be tested in the future if available for all considered elements.

All respective calculations were conducted with the ORCA 5.0.4,[31] xTB 6.5.1,[143] Molpro_2024.1,[47,48] AMS 2023.104,[144,145] DIRAC23,[146,147] Turbomole 7.6,[148] Q-Chem 5.4,[149] and MOPAC 2016,[150] program packages. Initially, all non-composite DFT functionals were evaluated in combination with Ahlrichs def2-QZVPP[151] basis sets. To further analyze the DFT basis set errors, PBE0-D4 calculations were also carried out with the awcvtz and awcvqz basis sets (*cf.* Section 3.3), because PBE0-D4 showed a significant improvement in the MAD for the covalent subset. The different basis sets yield MADs of $\text{MAD}_{\text{def2-QZVPP}}$ = 8.8, $\text{MAD}_{\text{awcvtz}}$ = 7.6, and $\text{MAD}_{\text{awcvqz}}$ = 7.4 kcal mol$^{-1}$. For the WDA subset, the basis sets perform equally and can be considered converged, making further counter-poise corrections obsolete. Nevertheless, no clear trend of general improvement is observed by using awcvtz for many other tested funtionals (see ESI†), which can in part be attributed to beneficial error compensation. Part of the different results obtained from both basis set families is because the calculations with awcvnz-pp basis sets use relativistic small-core pseudopotentials for all elements above $Z = 19$, and thus account implicitly for scalar relativistic effects. Unfortunately, for 4th row elements, this is not the case with the def2-QZVPP basis sets, which employ ECPs only for $Z > 36$ by default. This leads to rather large (up to 6.0 kcal mol$^{-1}$) relativistic errors for some systems containing the elements Ga–Se. Detailed tests using the PBE0-D4 functional showed that safely-converged DFT results are obtained with the awcvqz-pp basis sets (see ESI†), but obviously these basis sets are not designed for DFT and much larger than necessary (and thus too expensive). Using the standard valence avtz-pp or avqz-pp basis sets, however, still lead for some systems to significant (up to 1.0 kcal mol$^{-1}$) basis set errors relative to calculations with awcvqz. These errors are partly due to the basis set contractions, which are based on HF rather than DFT calculations. Recontracting the basis sets using atomic KS/PBE0 calculations significantly reduces the errors, in particular for systems containing group 13 and 14 elements. The contraction schemes were kept exactly as in the original avnz-pp basis sets. We denote the recontracted basis sets as aug-cc-pVnZ-PP-KS (short avnz-pp-ks, provided as additional ESI,† file). Calculations with the avqz-pp-ks sets (coefficients can be found in the ESI†) yielded excellent agreement with the awcvqz results. However, with the avtz-pp-ks sets errors of 0.5 kcal mol$^{-1}$ remained for some systems. It turned out that these errors are mainly due to missing tighter *f* and *g* functions. We therefore recommend to use the avqz-pp-ks basis sets for all elements

Table 4 Root mean square deviations (RMSDs) for all tested methods in kcal mol$^{-1}$. Except for the FF, SQM, and DFT composite methods, the def2-QZVPP basis set was applied. The values of further statistical descriptors are given in the ESI

| | | RMSD/kcal mol$^{-1}$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Covalent | | | | Weak donor–acceptor | | | |
| Class | Method | Plain | D3 | D4 | NL | Plain | D3 | D4 | NL |
| FF | GFN-FF[99] | | | 184.3 | | | | 9.1 | |
| | UFF[100] | 1416.3 | | | | 1.0 | | | |
| SQM | PM6-D3H4[23,101] [a] | | 158.8 | | | | 33.2 | | |
| | PM7[24] [a] | | 162.7 | | | | 17.9 | | |
| | GFN0-xTB | | | 52.0 | | | | 6.9 | |
| | GFN1-xTB[25] | | 46.7 | | | | 7.6 | | |
| | GFN2-xTB[26] | | | 36.8 | | | | 6.7 | |
| Composite | B97M-V-C[102] | | | | 12.7 | | | | 4.1 |
| | B97-3c[103] | | 9.3 | | | | 6.8 | | |
| | $r^2$SCAN-3c[30] | | | 5.7 | | | | 1.5 | |
| | PBEh-3c[104] | | 10.0 | | | | 1.8 | | |
| | $\omega$B97X-3c[105] | | | 8.0 | | | | 1.1 | |
| | HF-3c[106] | | 33.6 | | | | 9.0 | | |
| (*Meta-*)GGA | PBE[107] | 16.3 | 7.3 | 7.4 | | 8.3 | 1.0 | 2.2 | |
| | BP86[108,109] | 23.3 | 11.6 | 8.6 | | 10.9 | 10.3 | 7.7 | |
| | B97M[110] | | | | 8.8 | | | | 2.2 |
| | TPSS[111] | 14.9 | 7.0 | 9.6 | | 9.5 | 3.9 | 4.7 | |
| | $r^2$SCAN[112–115] | 8.1 | 4.3 | 4.8 | | 4.7 | 1.0 | 1.5 | |
| | M06-L[116] | 12.9 | 12.7 | 11.7 | | 1.8 | 1.7 | 1.6 | |
| | MN15-L[117] | 14.4 | 14.4 | | | 3.2 | 3.2 | | |
| Hybrid | PBE0[118] | 7.4 | 9.1 | 9.7 | 9.2 | 7.2 | 2.9 | 3.2 | 0.6 |
| | B3LYP[119,120] | 33.5 | 8.2 | 9.6 | | 12.9 | 4.2 | 3.9 | |
| | TPSSh[121] | 11.4 | 8.7 | 9.6 | | 9.0 | 4.8 | 4.6 | |
| | $r^2$SCAN0[122] | 5.0 | 5.6 | 5.1 | 6.2 | 4.5 | 1.3 | 1.0 | 1.0 |
| | M06[123] | 7.7 | 6.8 | 5.8 | | 1.2 | 1.6 | 2.2 | |
| | M06-2X[123] | 6.9 | 6.7 | | | 1.7 | 2.0 | | |
| | MN15[124] | 12.9 | 12.9 | | | 2.2 | 2.2 | | |
| | PW6B95[125] | 9.8 | 5.8 | 8.0 | | 6.1 | 3.4 | 3.6 | |
| | $\omega$B97X[126] | | | | 8.7 | | | | 1.1 |
| | $\omega$B97M[127] | | | | 4.8 | | | | 2.1 |
| Double-hybrid | revDSD-PBEP86[128] | 4.8 | 7.5 | 2.8 | | 2.4 | 1.5 | 0.8 | |
| | PWPB95[129] | 3.8 | 13.1 | 3.6 | | 3.4 | 4.8 | 1.7 | |
| | $\omega$B97M(2)[130] | | | | 9.7 | | | | 3.0 |
| | $\omega$B97X-2[131] | | | | 7.0 | | | | 2.1 |
| | P$r^2$SCAN50 | | | 5.0 | 6.7 | | | 1.1 | 1.0 |
| | $\kappa$P$r^2$SCAN50 | | | 7.2 | | | | 1.0 | |
| | $\omega$P$r^2$SCAN50 | | | 7.1 | 6.9 | | | 1.6 | 0.5 |
| | SOS0-PBE0-2[132,133] | 7.6 | 11.2 | | | 1.8 | 1.7 | | |
| | B2NC-PLYP[133,134] | 18.0 | 19.7 | | | 6.4 | 7.9 | | |

[a] As the PM$x$ methods are not parameterized for Po, no data for Po-containing systems are included.

containing elements of the 4th period, and def2-QZVPP basis sets for all other atoms. For a subset, which contains all systems with elements of the 4th period (IHD201), this choice gave excellent agreement with corresponding awcvqz calculations (*cf.* Table 5). For the WDA dimers, the improvement of the avqz-pp-ks set is negligible. As these tests were only done during our revision process we will discuss mostly "pure" def2-QZVPP results in this work. We note that redesigned basis sets for DFT were very recently proposed by Determan and Wilson,[152] but since their work appeared only after submission of our manuscript these were not tested in the current work.

The semi-numerical chain-of-spheres integration for the Fock exchange integrals (RIJCOSX)[153] has been applied with matching auxiliary basis sets (def2/J)[154] which is the default in the ORCA program package. Nevertheless, tests on the def2/J basis showed that a auxiliary basis set error of up to 0.6 kcal mol$^{-1}$ persists for the *IHD302* benchmark set. These errors are due to the comparably small size of the def2/J auxiliary basis when used with very large, polarized basis sets like def2-QZVPP. Employing the larger def2/JK auxiliary basis set for both the Coulomb and exchange part eliminates this error and yields very similar results (neglibible RI error) as obtained with RIJK density fitting using def2/JK (see ESI,† for details). Therefore, it is recommended to use the def2/JK fitting basis sets for both J and K for functionals containing exact exchange. Root mean square deviations (RMSDs) for all tested methods are given in Table 4. A detailed statistical analysis of the results is presented in Section 4 (for the detailed data, see ESI†).

Table 5 Mean deviation (MD), mean absolute deviation (MAD), standard deviation (SD) and absolute maximum error (AMAX) in kcal mol$^{-1}$ for the *IHD201* subset and relative wall times for the single point calculation of the 'Ga–N–Ga–N–Ga–N'$_{\text{COV}}$ dimer. The *IHD201* subset consists of all fused systems that contain elements of the 4th period

| | def2-QZVPP | avqz-pp-ks | awcvtz | awcvq |
|---|---|---|---|---|
| MD | −9.4 | −7.6 | −7.8 | −7.5 |
| MAD | 9.5 | 7.7 | 8.0 | 7.7 |
| SD | 3.9 | 3.7 | 3.9 | 3.8 |
| AMAX | 15.8 | 16.5 | 16.8 | 16.9 |
| **Relative wall time** | **1.00** | **0.77** | **0.49** | **1.46** |

### 4.1 Assessment of DFT methods

Results for selected DFT methods evaluated in the extended def2-QZVPP basis set are depicted in Fig. 9. For the covalent subset of the *IHD302*, several trends were observed: (i) climbing Perdew's "Jacob's ladder" of DFT results in increased (more negative) dimerization energies within the same underlying exchange–correlation functional and (ii) the indispensable combination of already attractive functionals with D3, D4, and NL London dispersion corrections results in overestimation of the dimerization energies (Fig. 9a). We note that in comparison to D3, the newer D4 correction additionally includes charge-dependency and the three-body ATM term

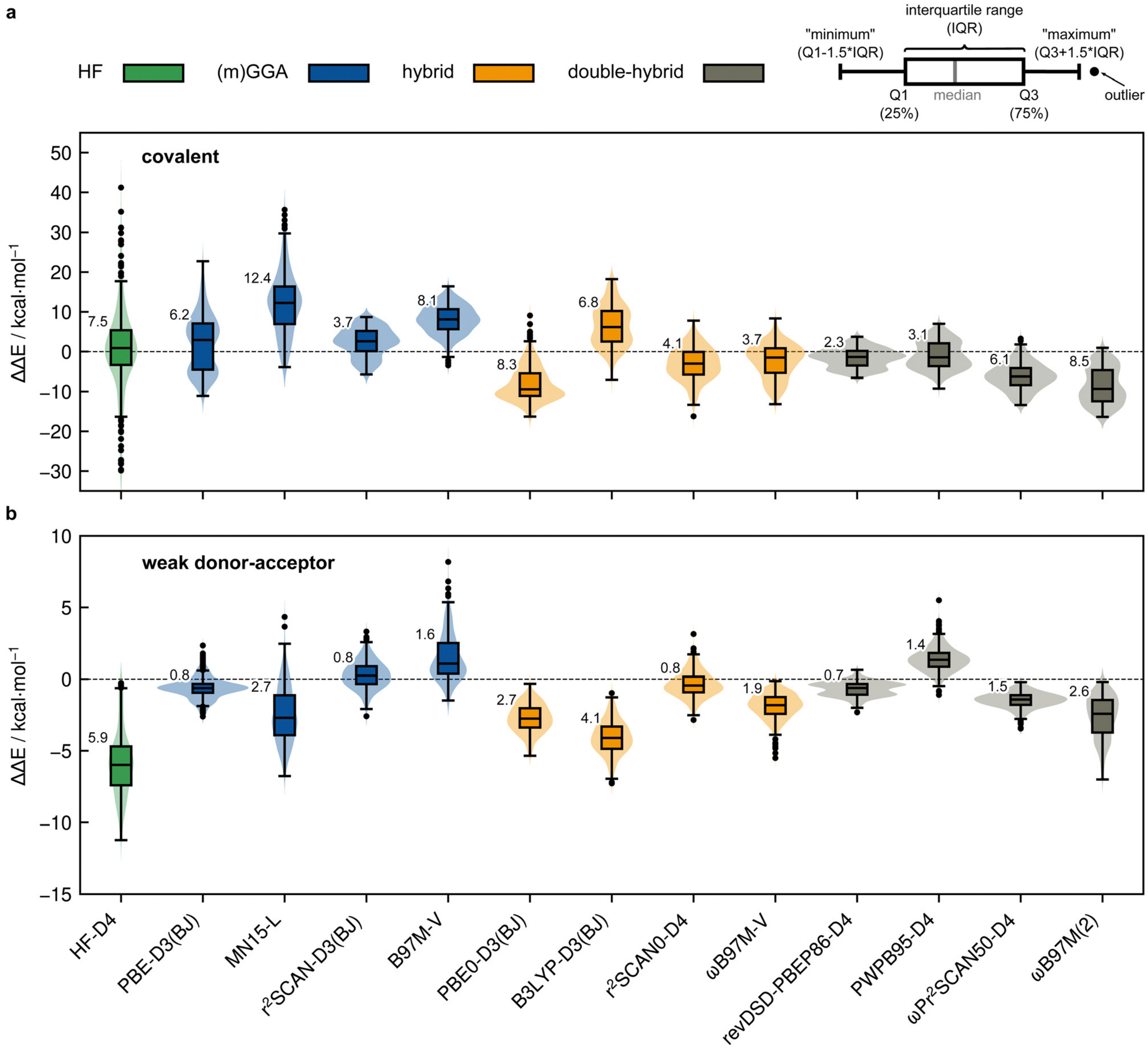


Fig. 9 Violinplots with boxplots of the deviations from the PNO-LCCSD(T)-F12/vtz-f12(corr.) reference for selected DFT methods for (a) covalent dimerization and (b) weak donor–acceptor interaction in kcal mol$^{-1}$ calculated with the def2-QZVPP basis set. Results for density functionals are shown for the most accurate combination between density functional and D3/D4 London dispersion correction. MN15-L results are not noticeably improved by the dispersion correction. HF-D4 was evaluated with the awcvtz basis set including the singles CABS correction (see Section 3.3). The numbers at the boxes are the MADs of the respective method.

but both are not relevant for the investigated structures. Therefore, there is no general preference for D4 over D3 for this benchmark set, nor over VV10, but it depends largely on the fitted parameters and how well they perform in combination with the functional parameters. The overestimation of the dimerization energies is manifested in a switch of the sign of the MD upon dispersion correction for functionals such as PBE0 ($\mathrm{MD_{plain}}$ = 5.9 kcal mol$^{-1}$, $\mathrm{MD_{D3(BJ)}}$ = −8.0 kcal mol$^{-1}$) or TPSSh-D4 ($\mathrm{MD_{plain}}$ = 10.9 kcal mol$^{-1}$, $\mathrm{MD_{D4}}$ = −8.4 kcal mol$^{-1}$). This results in mostly negative mean deviations for the dispersion-corrected functionals and is most pronounced for hybrid and double-hybrid functionals. This overbinding is also observed for NL ($\mathrm{MD_{PBE0\text{-}NL}}$ = −8.3 kcal mol$^{-1}$), thus largely excluding systematic issues of one of the dispersion corrections. An exception from this behavior is observed for very repulsive functionals like B3LYP. Some functionals that consider dispersion explicitly during the parameterization process such as Head-Gordon's ωB97M-V perform relatively well for the *IHD302* dimerization energies but still tend to overbind. The Minnesota-type functionals that indirectly cover some medium-ranged dispersion effects *via* their parameterization consistently underestimate the dimerization energies in the range of $\mathrm{MD_{M06\text{-}2X}}$ = 3.2 kcal mol$^{-1}$ to $\mathrm{MD_{MN15\text{-}L}}$ = 12.1 kcal mol$^{-1}$. The D3(0) correction only slightly reduces this underestimation ($\mathrm{MD_{M06\text{-}2X\text{-}D3(0)}}$ = 2.9 kcal mol$^{-1}$).

The observed overall overbinding trend of the explicitly dispersion-corrected functionals is mostly preserved for the weak donor–acceptor interacting dimers (*cf.* Fig. 9b) with the Minnesota functionals also adopting the behavior of the other functionals by yielding negative MDs. Nevertheless, the explicit treatment of London dispersion is highly recommended as it still significantly reduces the error for all tested (*meta-*)GGA functionals and most (double-)hybrid functionals. This is in line with the fact that DFT methods by construction do not cover London dispersion interactions (only potentially implicitly by parameterization), which is why corresponding corrections are indispensable for getting the right answer for the right reason which also holds true here. The D3, D4 and NL London dispersion corrections are also mostly in good agreement with each other. The most significant difference is found for PWPB95 which is an exception that can be traced back to its unsuited parameterization. It is hence recommended to check at least two different dispersion correction schemes for complicated systems. In the following, we only discuss dispersion-corrected approaches. Overall, both the covalent and also the weak donor–acceptor subset suffer from the over-attractive nature of many tested functionals with the latter being less affected. Here, the best-performing dispersion-corrected functionals regarding the covalent subset are $r^2$SCAN-D3(BJ) ($\mathrm{RMSD_{COV}}$ = 4.3 kcal mol$^{-1}$; $\mathrm{RMSD_{WDA}}$ = 1.0 kcal mol$^{-1}$) in the (*m*)GGA class, ωB97M-V ($\mathrm{RMSD_{COV}}$ = 4.8 kcal mol$^{-1}$; $\mathrm{RMSD_{WDA}}$ = 2.1 kcal mol$^{-1}$) and $r^2$SCAN0-D4 ($\mathrm{RMSD_{COV}}$ = 5.1 kcal mol$^{-1}$; $\mathrm{RMSD_{WDA}}$ = 1.0 kcal mol$^{-1}$) in the hybrid class, and revDSD-PBEP86-D4 ($\mathrm{RMSD_{COV}}$ = 2.8 kcal mol$^{-1}$; $\mathrm{RMSD_{WDA}}$ = 0.8 kcal mol$^{-1}$) in the double-hybrid class. Nevertheless, we note that the accuracy for the covalent subset is strongly functional dependent and even the best performing among the assessed functionals, which show reasonably small statistical errors, still yield significant outliers with several kcal mol$^{-1}$ deviation from the reference dimerization energies thus limiting their scope if highly accurate predictions are required.

In summary, the introduction of physically motivated ingredients into the density functional upon climbing Jacob's ladder in combination with a well-established dispersion correction yields an untypically small benefit for the covalent subset. This indicates less effective residual error compensation between the various contributions to the total energy for functionals that yield accurate results for other benchmark studies. Although the dispersion corrections tested at shorter distances (attenuation region) are somewhat less accurate than in the long-range London regime, they are still qualitatively correct and physically needed. Hence, the general trend for the tested DFT methods to overestimate the dimerization energies hints at an over-attractive nature of many usually well-behaved functionals that manifests for this congested situation of many spatially close bonds and partially repulsive regimes covered by the *IHD302* structures. In this case, only very repulsive functionals such as B3LYP still underestimate the covalent dimerization energies despite large dispersion corrections that cannot fully compensate for the functionals' comparably bad overall performance.

Fortuitous error compensation for less repulsive functionals applied without dispersion correction and thus misleading conclusions regarding their need and overall accuracy has been discussed previously.[155] These results further underline that the *IHD302* benchmark set challenges established and/or outdated design and parameterization strategies of density functionals. Specifically, fitting a robust dispersion correction directly in the process of the functional fit seems to be important to guarantee a well-balanced match of exchange–correlation functional and dispersion contributions, also for shorter distances than the usually evaluated London regime. This view is supported by the relatively good performance of Head-Gordon's ωB97M-V functional that is trained in the presence of the VV10 non-local dispersion correction. As ωB97M-V still systematically overbinds the investigated systems on average, incorporating more high-level benchmark sets that focus on the transition from non-covalent to covalent bonding as in the *IHD302* set may further improve its performance. Inclusion of explicitly repulsive contacts into the parameterization process as provided in the R160x6,[156] R739,[157] or SH250[158] test sets and the here introduced *IHD302* benchmark set may therefore be beneficial to avoid over-attractiveness of the respective functional for comparable systems even though the element composition may differ. Moreover, residual errors of dispersion corrections in the medium- to short-ranged regime (assumed to be about 10–20% due to additional uncertainty of the damping function and the asymptotic errors of about 5% in the $C_6$ coefficients)[138] may be further reduced by inclusion of accurate fit data also covering repulsive p-block contacts.

To classify the DFT results, dispersion-corrected Hartree–Fock (HF-D4) dimerization energies close to the HF basis set limit (awcvtz with CABS singles correction, see Section 3.3) will be discussed here. HF is slightly less repulsive than, *e.g.*, B3LYP, but cannot show over-attractiveness as it is parameter-free and medium to long-range correlation effects are only accounted for posteriorly by the D4 correction. Consistently, the dimerization energies for the WDA structures are overestimated more strongly than, *e.g.*, with B3LYP, which is due to the less accurate description of the short- and medium-range correlation effects (*i.e.*, in the damping range) with the D4 dispersion correction. The deviations from the reference values are negative in all cases up to $-11.2$ kcal mol$^{-1}$ for $\overline{\text{Pb–P–Pb–P–Pb–P}}_{\text{WDA}}$ (one of the test systems for the scans analyzed in the next section). In contrast to most of the tested DFT methods, HF-D4 does not produce a systematic error for the COV structures (MD $\approx$ 0), whereas the scatter of the deviations from the reference is significantly larger, because the D4 correction describes the formed σ bonds (to the H atoms or from the lone pairs) approximately well, but more difficult cases are not adequately taken into account. This is particularly the case for the IV–V systems with N (up to $-39.8$ kcal mol$^{-1}$ deviation from the reference for $\overline{\text{Sn–N–Sn–N–Sn–N}}_{\text{COV}}$, which also has the most negative HF contribution to the dimerization energy of all *IHD302* systems with $-164.6$ kcal mol$^{-1}$) and for III–VI systems with boron (up to +41.2 kcal mol$^{-1}$) deviation from the reference for $\overline{\text{B–Po–B–Po–B–Po}}_{\text{COV}}$ with a large positive HF contribution to the dimerization energy; this system also has the largest contribution of the correlation energy to $\Delta E$ (see ESI,† data_correlation_methods.xlsx).

To analyze the observed errors of the dispersion-corrected DFT methods in greater detail, specifically in the more repulsive

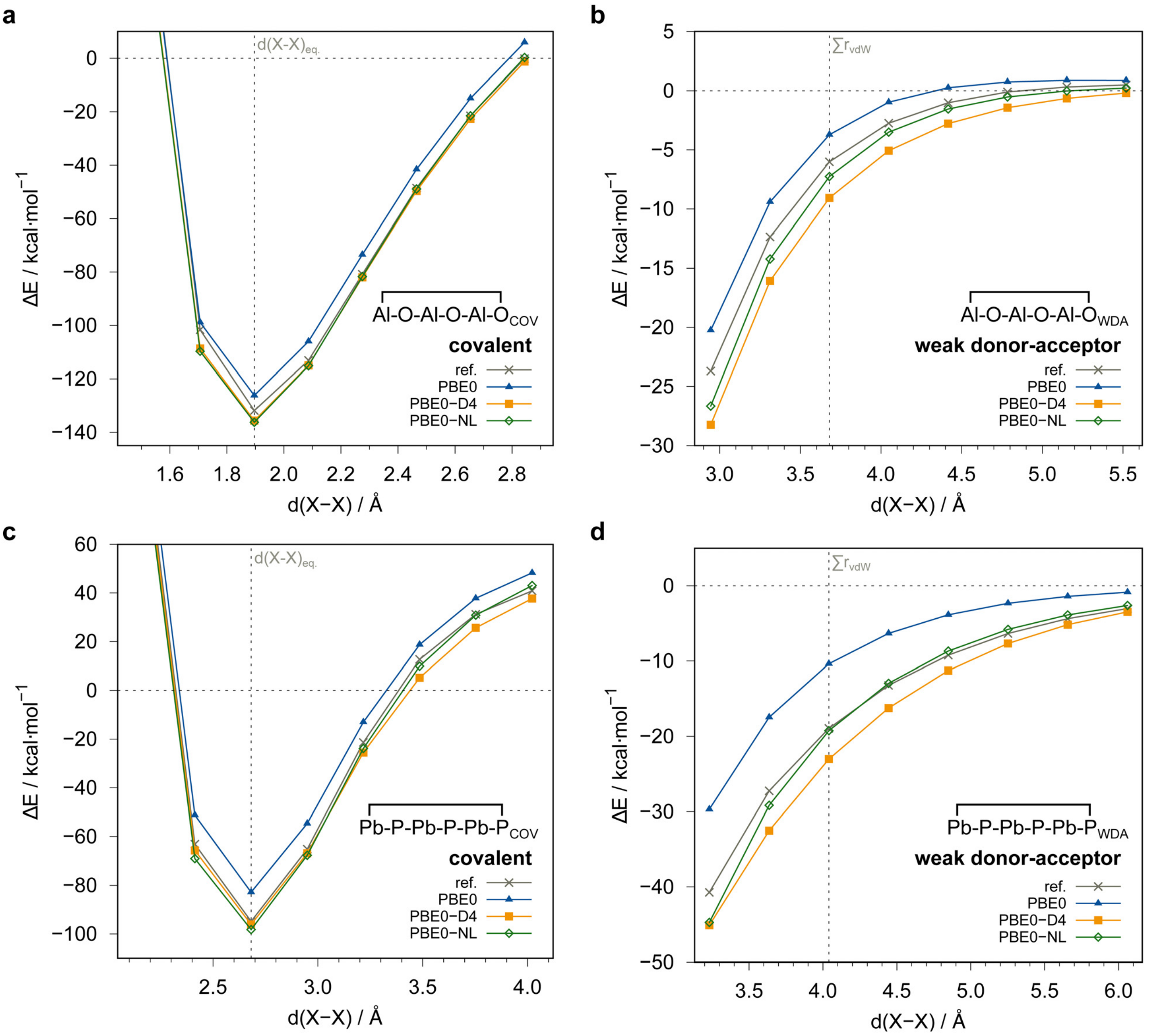


Fig. 10 Rigid potential energy scans of the covalent and weak donor–acceptor dimers with respect to the planar monomer fragments. $d$(X–X) is the center-of-ring-atoms inter-fragment distance. The grey reference curve was obtained with PNO-LCCSD(T)-F12/awcvtz with PROJECTOR=MIXED and TRHPNO_LMP2=1d-10. All energies in kcal mol$^{-1}$.

region of the PES, we conducted exemplary rigid potential energy surface scans of the inter-fragment distance for both dimer structures of $\overline{\mathrm{Al{-}O{-}Al{-}O{-}Al{-}O}}$ and $\overline{\mathrm{Pb{-}P{-}Pb{-}P{-}Pb{-}P}}$, respectively (Fig. 10). All scans clearly show that the PBE0 functional, even though not considered a particularly repulsive functional, systematically underestimates the interaction energies for both the covalently and the weak donor–acceptor bound dimers. This underlines the need for adding a (London) dispersion correction to achieve agreement with the reference (PNO-LCCSD(T)-F12/awcvtz with PROJECTOR=MIXED and TRHPNO_LMP2=1d-10) PES curve. Despite a slight over-correction with the NL and D4 dispersion corrections is observed for both systems, their asymptotic behavior is fully correct. These results confirm also for such rather unusual systems that the residual error of properly dispersion-corrected DFT methods is mainly originating from the DFT exchange–correlation functional and less from uncertainties of the dispersion corrections at the short- and medium-range distances. Overall, the overestimation is most pronounced at short inter-fragment distances, while the error decreases in the asymptotic region. Further, the scans of the weak donor–acceptor dimers underline their non-equilibrium nature by not showing a minimum at the sum of van der Waals radii due to the donor–acceptor character of the interaction that is distinct

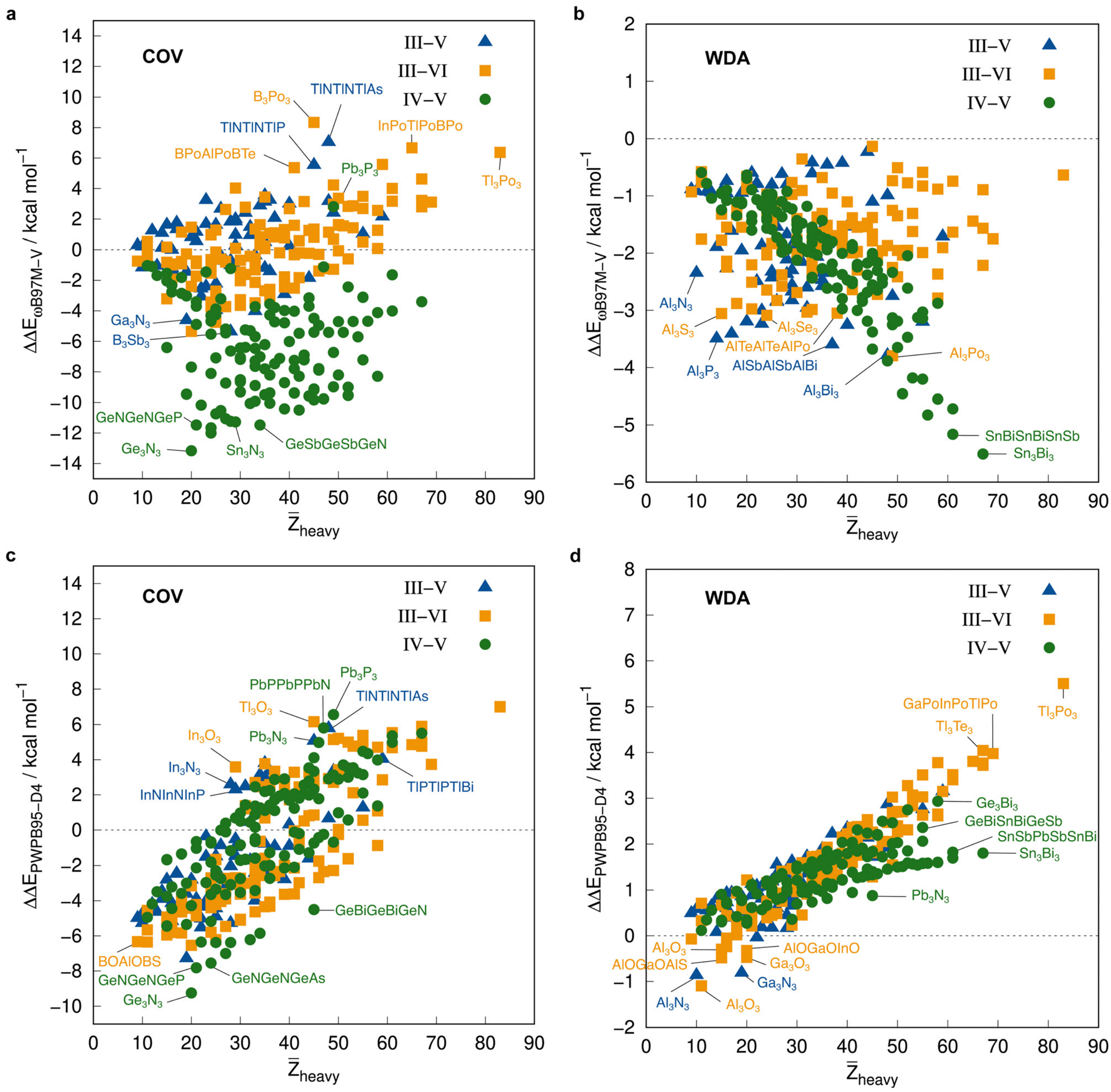


Fig. 11 Correlation plot of the DFT dimerization energies in kcal mol$^{-1}$ and the mean atomic number of the involved heavy atoms for covalent and weak donor–acceptor bound systems computed with $\omega$B97M-V (a) and (b), and PWPB95-D4 (c) and (d), respectively.

from typically discussed non-covalent interactions at larger separation (London regime).

Besides these systematic trends, it is observed that the functionals behave significantly differently with respect to the element composition of the system (Fig. 11). For example, the extent of the error of the dimerization energies obtained with PWPB95-D4, an overall well-performing DFA, does not follow the trends of the reference dimerization energies (*cf.* Fig. 8). Instead, the error becomes systematically more positive with increasing mean atomic number $\bar{Z}_{\text{heavy}}$ for both dimer types and all main group combinations. For $\omega$B97M-V, the error trends are much less systematic with the error distribution mostly following the trends of the reference dimerization energies for the covalent dimers, while mostly the IV–V element combination yields an inverted trend compared to PWPB95-D4.

### 4.2 Composite methods

As DFT calculations with large basis sets can become computationally unfeasible for large molecules, composite approaches that employ smaller basis sets together with tailored corrections have proven to yield accurate results at a fraction of computation time for various chemical systems. The most prominent composite methods are the so-called "3c" methods including PBEh-3c,[104] B97-3c,[103] $r^2$SCAN-3c,[30] and $\omega$B97X-3c.[105] The "3c" methods employ relatively small tailored basis sets and specifically adjusted corrections like D4 and a geometrical counter-poise correction to achieve good accuracy at much reduced computational cost. Another notable addition is Head-Gordon's B97M-V-C approach which combines the B97M-V functional with a def2-SVPD basis set and a tailored DFT-C counter-poise correction.[102] For the covalent subset of the *IHD302*, comparable trends as for the non-composite functionals evaluated with the extended def2-QZVPP basis are observed. The inclusion of Fock-exchange in PBEh-3c and in the recent $\omega$B97X-3c range-separated hybrid yields more negative covalent dimerization energies compared to the B97M-V-C, $r^2$SCAN-3c, and B97-3c *meta*-GGA approaches. Comparable results are obtained for the weak donor–acceptor subset, even though B97-3c tends to overbind these systems (Fig. 12). Overall, specifically the more recent composite methods $r^2$SCAN-3c and $\omega$B97X-3c perform quite well with RMSDs of 5.7 kcal mol$^{-1}$ and 8.0 kcal mol$^{-1}$ for the covalent and 1.5 kcal mol$^{-1}$ and 1.1 kcal mol$^{-1}$ for the weak donor–acceptor subset, respectively. Impressively, they perform similarly or even slightly better than the parent functionals evaluated in a quadruple-zeta basis set thus underlining their high robustness. The very good performance of $\omega$B97X-3c is specifically remarkable, as it employs large-core effective core potentials. The B97M-V-C approach is found to be less competitive, yielding significantly larger errors than its conventional DFT counterpart.

### 4.3 Semi-empirical quantum mechanical methods

Due to their low computational cost, semi-empirical quantum mechanical methods have become a central part of automated conformational sampling and reaction screening workflows.[159,160] Nevertheless, most of these methods are not or only sparsely parameterized for heavy p-block elements. Further, they typically yield limited accuracy for off-target elements and complicated electronic situations. Notable exceptions in terms of general applicability are the GFN$n$-xTB[19] and PM$x$[161] SQM families of methods. The corresponding error statistic is shown in Fig. 13. For the covalent subset of the *IHD302*, the GFN$n$-xTB methods yield RMSDs of 40.6 kcal mol$^{-1}$ and 36.8 kcal mol$^{-1}$, respectively for GNF1- and GFN2-xTB. The preliminary, non-self-consistent GFN0-xTB variant that is also used in the context of metallic clusters[162] yields an RMSD of 52.0 kcal mol$^{-1}$. These errors are one magnitude larger than those obtained with many well-performing DFT methods. This underlines the valuable focus of the *IHD302* on direct p-block element interactions that manifests the weaknesses of the respective SQM methods. In conventional benchmark sets that often employ organic

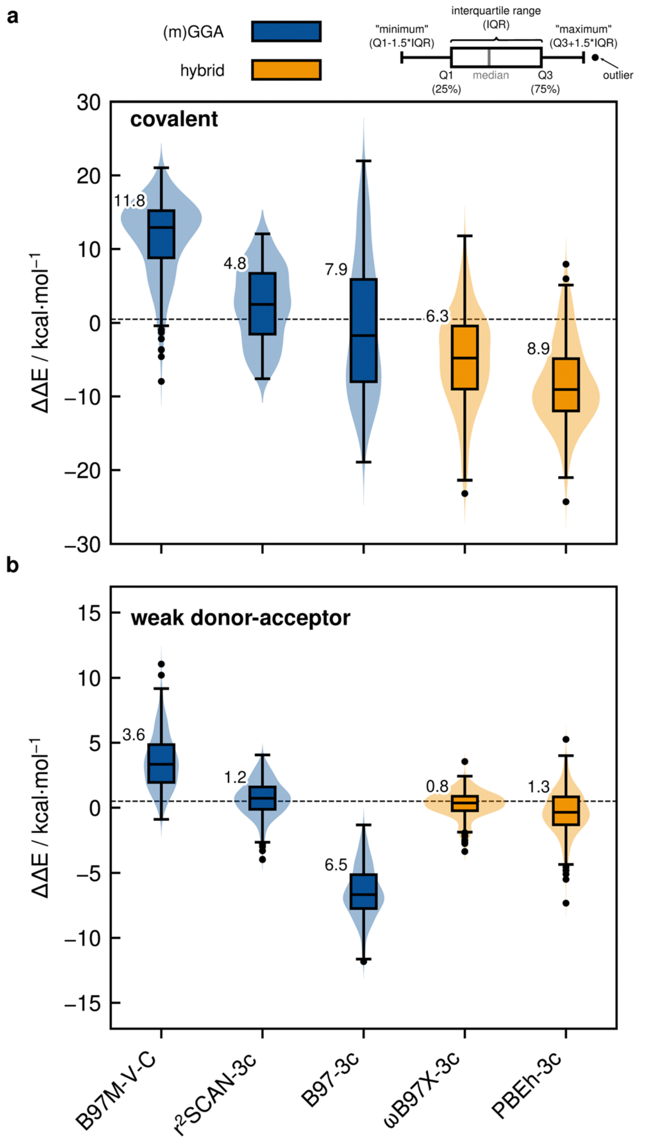


Fig. 12 Violinplots with boxplots of the deviations from PNO-LCCSD(T)-F12/vtz-f12(corr.) of five composite methods for (a) covalent dimerization and (b) weak donor–acceptor interaction in kcal mol$^{-1}$. The numbers at the boxes are the MADs of the respective method.

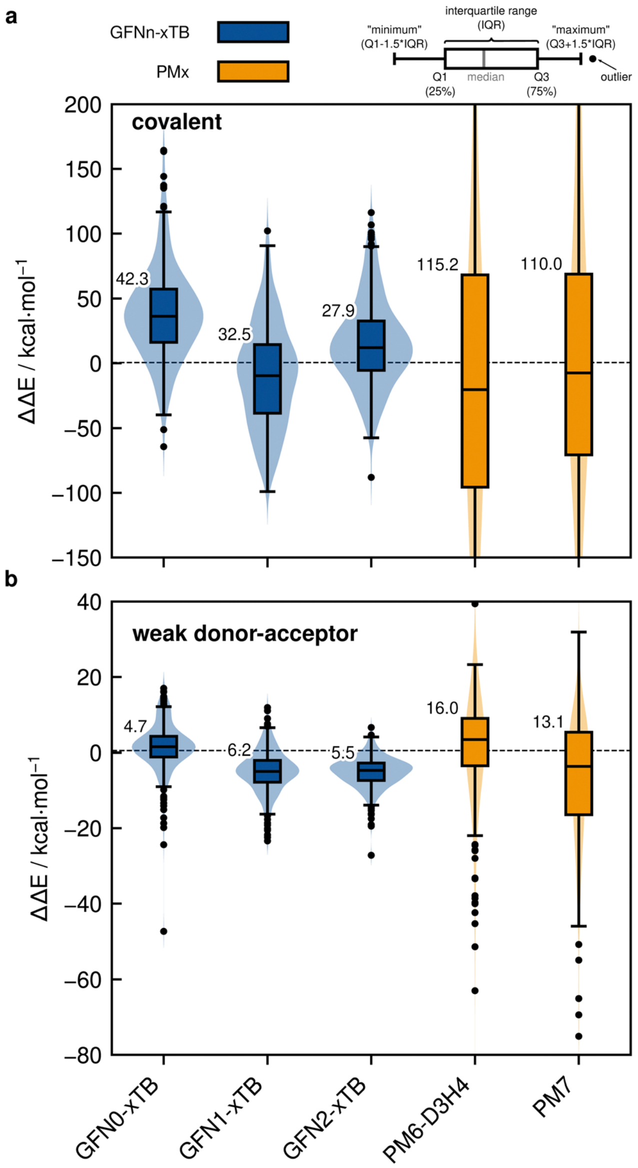


Fig. 13 Violinplots with boxplots of the deviations from PNO-LCCSD(T)-F12/vtz-f12(corr.) of five SQM methods for (a) covalent dimerization and (b) weak donor–acceptor interaction in kcal mol$^{-1}$. The numbers at the boxes are the MADs of the respective method.

substituents for saturation of the p-block element, these are compensated by the respective substituent interactions. Accordingly, the difference between DFT and SQM methods is usually considerably smaller. The Hartree–Fock based PM*x* methods yield much higher RMSDs of 158.8 kcal mol$^{-1}$ and 162.7 kcal mol$^{-1}$ for PM6 and PM7, respectively, even exceeding the mean dimerization energy of −95.5 kcal mol$^{-1}$ and therefore, these methods cannot be recommended. For the weak donor–acceptor subset, similar trends are observed for the GFN*n*-xTB and PM*x* methods. A comparison between the computational cost and the accuracy of different SQM and DFT methods can be found in the ESI† (Fig. S8). GFN2-xTB is, *e.g.*, up to 2700 times faster than the efficient $r^2$SCAN-3c composite DFT method in a single point calculation of $\overline{\text{Tl–Po–Tl–Po–Tl–Po}}_{\text{WDA}}$. Overall, none of the semi-empirical methods yields satisfying results for the complete *IHD302* benchmark. Nevertheless, GFN2-xTB may be considered a reasonable choice regarding the relatively large mean reference dimerization energies for the covalently bound dimers. Nevertheless, future improvements of SQM methods based on the presented benchmark data would be desirable, especially for modeling the metallic clusters of p-block elements.

## 5 Summary and perspective

In this work, we compiled a novel benchmark set of dimerization energies of six-membered purely p-block inorganic heterocycles. The benchmark set termed **I**norganic **H**eterocycle **D**imerizations **302** (*IHD302*) includes 302 energies for two distinct dimerization motifs, covalent and non-equilibrium weak donor–acceptor dimers. Reference values for these 604 dimerization energies were computed at the PNO-LCCSD(T)-F12 level. Due to large electron correlation contributions, the importance of core–valence correlation effects, and slow basis set convergence, these calculations turned out to be extremely challenging. We therefore carried out extensive benchmarks for 16 selected systems (eight of these were initially more or less randomly selected and further eight, which showed particularly large basis set errors, were added later). It was found that the dominant part of the errors were due to the slow basis set convergence, while errors caused by the local approximations are negligible.

At first, we employed the cc-pVTZ-PP-F12 basis sets of Hill and Peterson,[76] which were specifically optimized for F12 calculations on heavy p-block elements, including correlation of the $(n - 1)$d shell. It turned out, however, that these basis sets are too small for reliably obtaining the required accuracy. Even with F12 methods, deviations of up to about 3 kcal mol$^{-1}$ relative to more accurate calculations with the aug-cc-pwCVnZ (n = T, Q) basis sets were observed for the covalent binding energies. By employing a PNO-LMP2-F12 basis set correction denoted vtz-f12(corr.) and obtained with the aug-cc-pwCVTZ basis, the MADs and SDs of the 16 selected systems relative to the PNO-LCCSD(T)-F12/aug-cc-pwCVTZ and aug-cc-pwCVQZ results could be reduced to 0.26 ± 0.25 kcal mol$^{-1}$ (max error 0.87 kcal mol$^{-1}$) and 0.44 ± 0.28 kcal mol$^{-1}$ (max error 1.01 kcal mol$^{-1}$), respectively. For the weak donor–acceptor structures, the errors are even about three times smaller. The larger errors relative to the aug-cc-pwCVQZ basis mainly stem from the remaining basis set errors of the triples (T) contributions obtained with the TZ basis set. The errors of the vtz-f12(corr.) results are significantly smaller than the errors of the DFT methods. Therefore, we decided to compute the 602 reference values at the PNO-LCCSD(T)-F12/vtz-f12(corr.) level, which is computationally much cheaper than using the aug-cc-pwCVTZ or even the aug-cc-pwCVQZ basis sets.

Based on these reference data, the performance of 26 DFT functionals (evaluated with the def2-QZVPP basis set), five composite DFT approaches, and five semi-empirical quantum mechanical methods in combination with three different dispersion correction schemes (D3, D4, NL) were assessed. The best-performing dispersion-corrected functionals ordered by their rung on "Jacob's ladder" regarding the covalent subset are $r^2$SCAN-D4 (*meta*-GGA, $\text{RMSD}_{\text{COV}}$ = 4.8 kcal mol$^{-1}$; $\text{RMSD}_{\text{WDA}}$ = 1.5 kcal mol$^{-1}$), $\omega$B97M-V (hybrid, $\text{RMSD}_{\text{COV}}$ = 4.8 kcal mol$^{-1}$; $\text{RMSD}_{\text{WDA}}$ = 2.1 kcal mol$^{-1}$) and $r^2$SCAN0-D4 ($\text{RMSD}_{\text{COV}}$ = 5.1 kcal mol$^{-1}$; $\text{RMSD}_{\text{WDA}}$ = 1.0 kcal mol$^{-1}$), and revDSD-PBEP86-D4 (double-hybrid, $\text{RMSD}_{\text{COV}}$ = 2.8 kcal mol$^{-1}$; $\text{RMSD}_{\text{WDA}}$ = 0.8 kcal mol$^{-1}$). Further, the $r^2$SCAN-3c composite DFT method is found to yield good results regarding its reduced computational cost with $\text{RMSD}_{\text{COV}}$ = 5.7 kcal mol$^{-1}$ and $\text{RMSD}_{\text{WDA}}$ = 1.5 kcal mol$^{-1}$, respectively.

Furthermore, we found significant errors (up to 6 kcal mol$^{-1}$ in binding energies) due to missing scalar relativistic effects in calculations of molecules containing p-block elements of the 4th period, when using the def2-TZVPP or def2-QZVPP basis sets. This is because the def2 basis sets are associated to relativistic pseudo potentials only for elements with $Z > 36$. Further non-negligible errors (up to 1 kcal mol$^{-1}$) can occur in DFT calculations when using standard correlation consistent basis sets, such as aug-cc-pV*n*Z-PP, since these are contracted with coefficients from Hartree–Fock calculations. We therefore generated for the elements Ga–Se recontracted aug-cc-pVQZ-PP-KS basis sets, in which the contraction coefficients are taken from DFT (PBE0) calculations. Using these basis sets for 4th row elements in combination with the ECP10MDF pseudo potentials leads to significant improvements of the overall accuracy and hence, we recommend this basis sets for accurate DFT calculations of systems including 4th row elements.

The application of a London dispersion correction is found to be mandatory, even though rather attractive functionals tend to be over-corrected due to a less effective error compensation (underestimation of repulsive interactions out-weighted by missing attractive long-range dispersion interactions). This behavior underlines that London dispersion should be considered during the construction process of the respective functional to guarantee a well-balanced match of functional and dispersion correction. In this regard, the *IHD302* has proven to be a specifically indicative data set for the performance of dispersion-corrected functionals for p-block thermochemistry. The test set is also highly challenging for SQM methods as none of the tested (and currently available generally applicable) SQM methods is able to accurately reproduce the reference data, even though GFN2-xTB ($\text{RMSD}_{\text{COV}}$ = 36.8 kcal mol$^{-1}$; $\text{RMSD}_{\text{WDA}}$ = 6.7 kcal mol$^{-1}$) clearly outperforms, *e.g.*, PM7 ($\text{RMSD}_{\text{COV}}$ = 162.7 kcal mol$^{-1}$; $\text{RMSD}_{\text{WDA}}$ = 17.9 kcal mol$^{-1}$). Overall, the clear focus on the p-block element interactions without organic substituent bias allows for a unique and systematic assessment of quantum chemical methods for these elements. Regarding the high importance of p-block compounds, *e.g.*, in FLP chemistry or opto-electronics, the *IHD302* seems to represent a valuable assessment which may significantly impact future method development in this field, allowing for the development of more robust and transferable approximate quantum chemical methods. This is made possible not least by efficient and sophisticated implementations of explicit local correlation methods such as PNO-LCCSD(T)-F12, which was successfully applied as reference in this work.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

S. G. and M. B. gratefully acknowledge financial support of the Max Planck Society through the Max Planck fellow program. T. G. is thankful for fruitful discussions and technical support from Thomas Froitzheim and Lukas Wittmann. Open Access funding provided by the Max Planck Society.

## Notes and references

1 S. Aldridge and C. Jones, *Chem. Soc. Rev.*, 2016, **45**, 763–764.
2 L. Zhao, S. Pan, N. Holzmann, P. Schwerdtfeger and G. Frenking, *Chem. Rev.*, 2019, **119**, 8781–8845.
3 D. W. Stephan and G. Erker, *Angew. Chem., Int. Ed.*, 2015, **54**, 6400–6441.
4 A. R. Jupp and D. W. Stephan, *Trends Chem.*, 2019, **1**, 35–48.
5 M. Veith, *Chem. Rev.*, 1990, **90**, 3–16.
6 S. Nakarnura, *Semicond. Semimetals*, 1997, **48**, 391–443.
7 H. Asahi, *Infrared Detect. Emit. Mater. Devices*, Springer, Boston, MA, 2001, pp. 233–249.
8 K. Ota and R. Kinjo, *Chem. – Asian J.*, 2020, **15**, 2558–2574.
9 J. G. Ekerdt, Y. M. Sun, A. Szabo, G. J. Szulczewski and J. M. White, *Chem. Rev.*, 1996, **96**, 1499–1517.
10 F. C. Sauls and L. V. Interrante, *Coord. Chem. Rev.*, 1993, **128**, 193–207.
11 E. E. Foos, R. J. Jouet, R. L. Wells, A. L. Rheingold and L. M. Liable-Sands, *J. Organomet. Chem.*, 1999, **582**, 45–52.
12 A. M. Priegert, B. W. Rawe, S. C. Serin and D. P. Gates, *Chem. Soc. Rev.*, 2016, **45**, 922–953.
13 L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi and S. Grimme, *Phys. Chem. Chem. Phys.*, 2017, **19**, 32184–32215.
14 G. Bistoni, A. A. Auer and F. Neese, *Chem. – Eur. J.*, 2017, **23**, 865–873.
15 N. Mehta, T. Fellowes, J. M. White and L. Goerigk, *J. Chem. Theory Comput.*, 2021, **17**, 2783–2806.
16 J. Gorges, S. Grimme and A. Hansen, *Phys. Chem. Chem. Phys.*, 2022, **24**, 28831–28843.
17 T. Husch, A. C. Vaucher and M. Reiher, *Int. J. Quantum Chem.*, 2018, **118**, e25799.
18 F. Spiegelman, N. Tarrat, J. Cuny, L. Dontot, E. Posenitskiy, C. Martí, A. Simon and M. Rapacioli, *Adv. Phys. X*, 2020, **5**, 1710252.
19 C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher and S. Grimme, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2021, **11**, e1493.
20 A. Y. Timoshkin and G. Frenking, *Inorg. Chem.*, 2003, **42**, 60–69.

21 A. Mohajeri and M. Ebadi, *J. Phys. Chem. A*, 2012, **116**, 4678–4686.

22 H. Ni, D. M. York, L. Bartolotti, R. L. Wells and W. Yang, *J. Am. Chem. Soc.*, 1996, **118**, 5732–5736.

23 J. J. P. Stewart, *J. Mol. Model.*, 2007, **13**, 1173.

24 J. J. P. Stewart, *J. Mol. Model.*, 2013, **19**, 1–32.

25 S. Grimme, C. Bannwarth and P. Shushkov, *J. Chem. Theory Comput.*, 2017, **13**, 1989–2009.

26 C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652–1671.

27 E. Caldeweyher and J. G. Brandenburg, *J. Phys.: Condens. Matter*, 2018, **30**, 213001.

28 A. Bondi, *J. Phys. Chem.*, 1964, **68**, 441–451.

29 M. Mantina, A. C. Chamberlin, R. Valero, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. A*, 2009, **113**, 5806–5812.

30 S. Grimme, A. Hansen, S. Ehlert and J. M. Mewes, *J. Chem. Phys.*, 2021, **154**, 064103.

31 F. Neese, F. Wennmohs, U. Becker and C. Riplinger, *J. Chem. Phys.*, 2020, **152**, 224108.

32 F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2022, **12**, e1606.

33 H. Neugebauer, H. T. Vuong, J. L. Weber, R. A. Friesner, J. Shee and A. Hansen, *J. Chem. Theory Comput.*, 2023, **19**, 6208–6225.

34 S. Grimme and A. Hansen, *Angew. Chem., Int. Ed.*, 2015, **54**, 12308–12313.

35 C. A. Bauer, A. Hansen and S. Grimme, *Chem. – Eur. J.*, 2017, **23**, 6150–6164.

36 R. Nieman, J. R. Carvalho, B. Jayee, A. Hansen, A. J. A. Aquino, M. Kertesz and H. Lischka, *Phys. Chem. Chem. Phys.*, 2023, **25**, 27380–27393.

37 M. K. Kesharwani, N. Sylvetsky, A. Köhn, D. P. Tew and J. M. L. Martin, *J. Chem. Phys.*, 2018, **149**, 154109.

38 E. Ramos-Cordoba, P. Salvador and E. Matito, *Phys. Chem. Chem. Phys.*, 2016, **18**, 24015–24023.

39 H.-J. Werner, G. Knizia, C. Krause, M. Schwilk and M. Dornbach, *J. Chem. Theory Comput.*, 2015, **11**, 484–507.

40 Q. Ma and H.-J. Werner, *J. Chem. Theory Comput.*, 2015, **11**, 5291–5304.

41 M. Schwilk, Q. Ma, C. Köppl and H.-J. Werner, *J. Chem. Theory Comput.*, 2017, **13**, 3650–3675.

42 Q. Ma, M. Schwilk, C. Köppl and H.-J. Werner, *J. Chem. Theory Comput.*, 2017, **13**, 4871–4896.

43 Q. Ma and H.-J. Werner, *J. Chem. Theory Comput.*, 2017, **14**, 198–215.

44 Q. Ma and H.-J. Werner, *J. Chem. Theory Comput.*, 2020, **16**, 3135–3151.

45 Q. Ma and H.-J. Werner, *J. Chem. Theory Comput.*, 2021, **17**, 902–926.

46 H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby and M. Schütz, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 242–253.

47 H.-J. Werner, P. J. Knowles, F. R. Manby, J. A. Black, K. Doll, A. Heßelmann, D. Kats, A. Köhn, T. Korona, D. A. Kreplin, Q. Ma, I. Thomas, F. Miller, A. Mitrushchenkov, K. A. Peterson, I. Polyak, G. Rauhut and M. Sibaev, *J. Chem. Phys.*, 2020, **152**, 144107.

48 H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györffy, D. Kats, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, S. J. Bennie, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, S. J. R. Lee, Y. Liu, A. W. Lloyd, Q. Ma, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, T. F. Miller III, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang and M. Welborn, *MOLPRO, version 2023.1, a package of ab initio programs*.

49 F. Pavošević, P. Pinski, C. Riplinger, F. Neese and E. F. Valeev, *J. Chem. Phys.*, 2016, **144**, 144109.

50 F. Pavošević, C. Peng, P. Pinski, C. Riplinger, F. Neese and E. F. Valeev, *J. Chem. Phys.*, 2017, **146**, 174108.

51 H.-J. Werner and A. Hansen, *J. Chem. Theory Comput.*, 2023, **19**(20), 7007–7030.

52 M. Kallay, *J. Chem. Phys.*, 2015, **142**, 204105.

53 P. R. Nagy, G. Samu and M. Kallay, *J. Chem. Theory Comput.*, 2016, **12**, 4897–4914.

54 P. R. Nagy and M. Kállay, *J. Chem. Phys.*, 2017, **146**, 214106.

55 P. R. Nagy, G. Samu and M. Kállay, *J. Chem. Theory Comput.*, 2018, **14**, 4193–4215.

56 P. Tecmer and K. Boguslawski, *Phys. Chem. Chem. Phys.*, 2022, **24**, 23026–23048.

57 G. Knizia, *J. Chem. Theory Comput.*, 2013, **9**, 4834–4843.

58 J. Pipek and P. G. Mezey, *J. Chem. Phys.*, 1989, **90**, 4916–4926.

59 H.-J. Werner, C. Köppl, Q. Ma and M. Schwilk, in *Fragmentation: Towards Accurate Calculations on Complex Molecular Systems*, ed. M. S. Gordon, Wiley, Chichester, UK, 2017, pp. 1–79.

60 Q. Ma and H.-J. Werner, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2018, e1371.

61 W. Klopper and C. C. M. Samson, *J. Chem. Phys.*, 2002, **116**, 6397–6410.

62 H.-J. Werner, T. B. Adler and F. R. Manby, *J. Chem. Phys.*, 2007, **126**, 164102.

63 S. Ten-no, *Chem. Phys. Lett.*, 2004, **398**, 56–61.

64 S. Ten-no, *J. Chem. Phys.*, 2004, **121**, 117–129.

65 S. Kedžuch, M. Milko and J. Noga, *Int. J. Quantum Chem.*, 2005, **105**, 929–936.

66 T. B. Adler, G. Knizia and H.-J. Werner, *J. Chem. Phys.*, 2007, **127**, 221106.

67 G. Knizia, T. B. Adler and H.-J. Werner, *J. Chem. Phys.*, 2009, **130**, 054104.

68 H.-J. Werner and A. Hansen, *J. Chem. Theory Comput.*, 2023, **19**, 7007–7030.

69 K. Peterson, C. Krause, H. Stoll, J. G. Hill and H.-J. Werner, *Mol. Phys.*, 2011, **109**, 2607–2623.

70 H.-J. Werner, G. Knizia and F. R. Manby, *Mol. Phys.*, 2011, **109**, 407–417.

71 H.-J. Werner, *J. Chem. Phys.*, 2008, **129**, 101103.
72 T. B. Adler and H.-J. Werner, *J. Chem. Phys.*, 2011, **135**, 144117.
73 T. B. Adler and H.-J. Werner, *J. Chem. Phys.*, 2011, **135**, 144117.
74 C. Krause and H.-J. Werner, *Phys. Chem. Chem. Phys.*, 2012, **14**, 7591–7604.
75 Q. Ma, M. Schwilk, C. Köppl and H.-J. Werner, *J. Chem. Theory Comput.*, 2017, **13**, 4871–4896.
76 J. G. Hill and K. A. Peterson, *J. Chem. Phys.*, 2014, **141**, 094106.
77 K. A. Peterson, T. B. Adler and H.-J. Werner, *J. Chem. Phys.*, 2008, **128**, 084102.
78 B. Metz, M. Schweizer, H. Stoll, M. Dolg and W. Liu, *Theor. Chem. Acc.*, 2000, **104**, 22.
79 B. Metz, H. Stoll and M. Dolg, *J. Chem. Phys.*, 2000, **113**, 2563.
80 K. A. Peterson, D. Figgen, E. Goll, H. Stoll and M. Dolg, *J. Chem. Phys.*, 2003, **119**, 11113.
81 K. A. Peterson and K. E. Yousaf, *J. Chem. Phys.*, 2010, **133**, 174116.
82 E. van Lenthe, E. J. Baerends and J. G. Snijders, *J. Chem. Phys.*, 1993, **99**, 4597–4610.
83 W. Kutzelnigg and W. Liu, *J. Chem. Phys.*, 2005, **123**, 241102.
84 W. Liu and D. Peng, *J. Chem. Phys.*, 2006, **125**, 044102.
85 M. Iliaš and T. Saue, *J. Chem. Phys.*, 2007, **126**, 064102.
86 D. Peng, W. Liu, Y. Xiao and L. Cheng, *J. Chem. Phys.*, 2007, **127**, 104106.
87 M. K. Armbruster, F. Weigend, C. van Wüllen and W. Klopper, *Phys. Chem. Chem. Phys.*, 2008, **10**, 1748.
88 A. Baldes and F. Weigend, *Mol. Phys.*, 2013, **111**, 2617–2624.
89 C. Holzer, Y. J. Franzke and A. Pausch, *J. Chem. Phys.*, 2022, **157**, 204102.
90 D. Peng, N. Middendorf, F. Weigend and M. Reiher, *J. Chem. Phys.*, 2013, **138**, 184105.
91 Y. J. Franzke, L. Spiske, P. Pollak and F. Weigend, *J. Chem. Theory Comput.*, 2020, **16**, 5658–5674.
92 S. Höfener, R. Ahlrichs, S. Knecht and L. Visscher, *ChemPhysChem*, 2012, **13**, 3952–3957.
93 F. Weigend, *J. Comput. Chem.*, 2008, **29**, 167.
94 F. Weigend, *Phys. Chem. Chem. Phys.*, 2002, **4**, 4285–4291.
95 K. E. Yousaf and K. A. Peterson, *J. Chem. Phys.*, 2008, **129**, 184108.
96 E. F. Valeev, *Chem. Phys. Lett.*, 2004, **395**, 190–195.
97 M. S. Marshall, L. A. Burns and C. D. Sherrill, *J. Chem. Phys.*, 2011, **135**, 194102.
98 J. M. L. Martin and S. Parthiban, *W1 and W2 theories, and their variants: thermochemistry in the kJ mol$^{-1}$ accuracy range*, Springer, Basel, 2001, pp. 31–65.
99 S. Spicher and S. Grimme, *Angew. Chem., Int. Ed.*, 2020, **59**, 15665–15673.
100 A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.
101 P. S. Brahmkshatriya, P. Dobeš, J. Fanfrlík, J. Řezáč, K. Paruch, A. Bronowska, M. Lepšík and P. Hobza, *Curr. Comput.-Aid. Drug*, 2013, **9**, 118–129.
102 J. Witte, J. B. Neaton and M. Head-Gordon, *J. Chem. Phys.*, 2017, **146**, 234105.
103 J. G. Brandenburg, C. Bannwarth, A. Hansen and S. Grimme, *J. Chem. Phys.*, 2018, **148**, 064104.
104 S. Grimme, J. G. Brandenburg, C. Bannwarth and A. Hansen, *J. Chem. Phys.*, 2015, **143**, 054107.
105 M. Müller, A. Hansen and S. Grimme, *J. Chem. Phys.*, 2023, **158**, 14103.
106 R. Sure and S. Grimme, *J. Comput. Chem.*, 2013, **34**, 1672–1685.
107 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
108 A. D. Becke, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**, 3098–3100.
109 J. P. Perdew, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1986, **33**, 8822–8824.
110 N. Mardirossian and M. Head-Gordon, *J. Chem. Phys.*, 2015, **142**, 74111.
111 J. Tao, J. P. Perdew, V. N. Staroverov and G. E. Scuseria, *Phys. Rev. Lett.*, 2003, **91**, 146401.
112 J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew and J. Sun, *J. Phys. Chem. Lett.*, 2020, **11**, 8208–8215.
113 J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew and J. Sun, *J. Phys. Chem. Lett.*, 2020, 9248.
114 S. Ehlert, U. Huniar, J. Ning, J. W. Furness, J. Sun, A. D. Kaplan, J. P. Perdew and J. G. Brandenburg, *J. Chem. Phys.*, 2021, **154**, 061101.
115 J. Ning, M. Kothakonda, J. W. Furness, A. D. Kaplan, S. Ehlert, J. G. Brandenburg, J. P. Perdew and J. Sun, *Phys. Rev. B*, 2022, **106**, 075422.
116 Y. Zhao and D. G. Truhlar, *J. Chem. Phys.*, 2006, **125**, 194101.
117 H. S. Yu, X. He and D. G. Truhlar, *J. Chem. Theory Comput.*, 2016, **12**, 1280–1293.
118 C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158–6170.
119 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5656.
120 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
121 V. N. Staroverov, G. E. Scuseria, J. Tao and J. P. Perdew, *J. Chem. Phys.*, 2003, **119**, 12129–12137.
122 M. Bursch, H. Neugebauer, S. Ehlert and S. Grimme, *J. Chem. Phys.*, 2022, **156**, 10–12.
123 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.
124 H. S. Yu, X. He, S. L. Li and D. G. Truhlar, *Chem. Sci.*, 2016, **7**, 5032–5051.
125 Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A*, 2005, **109**, 5656–5667.
126 N. Mardirossian and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2014, **16**, 9904–9924.
127 N. Mardirossian and M. Head-Gordon, *J. Chem. Phys.*, 2016, **144**, 214110.
128 G. Santra, M. Cho and J. M. Martin, *J. Phys. Chem. A*, 2021, **125**, 4614–4627.
129 L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2011, **7**, 291–309.

130 N. Mardirossian and M. Head-Gordon, *J. Chem. Phys.*, 2018, **148**, 241736.

131 J.-D. Chai and M. Head-Gordon, *J. Chem. Phys.*, 2009, **131**, 174105.

132 M. Alipour, *Chem. Phys. Lett.*, 2017, **684**, 423–426.

133 N. Mehta, M. Casanova-Páez and L. Goerigk, *Phys. Chem. Chem. Phys.*, 2018, **20**, 23175–23194.

134 F. Yu, *J. Phys. Chem. A*, 2014, **118**, 3175–3182.

135 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

136 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.

137 E. Caldeweyher, C. Bannwarth and S. Grimme, *J. Chem. Phys.*, 2017, **147**, 034112.

138 E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth and S. Grimme, *J. Chem. Phys.*, 2019, **150**, 154122.

139 E. Caldeweyher, J. M. Mewes, S. Ehlert and S. Grimme, *Phys. Chem. Chem. Phys.*, 2020, **22**, 8499–8512.

140 O. A. Vydrov and T. Van Voorhis, *Phys. Rev. Lett.*, 2009, **103**, 063004.

141 W. Hujo and S. Grimme, *J. Chem. Theory Comput.*, 2011, **7**, 3866–3871.

142 O. A. Vydrov and T. Van Voorhis, *J. Chem. Phys.*, 2010, **133**, 244103.

143 C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher and S. Grimme, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2021, **11**, e1493.

144 *AMS 2023.104*, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, 2020, **https://www.scm.com/**.

145 G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders and T. Ziegler, *J. Comput. Chem.*, 2001, **22**, 931–967.

146 DIRAC, a relativistic ab initio electronic structure program, Release DIRAC23 (2023), written by R. Bast, A. S. P. Gomes, T. Saue and L. Visscher and H. J. Aa. Jensen, with contributions from I. A. Aucar, V. Bakken, C. Chibueze, J. Creutzberg, K. G. Dyall, S. Dubillard, U. Ekström, E. Eliav, T. Enevoldsen, E. Faßhauer, T. Fleig, O. Fossgaard, L. Halbert, E. D. Hedegård, T. Helgaker, B. Helmich-Paris, J. Henriksson, M. van Horn, M. Iliaš, Ch. R. Jacob, S. Knecht, S. Komorovský, O. Kullie, J. K. Lærdahl, C. V. Larsen, Y. S. Lee, N. H. List, H. S. Nataraj, M. K. Nayak, P. Norman, A. Nyvang, G. Olejniczak, J. Olsen, J. M. H. Olsen, A. Papadopoulos, Y. C. Park, J. K. Pedersen, M. Pernpointner, J. V. Pototschnig, R. di Remigio, M. Repisky, K. Ruud, P. Sałek, B. Schimmelpfennig, B. Senjean, A. Shee, J. Sikkema, A. Sunaga, A. J. Thorvaldsen, J. Thyssen, J. van Stralen, M. L. Vidal, S. Villaume, O. Visser, T. Winther, S. Yamamoto and X. Yuan (available at, DOI: **10.5281/zenodo.7670749**), see also (**https://www.diracprogram.org**).

147 T. Saue, R. Bast, A. S. P. Gomes, H. J. A. Jensen, L. Visscher, I. A. Aucar, R. Di Remigio, K. G. Dyall, E. Eliav, E. Fasshauer, T. Fleig, L. Halbert, E. D. HedegÅrd, B. Helmich-Paris, M. Iliaš, C. R. Jacob, S. Knecht, J. K. Laerdahl, M. L. Vidal, M. K. Nayak, M. Olejniczak, J. M. H. Olsen, M. Pernpointner, B. Senjean, A. Shee, A. Sunaga and J. N. P. van Stralen, *J. Chem. Phys.*, 2020, **152**, 204104.

148 TURBOMOLE V7.6 2023, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from **https://www.turbomole.com**.

149 E. Epifanovsky, A. T. B. Gilbert, X. Feng, J. Lee, Y. Mao, N. Mardirossian, P. Pokhilko, A. F. White, M. P. Coons, A. L. Dempwolff, Z. Gan, D. Hait, P. R. Horn, L. D. Jacobson, I. Kaliman, J. Kussmann, A. W. Lange, K. U. Lao, D. S. Levine, J. Liu, S. C. McKenzie, A. F. Morrison, K. D. Nanda, F. Plasser, D. R. Rehn, M. L. Vidal, Z.-Q. You, Y. Zhu, B. Alam, B. J. Albrecht, A. Aldossary, E. Alguire, J. H. Andersen, V. Athavale, D. Barton, K. Begam, A. Behn, N. Bellonzi, Y. A. Bernard, E. J. Berquist, H. G. A. Burton, A. Carreras, K. Carter-Fenk, R. Chakraborty, A. D. Chien, K. D. Closser, V. Cofer-Shabica, S. Dasgupta, M. de Wergifosse, J. Deng, M. Diedenhofen, H. Do, S. Ehlert, P.-T. Fang, S. Fatehi, Q. Feng, T. Friedhoff, J. Gayvert, Q. Ge, G. Gidofalvi, M. Goldey, J. Gomes, C. E. González-Espinoza, S. Gulania, A. O. Gunina, M. W. D. Hanson-Heine, P. H. P. Harbach, A. Hauser, M. F. Herbst, M. Hernández Vera, M. Hodecker, Z. C. Holden, S. Houck, X. Huang, K. Hui, B. C. Huynh, M. Ivanov, Á. Jász, H. Ji, H. Jiang, B. Kaduk, S. Kähler, K. Khistyaev, J. Kim, G. Kis, P. Klunzinger, Z. Koczor-Benda, J. H. Koh, D. Kosenkov, L. Koulias, T. Kowalczyk, C. M. Krauter, K. Kue, A. Kunitsa, T. Kus, I. Ladjánszki, A. Landau, K. V. Lawler, D. Lefrancois, S. Lehtola, R. R. Li, Y.-P. Li, J. Liang, M. Liebenthal, H.-H. Lin, Y.-S. Lin, F. Liu, K.-Y. Liu, M. Loipersberger, A. Luenser, A. Manjanath, P. Manohar, E. Mansoor, S. F. Manzer, S.-P. Mao, A. V. Marenich, T. Markovich, S. Mason, S. A. Maurer, P. F. McLaughlin, M. F. S. J. Menger, J.-M. Mewes, S. A. Mewes, P. Morgante, J. W. Mullinax, K. J. Oosterbaan, G. Paran, A. C. Paul, S. K. Paul, F. Pavošević, Z. Pei, S. Prager, E. I. Proynov, Á. Rák, E. Ramos-Cordoba, B. Rana, A. E. Rask, A. Rettig, R. M. Richard, F. Rob, E. Rossomme, T. Scheele, M. Scheurer, M. Schneider, N. Sergueev, S. M. Sharada, W. Skomorowski, D. W. Small, C. J. Stein, Y.-C. Su, E. J. Sundstrom, Z. Tao, J. Thirman, G. J. Tornai, T. Tsuchimochi, N. M. Tubman, S. P. Veccham, O. Vydrov, J. Wenzel, J. Witte, A. Yamada, K. Yao, S. Yeganeh, S. R. Yost, A. Zech, I. Y. Zhang, X. Zhang, Y. Zhang, D. Zuev, A. Aspuru-Guzik, A. T. Bell, N. A. Besley, K. B. Bravaya, B. R. Brooks, D. Casanova, J.-D. Chai, S. Coriani, C. J. Cramer, G. Cserey, A. E. DePrince, R. A. DiStasio, A. Dreuw, B. D. Dunietz, T. R. Furlani, W. A. Goddard, S. Hammes-Schiffer, T. Head-Gordon, W. J. Hehre, C.-P. Hsu, T.-C. Jagau, Y. Jung, A. Klamt, J. Kong, D. S. Lambrecht, W. Liang, N. J. Mayhall, C. W. McCurdy, J. B. Neaton, C. Ochsenfeld,

J. A. Parkhill, R. Peverati, V. A. Rassolov, Y. Shao, L. V. Slipchenko, T. Stauch, R. P. Steele, J. E. Subotnik, A. J. W. Thom, A. Tkatchenko, D. G. Truhlar, T. Van Voorhis, T. A. Wesolowski, K. B. Whaley, H. L. Woodcock, P. M. Zimmerman, S. Faraji, P. M. W. Gill, M. Head-Gordon, J. M. Herbert and A. I. Krylov, *J. Chem. Phys.*, 2021, **155**, 084801.

150 J. P. James, *MOPAC2016*, Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, 2016, **https://OpenMOPAC.net**.

151 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.

152 J. J. Determan and A. K. Wilson, *J. Chem. Phys.*, 2024, **160**, 084105.

153 F. Neese, F. Wennmohs, A. Hansen and U. Becker, *Chem. Phys.*, 2009, **356**, 98–109.

154 F. Weigend, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1057–1065.

155 S. Grimme, A. Hansen, J. G. Brandenburg and C. Bannwarth, *Chem. Rev.*, 2016, **116**, 5105–5154.

156 V. M. Miriyala and J. Řezáč, *J. Phys. Chem. A*, 2018, **122**, 2801–2808.

157 K. Kříž, M. Nováček and J. Řezáč, *J. Chem. Theory Comput.*, 2021, **17**, 1548–1561.

158 K. Kříž and J. Řezáč, *Phys. Chem. Chem. Phys.*, 2022, **24**, 14794–14804.

159 P. Pracht, F. Bohle and S. Grimme, *Phys. Chem. Chem. Phys.*, 2020, **22**, 7169–7192.

160 S. Dohm, M. Bursch, A. Hansen and S. Grimme, *J. Chem. Theory Comput.*, 2020, **16**, 2002–2012.

161 T. Husch, A. C. Vaucher and M. Reiher, *Int. J. Quantum Chem.*, 2018, **118**, e25799.

162 A. Ricchebuono, E. Vottero, A. Piovano, E. Groppo, P. Raybaud and C. Chizallet, *J. Phys. Chem. C*, 2023, **127**, 18454–18465.

PCCP

CORRECTION

View Article Online
View Journal | View Issue

Check for updates

Cite this: Phys. Chem. Chem. Phys., 2025, 27, 8572

# Correction: The p-block challenge: assessing quantum chemistry methods for inorganic heterocycle dimerizations

Thomas Gasevic,[c] Markus Bursch,*[ad] Qianli Ma,[b] Stefan Grimme,[c]
Hans-Joachim Werner*[b] and Andreas Hansen*[c]

DOI: 10.1039/d5cp90062j

rsc.li/pccp

Correction for 'The p-block challenge: assessing quantum chemistry methods for inorganic heterocycle dimerizations' by Thomas Gasevic *et al.*, *Phys. Chem. Chem. Phys.*, 2024, **26**, 13884–13908, **https://doi.org/10.1039/D3CP06217A**.

The root mean square deviations (RMSDs) reported in Table 4 of the original publication are incorrect for revDSD-PBEP86-D3(BJ) and B2NC-PLYP due to faulty inputs in the calculations. The correct values are now provided here in Table 1. Further, PM6-D3H4X was used instead of PM6-D3H4.

In our manuscript, we used the final single point energies of revDSD-PBEP86-D4(2021) and replaced the D4 London dispersion correction with the D3(BJ) correction of the original revDSD-PBEP86-D3(BJ) publication. However, as the density functional itself was also re-parameterized in 2021, this approach is not valid.[1,2] The correct parameters for revDSD-PBEP86-D3(BJ) yield slightly larger deviations ($\text{RMSD}_{\text{cov}}$ = 7.5 *vs.* 8.0 kcal $\text{mol}^{-1}$).

In the calculations with B2NC-PLYP, Hartree Fock was applied instead of Density Functional Theory (DFT). Using the correct settings significantly reduces the errors for this test set, making B2NC-PLYP a viable choice for the computation of inorganic heterocycle dimerizations ($\text{RMSD}_{\text{cov}}$ = 3.4 kcal $\text{mol}^{-1}$; $\text{RMSD}_{\text{wda}}$ = 0.8 kcal $\text{mol}^{-1}$).

The Royal Society of Chemistry apologises for these errors and any consequent inconvenience to authors and readers.

---

[a] Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany. E-mail: bursch@kofo.mpg.de
[b] Institut für Theoretische Chemie, Universität Stuttgart, Pfaffenwaldring 55, D-70569 Stuttgart, Germany. E-mail: werner@theochem.uni-stuttgart.de
[c] Mulliken Center for Theoretical Chemistry, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstr. 4, 53115 Bonn, Germany. E-mail: hansen@thch.uni-bonn.de
[d] FACCTs GmbH, 50677, Koeln, Germany

**Table 1** Root mean square deviations (RMSDs) for all tested methods in kcal mol$^{-1}$. Except for the FF, SQM, and DFT composite methods, the def2-QZVPP basis set was applied. The values of further statistical descriptors are given in the ESI

| | | RMSD/kcal mol$^{-1}$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Covalent | | | | Weak donor–acceptor | | | |
| Class | Method | Plain | D3 | D4 | NL | Plain | D3 | D4 | NL |
| FF | GFN-FF[3] | | | 184.3 | | | | 9.1 | |
| | UFF[4] | 1416.3 | | | | 1.0 | | | |
| SQM | PM6-D3H4X[5,6] [a] | | 158.8 | | | | 33.2 | | |
| | PM7[7] [a] | | 162.7 | | | | 17.9 | | |
| | GFN0-xTB | | | 52.0 | | | | 6.9 | |
| | GFN1-xTB[8] | | 46.7 | | | | 7.6 | | |
| | GFN2-xTB[9] | | | 36.8 | | | | 6.7 | |
| Composite | B97M-V-C[10] | | | | 12.7 | | | | 4.1 |
| | B97-3c[11] | | 9.3 | | | | 6.8 | | |
| | r$^2$SCAN-3c[12] | | | 5.7 | | | | 1.5 | |
| | PBEh-3c[13] | | 10.0 | | | | 1.8 | | |
| | ωB97X-3c[14] | | | 8.0 | | | | 1.1 | |
| | HF-3c[15] | | 33.6 | | | | 9.0 | | |
| (meta-)GGA | PBE[16] | 16.3 | 7.3 | 7.4 | | 8.3 | 1.0 | 2.2 | |
| | BP86[17,18] | 23.3 | 11.6 | 8.6 | | 10.9 | 10.3 | 7.7 | |
| | B97M[19] | | | | 8.8 | | | | 2.2 |
| | TPSS[20] | 14.9 | 7.0 | 9.6 | | 9.5 | 3.9 | 4.7 | |
| | r$^2$SCAN[21–24] | 8.1 | 4.3 | 4.8 | | 4.7 | 1.0 | 1.5 | |
| | M06-L[25] | 12.9 | 12.7 | 11.7 | | 1.8 | 1.7 | 1.6 | |
| | MN15-L[26] | 14.4 | 14.4 | | | 3.2 | 3.2 | | |
| Hybrid | PBE0[27] | 7.4 | 9.1 | 9.7 | 9.2 | 7.2 | 2.9 | 3.2 | 0.6 |
| | B3LYP[28,29] | 33.5 | 8.2 | 9.6 | | 12.9 | 4.2 | 3.9 | |
| | TPSSh[30] | 11.4 | 8.7 | 9.6 | | 9.0 | 4.8 | 4.6 | |
| | r$^2$SCAN0[31] | 5.0 | 5.6 | 5.1 | 6.2 | 4.5 | 1.3 | 1.0 | 1.0 |
| | M06[32] | 7.7 | 6.8 | 5.8 | | 1.2 | 1.6 | 2.2 | |
| | M06-2X[32] | 6.9 | 6.7 | | | 1.7 | 2.0 | | |
| | MN15[33] | 12.9 | 12.9 | | | 2.2 | 2.2 | | |
| | PW6B95[34] | 9.8 | 5.8 | 8.0 | | 6.1 | 3.4 | 3.6 | |
| | ωB97X[35] | | | | 8.7 | | | | 1.1 |
| | ωB97M[36] | | | | 4.8 | | | | 2.1 |
| Double-hybrid | revDSD-PBEP86-D3(BJ)[1] | | 8.0 | | | | 3.7 | | |
| | revDSD-PBEP86-D4(2021)[2] | 4.8 | | 2.8 | | 2.4 | | 0.8 | |
| | PWPB95[37] | 3.8 | 13.1 | 3.6 | | 3.4 | 4.8 | 1.7 | |
| | ωB97M(2)[38] | | | | 9.7 | | | | 3.0 |
| | ωB97X-2[39] | | | | 7.0 | | | | 2.1 |
| | Pr$^2$SCAN50[40] | | | 5.0 | 6.7 | | | 1.1 | 1.0 |
| | κPr$^2$SCAN50[40] | | | 7.2 | | | | 1.0 | |
| | ωPr$^2$SCAN50[40] | | | 7.1 | 6.9 | | | 1.6 | 0.5 |
| | SOS0-PBE0-2[41,42] | 7.6 | 11.2 | | | 1.8 | 1.7 | | |
| | B2NC-PLYP[42,43] | 3.4 | 4.5 | | | 0.8 | 1.7 | | |

[a] As the PM*x* methods are not parameterized for Po, no data for Po-containing systems are included.

## Notes and references

1 G. Santra, N. Sylvetsky and J. M. L. Martin, *J. Phys. Chem. A*, 2019, **123**, 5129–5143.

2 G. Santra, M. Cho and J. M. Martin, *J. Phys. Chem. A*, 2021, **125**, 4614–4627.

3 S. Spicher and S. Grimme, *Angew. Chem., Int. Ed.*, 2020, **59**, 15665–15673.

4 A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.

5 J. J. P. Stewart, *J. Mol. Model.*, 2007, **13**, 1173.

6 P. S. Brahmkshatriya, P. Dobeš, J. Fanfrlík, J. Řezáč, K. Paruch, A. Bronowska, M. Lepšík and P. Hobza, *Curr. Comput.-Aid. Drug.*, 2013, **9**, 118–129.

7 J. J. P. Stewart, *J. Mol. Model.*, 2013, **19**, 1–32.

8 S. Grimme, C. Bannwarth and P. Shushkov, *J. Chem. Theory Comput.*, 2017, **13**, 1989–2009.

9 C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652–1671.

10 J. Witte, J. B. Neaton and M. Head-Gordon, *J. Chem. Phys.*, 2017, **146**, 234105.

11 J. G. Brandenburg, C. Bannwarth, A. Hansen and S. Grimme, *J. Chem. Phys.*, 2018, **148**, 064104.

12 S. Grimme, A. Hansen, S. Ehlert and J. M. Mewes, *J. Chem. Phys.*, 2021, **154**, 064103.

13 S. Grimme, J. G. Brandenburg, C. Bannwarth and A. Hansen, *J. Chem. Phys.*, 2015, **143**, 054107.

14 M. Müller, A. Hansen and S. Grimme, *J. Chem. Phys.*, 2023, **158**, 14103.

15 R. Sure and S. Grimme, *J. Comput. Chem.*, 2013, **34**, 1672–1685.
16 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
17 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.
18 J. P. Perdew, *Phys. Rev. B*, 1986, **33**, 8822–8824.
19 N. Mardirossian and M. Head-Gordon, *J. Chem. Phys.*, 2015, **142**, 74111.
20 J. Tao, J. P. Perdew, V. N. Staroverov and G. E. Scuseria, *Phys. Rev. Lett.*, 2003, **91**, 146401.
21 J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew and J. Sun, *J. Phys. Chem. Lett.*, 2020, **11**, 8208–8215.
22 J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew and J. Sun, *J. Phys. Chem. Lett.*, 2020, 9248.
23 S. Ehlert, U. Huniar, J. Ning, J. W. Furness, J. Sun, A. D. Kaplan, J. P. Perdew and J. G. Brandenburg, *J. Chem. Phys.*, 2021, **154**, 061101.
24 J. Ning, M. Kothakonda, J. W. Furness, A. D. Kaplan, S. Ehlert, J. G. Brandenburg, J. P. Perdew and J. Sun, *Phys. Rev. B*, 2022, **106**, 075422.
25 Y. Zhao and D. G. Truhlar, *J. Chem. Phys.*, 2006, **125**, 194101.
26 H. S. Yu, X. He and D. G. Truhlar, *J. Chem. Theory Comput.*, 2016, **12**, 1280–1293.
27 C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158–6170.
28 Axel D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5656.
29 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
30 V. N. Staroverov, G. E. Scuseria, J. Tao and J. P. Perdew, *J. Chem. Phys.*, 2003, **119**, 12129–12137.
31 M. Bursch, H. Neugebauer, S. Ehlert and S. Grimme, *J. Chem. Phys.*, 2022, **156**, 10–12.
32 Y. Zhao and D. G. Truhlar, *Theory Chem. Acc.*, 2008, **120**, 215–241.
33 H. S. Yu, X. He, S. L. Li and D. G. Truhlar, *Chem. Sci.*, 2016, **7**, 5032–5051.
34 Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A*, 2005, **109**, 5656–5667.
35 N. Mardirossian and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2014, **16**, 9904–9924.
36 N. Mardirossian and M. Head-Gordon, *J. Chem. Phys.*, 2016, **144**, 214110.
37 L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2011, 7, 291–309.
38 N. Mardirossian and M. Head-Gordon, *J. Chem. Phys.*, 2018, **148**, 241736.
39 J.-D. Chai and M. Head-Gordon, *J. Chem. Phys.*, 2009, **131**, 174105.
40 L. Wittmann, H. Neugebauer, S. Grimme and M. Bursch, *J. Chem. Phys.*, 2023, **159**, 224103.
41 M. Alipour, *Chem. Phys. Lett.*, 2017, **684**, 423–426.
42 N. Mehta, M. Casanova-Páez and L. Goerigk, *Phys. Chem. Chem. Phys.*, 2018, **20**, 23175–23194.
43 F. Yu, *J. Phys. Chem. A*, 2014, **118**, 3175–3182.

# APPENDIX D

# Chemical Space Exploration with Artificial "Mindless" Molecules

Thomas Gasevic,[†] Marcel Müller,[†] Jonathan Schöps,[†] Stephanie Lanius,[‡], Jan Hermann,[‡] Stefan Grimme,[†] and Andreas Hansen[†]

*Received: June 16, 2025*
*First published: September 2, 2025*

Reprinted in (adapted) with permission from:
T. Gasevic, M. Müller, J. Schöps, S. Lanius, J. Hermann, S. Grimme, and A. Hansen, *Chemical Space Exploration with Artificial "Mindless" Molecules*, J. Chem. Inf. Model. **65**.18 (2025) 9576, DOI: `10.1021/acs.jcim.5c01364`

**Own contributions**

- Conceptualization
- Formal Analysis
- Investigation – Generating the MB2061 set & computing all results, excluding the reference energies
- Visualization
- Writing – Original Draft Preparation, Review & Editing

---

[†] Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany

[‡] Microsoft Research AI for Science, Microsoft Research, D-10117 Berlin, Germany

JCIM JOURNAL OF
CHEMICAL INFORMATION
AND MODELING

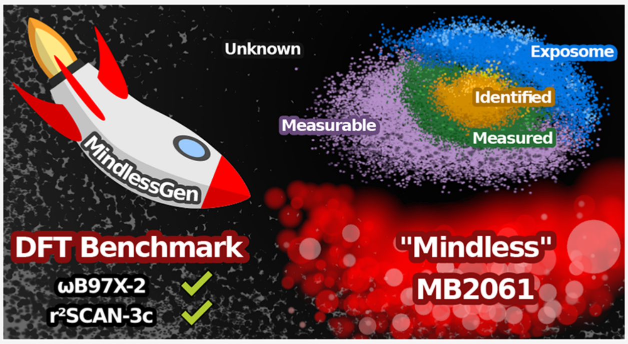
# Chemical Space Exploration with Artificial "Mindless" Molecules

*Published as part of Journal of Chemical Information and Modeling special issue "Chemical Compound Space Exploration by Multiscale High-Throughput Screening and Machine Learning".*

Thomas Gasevic, Marcel Müller, Jonathan Schöps, Stephanie Lanius, Jan Hermann, Stefan Grimme, and Andreas Hansen*

**Cite This:** *J. Chem. Inf. Model.* 2025, 65, 9576−9587 **Read Online**

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information



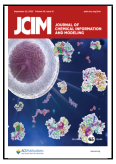
**ABSTRACT:** We introduce MindlessGen, a Python-based generator for creating chemically diverse, "mindless" molecules through random atomic placement and subsequent geometry optimization. Using this framework, we constructed the *MB2061* benchmark set, containing 2061 molecules with high-level PNO-LCCSD(T)-F12 reference data for $H_2$-promoted decomposition reactions. This set provides a challenging benchmark for testing, validating, and training density functional approximations (DFAs), semiempirical methods, force fields, and machine learning potentials using molecular structures beyond conventional chemical space. For DFAs, we initially hypothesized that highly parametrized functionals might perform poorly on this set. However, no consistent relationship between the fitting strategy and accuracy was observed. A clear Jacob's ladder trend emerges, with *ω*B97X-2 achieving the lowest mean absolute error (MAE) of 8.4 kcal·mol$^{-1}$ and r$^2$SCAN-3c offering a robust cost-efficient alternative (19.6 kcal·mol$^{-1}$). Furthermore, we discuss the performance of selected semiempirical methods and contemporary machine-learning interatomic potentials.

## INTRODUCTION

Exploring the vast, multidimensional landscape of chemical structures has become a highly desirable pursuit in modern chemistry.[1−3] Samanipour et al. divide this landscape into five main categories of chemical space: the identified, the measured, the measurable, the exposome, and the unknown.[4] Identifying novel compounds with suitable properties in the uncharted regions is required in fields such as de novo drug design, materials science, or catalysis, and can lead to great scientific and technological advancements.[5−11] However, the number of structures is nearly infinite and therefore cannot be assessed solely experimentally. To overcome this limitation, computational approaches have become indispensable. For example, cheminformatics methods are routinely used for high-throughput screening (HTS) or predicting molecular properties.[12−14] They typically rely on large databases and often apply machine learning (ML) models to extrapolate existing knowledge to new compounds. This works well for most synthetically accessible molecules, but their predictive power is inherently limited to the quality of data they are trained on.[15−17]

A significant aspect of chemical space exploration involves investigating structures in the unknown realm that are theoretically feasible but synthetically inaccessible by current techniques. In such cases, quantum chemical (QC) calculations provide a more fundamental approach to exploring uncharted regions of the chemical space. Specifically, density functional theory (DFT) has proven to yield accurate results at a moderate computational cost.[18−20] Nevertheless, given the vast number of available DFAs, selecting the most appropriate one for a specific application is far from straightforward. Each density functional is constructed differently to approximate the exact functional, often incorporating varying degrees of empirical fitting. As Medvedev et al. point out, the extent of empiricism typically influences both the generalizability and the performance of a functional.[21−23] For instance, the so-called "exact constraints," which are mathematical properties that the exact functional should satisfy, may be approximated to differing degrees or even explicitly violated. In general, achieving accuracy in both electron densities and total energies is essential. Still, most modern functionals are primarily fitted to reproduce only the energies of bonded systems, often at the expense of accurately modeling electron densities. Benchmark studies against high-

**Received:** June 16, 2025
**Revised:** August 19, 2025
**Accepted:** August 20, 2025
**Published:** September 2, 2025



https://doi.org/10.1021/acs.jcim.5c01364
*J. Chem. Inf. Model.* 2025, 65, 9576−9587

level wave function-based methods are therefore crucial for assessing and refining QC approaches such as DFT.[24] Yet, most existing and commonly applied data sets focus on conventional, synthetically accessible molecules, limiting their applicability to more "exotic" chemical systems.[25,26] Commonly employed quantum mechanical (QM) data sets[27] for machine learning and benchmarking studies include QM7,[28] QM7b,[29] QM7-X,[30] QM9,[31] ANI-1,[32] ANI-1x,[33] PubChemQC,[34] ZINC20,[35] GEOM,[36] OMol25,[37] and GMTKN55,[38] just to name a few. Several of these QM data sets were derived from widely used chemical databases such as PubChem,[25] ChEMBL,[26] and GDB.[39,40] Research into unconventional molecular structures has already been explored through data sets such as the "mindless" benchmarking set MB08-165,[41] its successor MB16-43,[38] the VQM24 data set by von Lilienfeld et al.,[42] as well as the QCML data set by Google DeepMind.[43] However, these sets are limited either in size or by their elemental composition. For instance, MB16-43 includes only 16-atom molecules composed of elements from hydrogen to chlorine, explicitly excluding noble gases. Similarly, the VQM24 data set comprises molecules with up to five heavy atoms selected from the elements C, N, O, F, Si, P, S, Cl, and Br, which are then saturated with hydrogen to yield neutral, closed-shell species.

In this work, we explore undiscovered chemical space by generating "mindless" molecules (MLMs)[41] in their relaxed ground-state geometries. The generated molecules consist of up to 20 randomly placed atoms with diverse binding motifs. Reaction energies for the decomposition of MLMs into hydrides and diatomic molecules serve here as a reliable benchmark for assessing the accuracy and the "robustness" of DFAs.[41] MLMs are particularly valuable for methods that rely heavily on the chemical diversity and reliability of their training data, such as contemporary ML models. By incorporating a wide range of binding motifs within a single molecule, they hypothetically provide a richer source of information than conventional training data sets. This enhanced chemical diversity should improve the generalizability of ML-based approaches while reducing the required number of training samples.[44] With fewer data points needed, more computationally demanding reference calculations with higher accuracy can be incorporated into the data generation process, enhancing the data's reliability. By additionally integrating already established lower-accuracy data sets within multifidelity approaches for training, predictive capabilities can be further improved. This is also valuable in the development of new QC codes that are still not optimized and, therefore, are limited to very slow computations, and as a consequence, rely on tests of small molecules.[45]

To assess the performance of computational methods and to reach an uncharted chemical space, specialized generative strategies are required. Commonly, ML approaches are employed to generate new molecules by extrapolating from known structures.[46−48] Here, we introduce a Python code for the efficient generation of MLMs and provide a collection of 2061 optimized structures with high-level PNO-LCCSD(T)-F12 reaction energies for reference (vide infra). We present the generation process of MLMs, introduce our benchmark set, and assess the performance of semiempirical quantum mechanical (SQM), machine learning interatomic potential (MLIP), and DFT methods. In this context, we also investigate the influence of relativistic effects. Through our DFT study, we aim to clarify how different fitting strategies and density functional approximation (DFA) parametrization schemes impact the accuracy and robustness of the computation of MLMs.

## ■ MINDLESS MOLECULE GENERATOR

The first attempts to use "mindless" molecules (MLMs) for validating DFAs beyond conventional chemical space in our lab relied on a semiautomatic Fortran program that randomly placed atoms in space. Although basic geometric constraints, such as avoiding close contacts, were applied, the program still required extensive manual intervention for subsequent optimization, filtering failed or poor-quality structures, and adjusting the elemental compositions. As part of this work and for generating validation data for a novel semiempirical tight-binding (TB) method, we developed a new open-source, Python-based implementation of a fully automated "mindless" molecule generator: MindlessGen. Its key features are summarized in Figure 1 and will be described in the following.

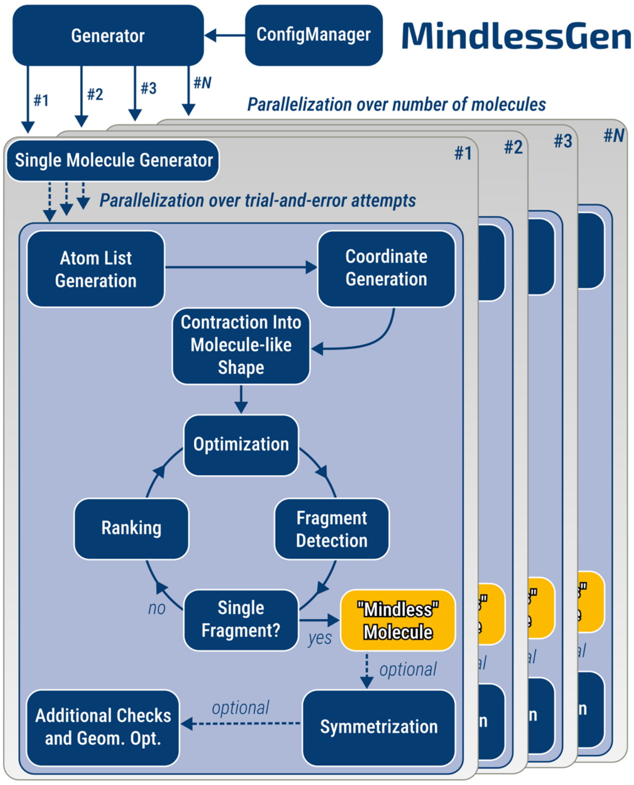


**Figure 1.** Outline of the working principle of MindlessGen.

**MindlessGen: How It Works.** The ConfigManager class stores all settings relevant to the execution of MindlessGen. Configurations can be specified in three ways: via a mindlessgen.toml file, through command-line arguments, or directly using the Python API. Users can define key parameters such as the number and size of molecules, molecular charge, symmetry constraints (see below), desired atom types and their quantities, and excluded elements. MindlessGen supports all elements with atomic numbers $Z = 1-103$, including lanthanides and actinides. For actinides, which are not supported by GFN$n$-xTB,[49−52] the atom types are temporarily substituted with the corresponding lanthanide analogs during optimization and subsequently reverted. Additional technical settings include the number of available CPU cores, configuration options for the interfaced (S)QM programs, and thresholds for molecule rejection during sanity checks. An overview of all configurable options can be found in the example mindlessgen.toml file available on the

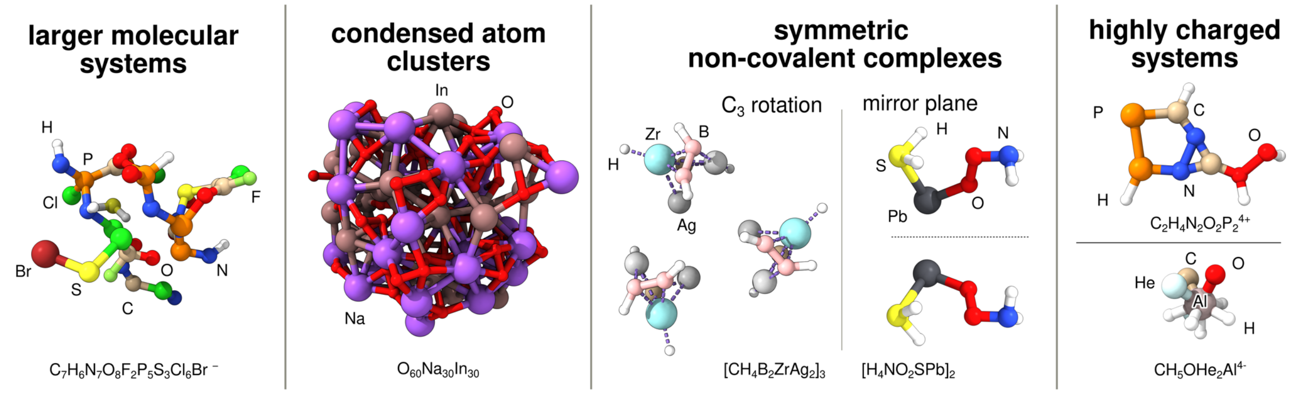


**Figure 2.** Exemplary "mindless" molecules generated using MindlessGen, clustered into different categories. The corresponding mindlessgen.toml input files used to generate these structures are available in the code repository at github.com/grimme-lab/MindlessGen.

GitHub repository (https://github.com/grimme-lab/MindlessGen).

The molecule generation proceeds in several steps: First, a randomized list of atoms adhering to the given constraints is generated, and initial atomic coordinates are assigned. Unless explicitly defined otherwise, the atom list and molecular charge are chosen to yield closed-shell species. An exception is made for lanthanides and actinides, which are assumed to adopt an oxidation state (III) and high-spin configuration. The atomic coordinates are then contracted into a molecule-like shape using only geometric rules before undergoing optimization, preferably using a fast SQM method such as GFN*n*-xTB. During optimization, the structure may have broken into multiple disconnected fragments. Using graph theory, the program identifies and ranks fragments by size, selecting the largest fragment for continued optimization. Graph nodes, reflecting covalent bonds, are set up by comparing interatomic distances with covalent radii by Pyykkö and Atsumi.[53] This cycle is repeated until a single connected molecule remains. As a basic test for multireference (MR) character, the HOMO−LUMO gap is compared against a user-defined threshold following optimization. Further optional postprocessing steps may then be applied:

- Generation of symmetric arrangements for supramolecular complexes based on noncovalent interactions (NCIs).
- Use of alternative (DFT) methods from various external programs for additional sanity checks or geometry optimizations.

Since the initial coordinate generation often leads to structures that cannot be tackled reliably with standard SQM or DFT methods, self-consistent field (SCF) procedures and geometry optimizations may fail to converge. Consequently, the structure suggestion, fragmentation handling, and optional postprocessing are conducted in a trial-and-error fashion. Depending on the specified composition, size, and charge, the number of cycles required to obtain a valid structure can range from $10^1$ to $10^3$. Upon success, the generated MLMs are returned, optionally also in .xyz format. MindlessGen offers a modular and extensible architecture with the following key features:

- Object-oriented design enabling intuitive access to all major modules.
- Comprehensive unit and validation testing for robust performance.
- Easy installation via PyPI (https://pypi.org/project/mindlessgen/) or from source via GitHub (https://github.com/grimme-lab/MindlessGen).
- Dynamic parallelization over the number of molecules and the trial cycles per molecule, ensuring high efficiency even for large-scale "mindless" molecule (MLM) generation.
- Interfaces to various (S)QM programs, including xtb-6.7.1,[50] TURBOMOLE-7.9.0,[54] and ORCA-6.0.1,[55] with support for easy extension through a standardized backend interface.

**Exemplary "Mindless" Molecules.** Figure 2 showcases exemplary "mindless" molecules (MLMs) generated by using MindlessGen. In addition to medium- to large-sized organic molecules, the generator also supports the creation of large atomic clusters such as $O_{60}Na_{30}In_{30}$, which may serve as model systems for sodium-ion battery solid electrolytes[56] or amorphous indium oxide semiconductors.[57] The option to generate symmetric copies ($C_s$ mirror planes, $C_i$ inversion centers, or $C_n$ rotational symmetry) enables the automated construction of NCI complexes from MLMs. When quantum chemistry packages able to exploit molecular symmetry are employed in the postprocessing step, these symmetry constraints can be utilized for more efficient geometry optimizations. Moreover, the $C_3$-symmetric example $[CH_4B_2ZrAg_2]_3$ highlights that MLMs are not restricted to main-group elements but may also include transition metals or *f*-block elements. Constraining the molecular charge to extreme values allows for the generation of highly charged species, which are particularly useful for identifying edge cases in the development of SQM methods, charge models, or machine learning force fields (MLFFs).

## ■ MOLECULE GENERATION FOR THE BENCHMARK SET

After the generation of the candidate molecules for our benchmark using MindlessGen, a few key steps were taken to ensure a representative and reliable structure collection. Although MindlessGen contains simple checks for multireference (MR) character, random molecule generation carries the risk of producing systems with such features. Molecules exhibiting pronounced MR character are difficult to describe accurately using commonly applied methods such as density functional or coupled-cluster theory and would require individual analysis for the generation of reliable reference data.

To mitigate this, transition metals and open-shell systems were excluded from the data set. Also, noble gases were not included, since they are chemically too inert. Total charges between −2 and +2 were randomly selected to ensure closed-shell systems. Additionally, various MR diagnostics were applied, including fractional occupation density (FOD)-based static correlation diagnostics[58−60] or the SCF stability analysis[61,62] (see the Supporting Information for further details).[63] While the FOD quantifies static electron correlation effects through fractional orbital occupations, the SCF stability analysis examines the electronic Hessian with respect to orbital rotations, determining whether the obtained solution represents a true energy minimum or if alternative solutions exist, making both methods valuable MR diagnostics. During inspection of the generated MLMs, we encountered further challenges concerning our design philosophy that need to be addressed: Group 13 elements, especially B and Al, tend to form minimum structures more easily and therefore could dominate the data set if this is not restricted. Moreover, many generated MLMs also contain linear chains that offer little structural diversity. Some structures that were not sorted out by the graph algorithm also resemble conventional molecules too closely or contain small non-covalently bound molecular fragments such as $H_2O$, $CO^-$, HCl, HBr, or others. This is especially problematic in the context of the decomposition reactions that are investigated in the original MB08-165 and MB16-43 benchmark sets:

$$\begin{aligned}&\square C_{\square}H_{\square}B_{\square}N_{\square}\ldots + \square H_2\\&\quad\rightarrow \square CH_4 + \square BH_3 + \square N_2 + \ldots\end{aligned}\tag{1}$$

If multiple molecular fragments are present within a single structure, then such transformations no longer represent proper decomposition reactions and would bias the benchmark statistics.

In the new *MB2061* set, introduced in this work, each MLM is decomposed by adding $H_2$, forming either hydrides or homonuclear diatomics for group 15 and 17 elements ($X_2$). To balance the charges, charged hydride species were generated by protonating or deprotonating the neutral products. The charged species include $BH_2^-$, $CH_3^-$, $NH_2^-$, $OH^-$, $PH_2^-$, $SH^-$, $SiH_3^-$, $CH_3^+$, $NH_4^+$, $H_3O^+$, $PH_4^+$, $SH_3^+$, and $SiH_3^+$. We chose this decomposition reaction because we believe it better reflects everyday chemistry, unlike conventional atomization energies, which are biased toward open-shell atoms due to their reliance on isolated atomic reference states. Nevertheless, it is important to keep in mind that these reactions may involve large stoichiometric coefficients for the products, which can amplify any associated errors. Further, there is a bias toward hydride species, and while conventional molecules are used as products, "mindless" molecules are only applied as reactants in the investigated reactions. Notably, our intention is not to diminish the importance of atomization energy data sets. On the contrary, we want to emphasize that accurately representing the field of thermochemistry requires the incorporation of a diverse range of molecular properties during method development.

As MindlessGen was still under active development during the generation of our benchmark set, we evaluated how many structures needed to be discarded to obtain a high-quality final set with the current MindlessGen-0.6.0. For this purpose, we generated 20 molecules for each system size between 10 and 20 atoms, yielding a total of 220 molecules optimized with the semiempirical GFN2-xTB tight-binding method. These were then filtered using all necessary checks (see Figure 3).

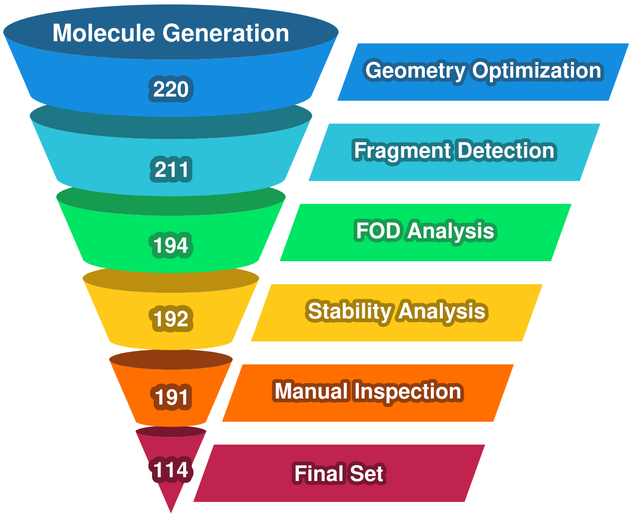


**Figure 3.** Exemplary screening process of MLMs after molecule generation. The respective volume elements of the funnel do not represent the actual set size. See the text for more details.

Final geometry optimizations on the DFT level (PBE0-D4/def2-TZVP)[64−66] failed to converge for 9 systems, and 17 structures dissociated during this geometry optimization. Two more structures were excluded due to the FOD analysis, and one more was excluded due to the SCF stability analysis. A final manual inspection led to the removal of 77 further structures. In total, approximately half of the generated geometries were discarded. Even after applying these filter steps, the final set should still be reviewed for elemental composition and bonding motifs to ensure sufficient chemical diversity across the data set. However, we note that some of these filtering steps could also be included in the fully automated process.

## ■ *MB2061* BENCHMARK SET

To evaluate various quantum-chemical methods for the accurate description of MLMs, we developed the *MB2061* "mindless" benchmark set using high-quality reference data. All structures for the benchmark set were generated with a development version of MindlessGen and contain all main-group elements up to iodine, excluding noble gases (Figure 4). Duplicates can be ruled out due to unique elemental compositions. The total set comprises 2061 MLMs in their singlet ground state geometries.

All structures were screened according to our workflow described in the previous section. Potential MR cases were excluded based on conclusive diagnostics (see Supporting Information). Final geometries were optimized at the PBE0-D4/def2-TZVP level. Final reaction energies were computed for the decomposition of MLMs into hydrides and diatomic molecules (eq 1) with the high-level local correlation method PNO-LCCSD(T)-F12[67−73] (as implemented in Molpro 2025.1)[71,74,75] applied with TIGHT domain settings, sub-valence correlation, and large basis sets (see Supporting Information for details). This approach has been extensively validated and proven reliable in our theoretical study on dimerizations of cyclic molecules with *p*-block elements[76] and several further studies.[77−80] The reaction energies range from −1232.4 to 1607.8 kcal·mol$^{-1}$ and have a mean absolute value of 306.3 kcal·mol$^{-1}$ (Figure 5). The estimated residual error of the reference values is—based on the experience of previous studies—below ± 0.5 and 1 kcal·mol$^{-1}$ for the light and heavy

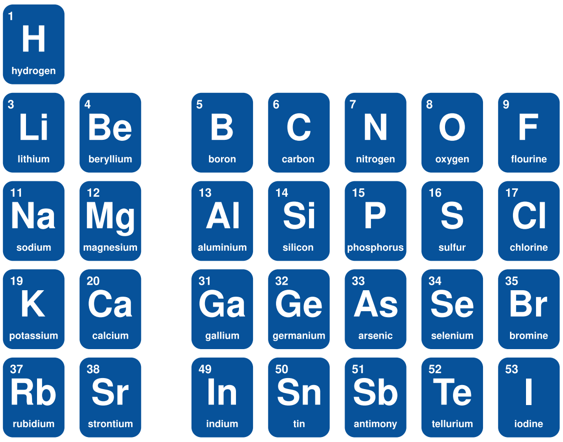


**Figure 4.** All elements included in the *MB2061* benchmark set.

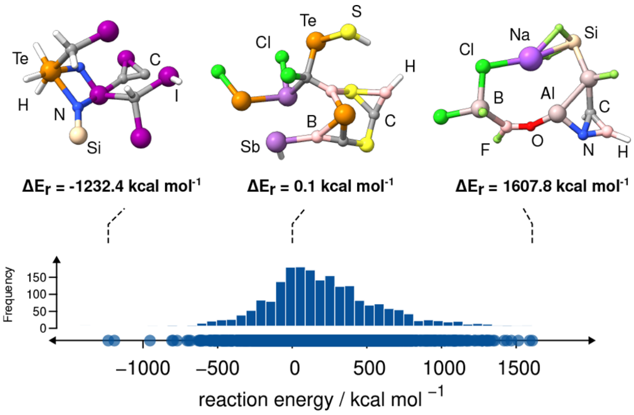


**Figure 5.** Range of reference reaction energies ($\Delta E_r$) in the *MB2061* set with three example MLMs.

*p*-block reactions, respectively, and may be a little bit larger for the heavier *s*-block reactions due to a slightly less accurate basis set setup (see the Supporting Information).

As the already established MB16-43 "mindless" benchmark set only contains neutral MLMs with exactly 16 atoms, we chose to introduce significantly greater structural and compositional diversity in terms of system size and element choice in our data set. The new *MB2061* set is designed to be chemically diverse by incorporating various system charges, sizes, elements, binding partners, and binding motifs (cf. Figure 6). Notably, this set is not intended to serve as a comprehensive QM data set covering broad regions of chemical space with millions of data points at a low to moderate level of theory. Instead, it is designed as a benchmark set featuring accurate reference energies for chemically unconventional structures.

As each MLM inherently hypothetically contains substantial chemical information due to its diverse bonding patterns and environments, very large molecular systems are not required, and hence we limited the system size to 10−20 atoms. A few smaller molecules with 8 or 9 atoms remained in the set from initial tests, as we had already generated high-quality reference data for them. The system sizes are almost evenly distributed, but smaller molecules are slightly overrepresented.

Most structures contain at least one main-group element of the first three rows of the periodic table, with C (74%), N (56%), and O (54%) being the most prevalent. As organic and other conventional molecules usually consist of these elements, this set is also of high interest for the general assessment of QM methods. The remaining elements occur in 4−9% of all structures. The entire set can be divided into individual subsets if only a specific group of elements is desired. Since benchmark data sets with *s*-block and heavy *p*-block elements are relatively rare, and heavier atoms may require an individual assessment of relativistic effects, we created three subsets: *MB727-Light* with all main-group elements from H—Cl excluding group 1 and 2 elements, *MB782-s-Block* with all main-group elements of H—Cl plus K, Ca, Rb, and Sr, and *MB552-Heavy* with all main-group elements from H—I excluding group 1 and 2 elements. The sets include 727, 782, and 552 MLMs, respectively, and, as noted before, exclude transition metals, as well as noble gases.

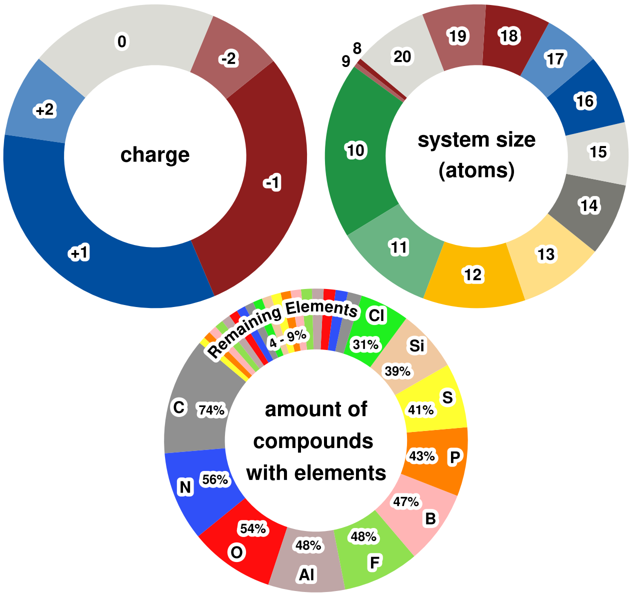


**Figure 6.** Composition of the *MB2061* set. The "amount of compounds with elements" indicates, for example, that 74% of all compounds contain at least one carbon atom.

All molecular geometries, reference values, and reaction energy coefficients are provided in the Supporting Information.

While we referred to MB2061 being a representative data set for testing quantum chemical methods beyond conventional molecules, it is important to clarify what this representativeness includes. Here, it primarily refers to the chemical diversity of the molecules, covering a broad range of main-group elements, bonding motifs, charges, and system sizes. Our data set, being generated from scratch, avoids availability bias that is typically found in literature- or database-derived sets, but other design choices still limit its scope. The exclusion of transition metals and noble gases, as well as the prescreening process (e.g., MR diagnostics and SCF stability checks), constrains the chemical space. In addition, biases related to the applied methods must be considered: geometry optimizations and initial filtering steps rely on GFN2-xTB and DFT (PBE0-D4/def2-TZVP), both of which may favor certain bonding motifs or fail in challenging electronic situations, leading to a survivor-ship bias. Although the final reference values are computed at the high-accuracy PNO-LCCSD(T)-F12 level, these choices during structure selection shape the final data set.

**Computational Details.** Using the *MB2061* benchmark set, we assess the performance and robustness of one force field (FF),[81] six SQM methods,[49,50,82−86] seven (meta-) generalized gradient approximations (GGAs),[87−96] 11 (range-separated)

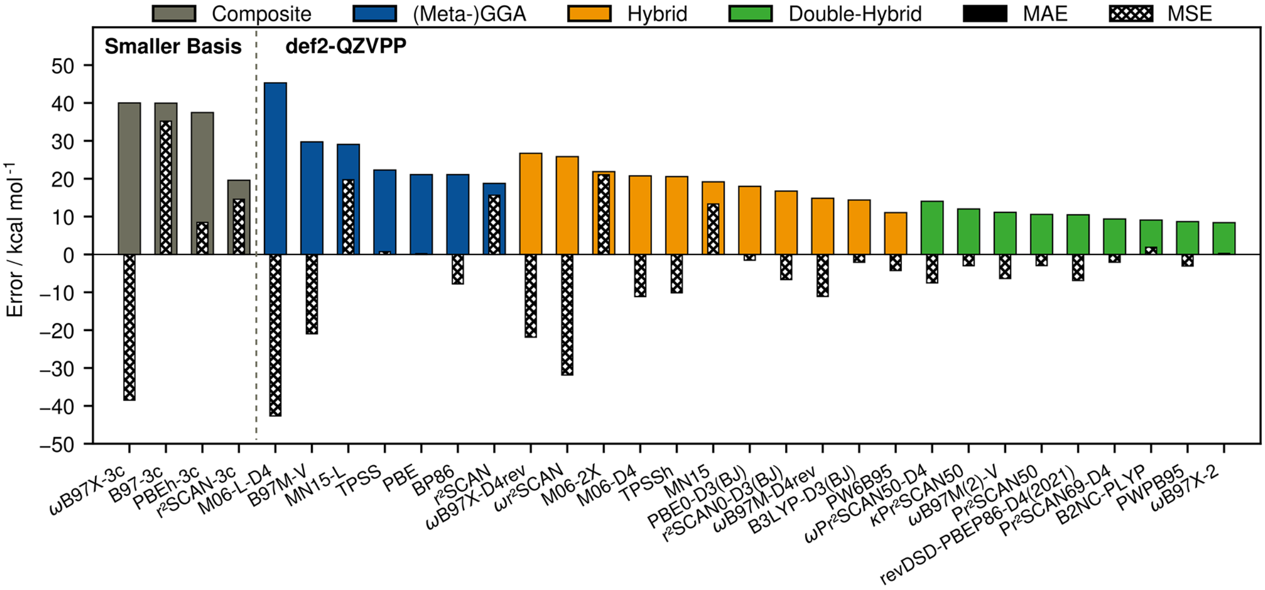


**Figure 7.** Mean absolute error (MAE) and mean signed error (MSE) of the best-performing combination of DFA and London dispersion correction for the whole *MB2061* set.

hybrids,[64,97−106] nine double-hybrid,[107,107−112] five composite methods,[113−117] and four MLIPs.[37,119,120] All DFAs (apart from composite methods) are used in combination with Ahlrichs' quadruple-ζ def2-QZVPP[66] basis set (with def2-ECPs) or with the aug-cc-pVQZ-PP-KS[76] (AVQZ-PP-KS) basis set, which is specifically recontracted for the use of ECP10MDF[121−123] pseudopotentials (PPs) for the elements Ga−Br.[76] As density functionals do not sufficiently cover London dispersion effects, we tested the D3 London dispersion correction with zero damping[124] for Minnesota functionals and Becke-Johnson (BJ) damping[125] for the remaining functionals, the D4 dispersion correction[65,126−128] as well as the nonlocal DFT-NL dispersion correction[129,130] as introduced by Vydrov and van Voorhis (VV10)[131] in a post-SCF way. All respective calculations were conducted with the ORCA 6.0.0,[55] xTB 6.7.1,[51] Molpro_2025.1,[74,75] MOPAC 2016,[132] and fairchem 2.1.0[133] program packages. All tested methods are listed in Tables S1 and S3 (see the Supporting Information). D3 and D4 London dispersion corrections were computed with the respective standalone programs s-dftd3[134,135] and dftd4.[136] For more detailed computational settings, see the Supporting Information.

## ■ RESULTS AND DISCUSSION

**Assessment of DFT.** The mean signed errors (MSEs) and mean absolute errors (MAEs) for each DFA are shown in Figure 7. Since London dispersion effects are expected to have little influence on the small decomposition reactions studied here, we evaluate the best combinations of DFA and London dispersion corrections accordingly.

The statistical trends align well with the Jacob's ladder framework proposed by Perdew et al.,[137] demonstrating that accuracy generally improves with each successive rung. In line with this view, double-hybrid functionals yield the highest accuracy. Among all tested methods, M06-L-D4 (MAE = 45.3 kcal·mol$^{-1}$) is the least accurate, while ωB97X-2 (MAE = 8.4 kcal·mol$^{-1}$) is the most accurate. However, r$^2$SCAN-3c (MAE = 19.6 kcal·mol$^{-1}$), B3LYP-D3(BJ) (MAE = 14.4 kcal·mol$^{-1}$), and PW6B95 (MAE = 11.0 kcal·mol$^{-1}$) also perform well, with r$^2$SCAN-3c being especially compelling given its modest computational cost.

The heavily parametrized Minnesota functionals MN15-L, M06-2X, and MN15L, as well as the composite methods B97-3c, PBEh-3c, and r$^2$SCAN-3c, exhibit a positive MSE, corresponding to an overbinding of the MLM. In contrast, most other DFAs show a negative MSE, indicating weaker bonding. Notably, the methods with the largest MAEs within each rung of the Jacob's ladder also feature an MSE of similar magnitude, implying systematic errors. Furthermore, it is worth mentioning that the Minnesota functionals sometimes converged to incorrect SCF minima, leading to large errors that were mitigated by using converged orbitals from B2NC-PLYP calculations as initial guesses. Similar convergence problems for these functionals have already been reported in the first "mindless" benchmark study.[41]

The tested density functionals are based on distinct design philosophies. For example, r$^2$SCAN obeys almost all known exact constraints for meta-GGA functionals, while ωB97M-V incorporates 12 empirically optimized parameters and B3LYP relies on just three. Despite these differences in the degree of empiricism, their overall performance on this data set is remarkably similar. We initially expected that functionals with extensive parameter fitting, such as MN15 or ωB97M, would perform poorly due to the presence of uncommon binding motifs in MLMs. However, no consistent trend emerges to suggest that one design strategy systematically outperforms the others, which represents a key finding in this study.

As mentioned earlier, the full data set can be divided into three subsets: *MB727-Light*, *MB782-s-Block*, and *MB552-Heavy*. The performance of the selected DFAs for each subset is shown in Figure S3 of the Supporting Information. Overall, the performance trends of the DFAs within each subset closely resemble those observed for the complete set.

The *Light* subset contains systems with a relatively simple electronic structure, resulting in slightly lower errors across all functionals. In contrast, the *s-Block* subset, which includes alkali and alkaline earth metals, presents greater challenges for the DFAs, as reflected by an increase in the number of outliers. This is particularly evident for functionals such as PW6B95 and ωB97M-D4rev, which show more pronounced deviations compared with their performance on the *Light* subset.

For the *Heavy* subset, which involves heavier elements, larger errors are observed overall compared to the *Light* and *s-Block* subsets. For instance, the MAE of PBE0-D3(BJ) increases from 15.7 kcal·mol$^{-1}$ for the *Light* subset to 22.9 kcal·mol$^{-1}$ for the *Heavy* subset, and from 5.9 to 13.7 kcal·mol$^{-1}$ for ωB97X-2.

**SQM.** As demonstrated in this work, density functional theory has proven to be a reliable choice for describing MLMs. However, it remains significantly more computationally demanding than more approximate methods such as classical FFs, SQM, or machine learning potential (MLP) approaches. Particularly for high-throughput screening applications, faster alternatives are desirable. Here, we evaluated the performance of the generic GFN force field (GFN-FF),[81] all tight-binding methods from the GFN*n*-xTB family ($n$ = 0, 1, 2),[49,50,85] the aTB[86] atomic density-based tight-binding model, the PM6-D3H4X[82,83] and PM7[84] SQM methods (see Figure S1 in the Supporting Information).

As expected, the SQM methods fail to describe the "exotic" MLMs in the test set with sufficient accuracy. The MAEs range from 183.1 kcal·mol$^{-1}$ (GFN1-xTB) to 606.5 kcal·mol$^{-1}$ (PM7), and are therefore nearly five times higher than that of the worst-performing DFAs. Notably, however, GFN-FF performs exceptionally well for an FF among the approximate methods, yielding an MAE of 218.9 kcal·mol$^{-1}$, outperforming GFN0-xTB, aTB, PM6-D3H4X, and PM7. Only a preliminary (unpublished) version of our third-generation tight-binding model (g-xTB)[118] performs very well, achieving accuracy and robustness on par with common DFT methods, with an MAE of about 35 kcal·mol$^{-1}$, a relatively low absolute maximum error (AMAX) of 162 kcal·mol$^{-1}$, and no convergence failure.

Furthermore, the data set serves as a challenging benchmark for evaluating the robustness of computational methods and their underlying implementations. This is reflected by the number of nonconverging systems or cases that converge to incorrect SCF minima. For example, aTB failed for 61 systems, GFN1-xTB failed for 67, and PM6-D3H4X failed for 90 systems.

**ML-Based Models.** By employing MLIP models, such as Aimnet2,[119,138] the accuracy can be greatly improved while maintaining a similar computational cost. Since AIMNet2 is only available for a limited but chemically relevant set of elements (H, B, C, N, O, F, Si, P, S, Cl, As, Se, Br, I), we constructed the *MB306-AIMNet2* subset, which includes all 306 structures from the *MB2061* data set that can be described using AIMNet2. We then compared its performance against DFT results (Figure S2 in the Supporting Information). The performance of AIMNet2 on this subset is great, particularly in light of its low computational cost. With an MAE of 15.5 kcal·mol$^{-1}$, it rivals those of many of the tested DFAs. For example, PBE0 yields an MAE of 16.2 kcal·mol$^{-1}$, while PW6B95 achieves 10.4 kcal·mol$^{-1}$ for the same subset. This indicates that the training set of AIMNet2 is already sufficiently broad and the model is flexible enough to describe the behavior of MLMs, at least for the included elements. However, it is worth noting that this subset does not include any metals, which are often electronically more complex than nonmetals. As a result, AIMNet2 cannot yet explore certain regions of the "mindless" chemical space that are included in the *MB2061*.

Very recently, the Open Molecules 2025 (*OMol25*)[37] data set was published, alongside some MLIPs that were trained on it.[37,120] Note that this field is currently very active, and the methods tested should be regarded as 'snapshots' of ongoing development. We evaluated the performance of the eSEN models trained exclusively on *OMol25* (esen_sm_direct_all.pt and esen_sm_conserving_all.pt), as well as the UMA-sm model (uma_sm.pt), which was trained on *OMol25* in combination with additional data sets, using the *MB2061* benchmark (see Table S2 in the Supporting Information). Unlike Aimnet2, these models can handle all elements present in *MB2061* and yield promising results. The MAEs are 26.5 kcal·mol$^{-1}$ (OMol25-eSEN-direct), 33.4 kcal·mol$^{-1}$ (OMol25-eSEN-conserving), and 23.8 kcal·mol$^{-1}$ (UMA-sm), respectively, approaching the accuracy of r$^2$SCAN-3c (19.6 kcal·mol$^{-1}$). Nevertheless, we also observe significant outliers, particularly in the eSEN models, with AMAX values reaching 892.3 kcal·mol$^{-1}$. The UMA-sm model, benefiting from a broader training base, achieves lower overall errors but still exhibits an AMAX of 374.5 kcal·mol$^{-1}$. An analysis of the subsets discussed in the DFT section reveals that the largest UMA-sm errors originate from the *MB552-Heavy* subset, which alone has an MAE of 51.2 kcal·mol$^{-1}$ (see Figure S3 in the Supporting Information). Despite these challenges, the results remain impressive and highlight the promise of MLIPs trained on diverse and chemically broad data sets.

## ■ INFLUENCE OF RELATIVISTIC EFFECTS

Since relativistic effects become more significant for heavier elements, we also examined their impact in our data set. It is important to note that relativistic effects are already accounted for—albeit approximately—through the use of effective core potentials (ECPs) in both the reference energies and the tested DFA results. Nevertheless, one important difference remains: the reference energies were obtained using the ECP10MDF pseudopotentials (PPs) for Ga–Br, whereas the DFT calculations employed def2-ECPs, which are applied only for elements heavier than Kr. Since ECPs implicitly include relativistic effects, this leads to an imbalance between the two approaches. To assess the potential impact of this discrepancy, and particularly the role of an explicit relativistic treatment, we additionally recalculated the *MB552-Heavy* subset using PBE0-D4 in combination with the AVQZ-PP-KS[76] basis set, as well as with the exact two-component (X2C)[139−141] relativistic method and the x2c-QZVPPall(−2c)[142] basis sets (cf. Figure 8). Since the X2C approach can be applied in either a one- or two-component fashion, it enables the separate evaluation of scalar relativistic (SR) and spin–orbit coupling (SOC) effects. The AVQZ-PP-KS basis set is essentially the def2-QZVPP basis set, but with a recontracted basis for Ga–Br to support the ECP10MDF PP, ensuring a more consistent comparison in this context (see ref 76 for further details). Using the full X2C approach as a reference (PBE0-D4/x2c-QZVPPall-2c with SOC), def2-QZVPP with def2-ECPs exhibits an MAE of 6.2, AVQZ-PP-KS an MAE of 4.6, and x2c-QZVPPall at the SR level exhibits an MAE of 1.3 kcal·mol$^{-1}$. These results indicate that including ECPs for Ga–Br significantly reduces the number of negative outliers and thereby lowers the overall error. Nevertheless, the deviations also show that ECPs do not fully account for relativistic effects. While an explicit treatment of relativistic effects is essential for highly accurate total electronic energies, their influence on the benchmarked reaction energies is small relative to the overall energy range and is therefore neglected in this work. Furthermore, the minimal difference between the scalar-relativistic and full SOC-X2C results demonstrates that spin–orbit coupling contributes only marginally in this context. To further assess the suitability of the AVQZ-PP-KS basis set, we

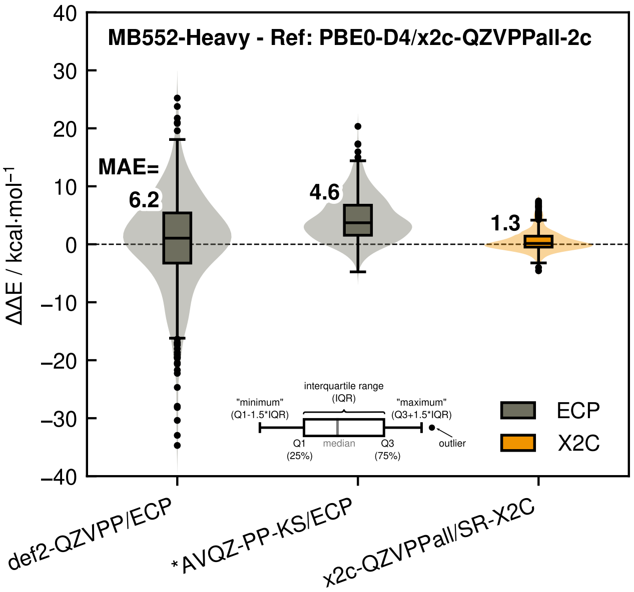


**Figure 8.** Violin plot showing the performance of PBE0-D4 in combination with ECPs and the explicit treatment of relativity using X2C. All errors are given relative to PBE0-D4/x2c-QZVPPall-2c (SO-X2C) reaction energies. *AVQZ-PP-KS is identical to def2-QZVPP, except for Ga−Br, where the basis functions were recontracted for use with ECPs (see Supporting Information or ref 76.).

also tested its performance with other density functionals. However, we observed SCF convergence issues with certain functionals such as r$^2$SCAN or $\omega$B97M-V, and the overall accuracy did not consistently improve. Owing to its robustness, we hence recommend using the def2-QZVPP basis set in combination with def2-ECPs for the computation of MLMs.

## ■ CONCLUSIONS

In this work, we explored the chemical space by systematically generating and analyzing "mindless" molecules. These structures were created through random atomic placement, followed by geometry optimization. To enable this, we developed MindlessGen, a Python-based molecular generator of MLMs, and used it to construct the *MB2061* benchmark set. This data set contains 2061 high-level reference data points computed at the PNO-LCCSD(T)-F12 level, covering $H_2$-promoted decomposition reactions of MLMs yielding small hydrides and diatomic molecules. The molecules include up to 20 atoms and span all main-group elements from hydrogen to iodine, excluding noble gases and transition metals, as well as difficult multireference cases. It includes only closed-shell singlet structures, limiting the transferability to more complicated systems.

We evaluated the performance of various computational methods, including DFT and SQM, and ML-based methods, on the *MB2061* database. The DFT results exhibit a clear Jacob's ladder trend, with accuracy improving at each successive rung. Among the tested functionals, double-hybrids proved to be the most accurate, with $\omega$B97X-2 achieving the lowest MAE (8.4 kcal·mol$^{-1}$), while r$^2$SCAN-3c (19.6 kcal·mol$^{-1}$) emerged as a cost-effective and robust alternative. Although we initially expected highly parametrized density functionals to perform poorly on this chemically unconventional set, our results show no consistent relationship between the number of empirical parameters or the fitting strategy and the overall accuracy of a given DFA.

The tested SQM methods provided mostly highly erratic results. Even the best-performing standard method, GFN1-xTB, yields a relatively large MAE of 183.1 kcal·mol$^{-1}$. However, preliminary results obtained with a development version of g-xTB indicate that DFT-level accuracy can also be achieved with SQM approaches. Similarly, modern MLIPs, such as UMA, perform well, indicating that seemingly (for a chemist) very complex molecules are less difficult to describe at an intermediate accuracy level than previously thought. The inclusion of heavier main-group elements allowed us to investigate relativistic effects. We found that the implicit relativistic corrections in ECPs already provide a sufficient description for the types of molecules studied here.

Overall, the *MB2061* benchmark set provides a chemically diverse and challenging test for quantum chemical methods. It enhances our understanding of method accuracy and robustness on unconventional molecular systems, thereby supporting continued exploration of uncharted regions in the chemical space of closed-shell systems. While DFT methods can describe MLMs with high accuracy, the greatest potential for improvement lies in the development of SQM, FF, and ML approaches. *MB2061* is well-suited not only as a validation benchmark but also as training data for data-efficient learning strategies. In particular, it offers a promising foundation for active learning workflows, such as the regAL framework by Proppe and coworkers.[143] Looking ahead, we anticipate that MindlessGen and the *MB2061* set will serve as a valuable resource for the parametrization and testing of next-generation models, ultimately contributing to more robust and transferable methods across the chemical space.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

The Python-based source code of MindlessGen is openly available at: https://github.com/grimme-lab/MindlessGen. For setup and usage instructions, please consult the documentation provided in the repository. All geometries (XYZ files), reference energies, and the coefficients used to compute reaction energies for the *MB2061* benchmark set are provided in the mb2061.zip archive, available as Supporting Information with this work. The decomposition reaction schemes can be found in the .res file included in mb2061.zip, where each line beginning with $tmer specifies one reaction. These lines list the reaction partners, their respective coefficients, and the reference reaction energy in sequence.

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.5c01364.

Reference energies, and reaction energy coefficients of the *MB2061* benchmark set (ZIP)

Additional technical details and supporting figures (PDF)

Geometries and calculated reaction energies with SQM, DFT, and MLIPs, as well as total energies computed with the reference level (XLSX)

## ■ AUTHOR INFORMATION

### Corresponding Author

**Andreas Hansen** − *Mulliken Center for Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany;* orcid.org/0000-0003-1659-8206; Email: hansen@thch.uni-bonn.de

### Authors

**Thomas Gasevic** − *Mulliken Center for Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany;* orcid.org/0000-0003-4864-1758
**Marcel Müller** − *Mulliken Center for Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany;* orcid.org/0000-0002-5398-4192
**Jonathan Schöps** − *Mulliken Center for Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany;* orcid.org/0009-0003-4163-646X
**Stephanie Lanius** − *Microsoft Research AI for Science, 10178 Berlin, Germany;* orcid.org/0009-0004-8212-3238
**Jan Hermann** − *Microsoft Research AI for Science, 10178 Berlin, Germany;* orcid.org/0000-0002-2779-0749
**Stefan Grimme** − *Mulliken Center for Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany;* orcid.org/0000-0002-5844-4371

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jcim.5c01364

### Author Contributions

A.H. proposed and developed the initial concept for the project. The project was then conceptualized by A.H., J.H., S.G., T.G., and M.M., M.M., and J.S., developed the MindlessGen and coauthored the corresponding manuscript section. T.G. generated the geometries for the MB2061 set, carried out all subsequent DFT, SQM, and MLIP calculations, and wrote the remaining parts of the manuscript. S.L. contributed to data curation. T.G., A.H., and J.H. analyzed the data. While S.G. supervised the MindlessGen part of the project, J.H. and A.H. supervised the benchmark creation and evaluation part. All authors contributed to the review and editing process.

### Notes

The authors declare the following competing financial interest(s): S.L. and J.H. are employees of Microsoft Research AI for Science. This work was supported in part by Microsoft Research. The authors declare no other competing financial interests.

## ■ ACKNOWLEDGMENTS

The authors thank L. Seidler for implementing a sophisticated parallelization scheme for the MindlessGen, M. Friede for fitting the D4 London dispersion correction for B2NC-PLYP, and T. Froitzheim for helpful discussions during the development of the MindlessGen program. The authors gratefully acknowledge access to the Marvin Computing Cluster of the University of Bonn. The authors also acknowledge the Microsoft Accelerating Foundation Models Research Program for access to Microsoft Azure cloud computing. T.G. thanks Microsoft Research for funding. M.M. thanks the Fonds der Chemischen Industrie (FCI) for funding via a Kekulé scholarship.

## ■ REFERENCES

(1) Reymond, J.-L. The Chemical Space Project. *Acc. Chem. Res.* **2015**, *48*, 722−730.

(2) Cleaves, H. J. I.; Butch, C.; Burger, P. B.; Goodwin, J.; Meringer, M. One Among Millions: The Chemical Space of Nucleic Acid-Like Molecules. *J. Chem. Inf. Model.* **2019**, *59*, 4266−4277.

(3) Reymond, J. L. Chemical space as a unifying theme for chemistry. *J. Cheminform.* **2025**, *17*, 6.

(4) Samanipour, S.; Barron, L. P.; van Herwerden, D.; Praetorius, A.; Thomas, K. V.; O'Brien, J. W. Exploring the Chemical Space of the Exposome: How Far Have We Gone? *JACS Au* **2024**, *4*, 2412−2425.

(5) Lu, C.; Liu, S.; Shi, W.; Yu, J.; Zhou, Z.; Zhang, X.; Lu, X.; Cai, F.; Xia, N.; Wang, Y. Systemic evolutionary chemical space exploration for drug discovery. *J. Cheminform.* **2022**, *14*, 19.

(6) Reymond, J.-L.; Awale, M. Exploring Chemical Space for Drug Discovery Using the Chemical Universe Database. *ACS Chem. Neurosci.* **2012**, *3*, 649−657.

(7) Khalak, Y.; Tresadern, G.; Hahn, D. F.; de Groot, B. L.; Gapsys, V. Chemical Space Exploration with Active Learning and Alchemical Free Energies. *J. Chem. Theory Comput.* **2022**, *18*, 6259−6270.

(8) Cheng, C. Y.; Campbell, J. E.; Day, G. M. Evolutionary chemical space exploration for functional materials: computational organic semiconductor discovery. *Chem. Sci.* **2020**, *11*, 4922−4933.

(9) Mroz, A. M.; Posligua, V.; Tarzia, A.; Wolpert, E. H.; Jelfs, K. E. Into the Unknown: How Computation Can Help Explore Uncharted Material Space. *J. Am. Chem. Soc.* **2022**, *144*, 18730−18743.

(10) Park, H.; Onwuli, A.; Butler, K. T.; Walsh, A. Mapping inorganic crystal chemical space. *Faraday Discuss.* **2025**, *256*, 601−613.

(11) Meyers, J.; Fabian, B.; Brown, N. De novo molecular design and generative models. *Drug Discovery Today* **2021**, *26*, 2707−2715.

(12) Zhu, H.; Zhang, J.; Kim, M. T.; Boison, A.; Sedykh, A.; Moran, K. Big data in chemical toxicity research: the use of high-throughput screening assays to identify potential toxicants. *Chem. Res. Toxicol.* **2014**, *27*, 1643−1651.

(13) Wagen, C. C.; McMinn, S. E.; Kwan, E. E.; Jacobsen, E. N. Screening for generality in asymmetric catalysis. *Nat.* **2022**, *610*, 680−686.

(14) Rodríguez-Pérez, R.; Miljković, F.; Bajorath, J. Machine Learning in Chemoinformatics and Medicinal Chemistry. *Annu. Rev. Biomed. Data Sci.* **2022**, *5*, 43−65.

(15) Chan, B.; Dawson, W.; Nakajima, T. Data Quality in the Fitting of Approximate Models: A Computational Chemistry Perspective. *J. Chem. Theory Comput.* **2024**, *20*, 10468−10476.

(16) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nat.* **2018**, *559*, 547−555.

(17) Huang, B.; von Lilienfeld, O. A. Ab Initio Machine Learning in Chemical Compound Space. *Chem. Rev.* **2021**, *121*, 10001−10036.

(18) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Challenges for Density Functional Theory. *Chem. Rev.* **2012**, *112*, 289−320.

(19) Bursch, M.; Mewes, J.; Hansen, A.; Grimme, S. Best-Practice DFT Protocols for Basic Molecular Computational Chemistry. *Angew. Chem., Int. Ed.* **2022**, *61*, No. e202205735.

(20) Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **2017**, *115*, 2315−2372.

(21) Medvedev, M. G.; Bushmarinov, I. S.; Sun, J.; Perdew, J. P.; Lyssenko, K. A. Density functional theory is straying from the path toward the exact functional. *Sci.* **2017**, *355*, 49−52.

(22) Kepp, K. P. Comment on "Density functional theory is straying from the path toward the exact functional". *Sci.* **2017**, *356*, 496−496.

(23) Medvedev, M. G.; Bushmarinov, I. S.; Sun, J.; Perdew, J. P.; Lyssenko, K. A. Response to Comment on "Density functional theory is straying from the path toward the exact functional". *Sci.* **2017**, *356*, 496−496.

(24) Karton, A.; de Oliveira, M. T. Good Practices in Database Generation for Benchmarking Density Functional Theory. *WIREs Comput. Mol. Sci.* **2025**, *15*, No. e1737.

(25) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; et al. PubChem 2023 update. *Nucleic Acids Res.* **2023**, *51*, D1373−D1380.

(26) Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M. P.; Mosquera, J. F.; Mutowo, P.; Nowotka, M.; et al. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* **2019**, *47*, D930−D940.

(27) Ullah, A.; Chen, Y.; Dral, P. O. Molecular quantum chemical data sets and databases for machine learning potentials. *Mach. Learn.: Sci. Technol.* **2024**, *5*, No. 041001.

(28) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; Von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **2012**, *108*, No. 058301.

(29) Montavon, G.; Rupp, M.; Gobre, V.; Vazquez-Mayagoitia, A.; Hansen, K.; Tkatchenko, A.; Müller, K.-R.; Anatole von Lilienfeld, O. Machine learning of molecular electronic properties in chemical compound space. *New J. Phys.* **2013**, *15*, No. 095003.

(30) Hoja, J.; Medrano Sandonas, L.; Ernst, B. G.; Vazquez-Mayagoitia, A.; DiStasio, R. A.; Tkatchenko, A. QM7-X, a comprehensive dataset of quantum-mechanical properties spanning the chemical space of small organic molecules. *Sci. Data* **2021**, *8*, 43.

(31) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **2014**, *1*, 140022.

(32) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **2017**, *8*, 3192−3203.

(33) Smith, J. S.; Zubatyuk, R.; Nebgen, B.; Lubbers, N.; Barros, K.; Roitberg, A. E.; Isayev, O.; Tretiak, S. The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules. *Sci. Data* **2020**, *7*, 134.

(34) Nakata, M.; Shimazaki, T. PubChemQC project: a large-scale first-principles electronic structure database for data-driven chemistry. *J. Chem. Inf. Model.* **2017**, *57*, 1300−1308.

(35) Irwin, J. J.; Tang, K. G.; Young, J.; Dandarchuluun, C.; Wong, B. R.; Khurelbaatar, M.; Moroz, Y. S.; Mayfield, J.; Sayle, R. A. ZINC20—A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *J. Chem. Inf. Model.* **2020**, *60*, 6065−6073.

(36) Axelrod, S.; Gómez-Bombarelli, R. GEOM, energy-annotated molecular conformations for property prediction and molecular generation. *Sci. Data* **2022**, *9*, 185.

(37) Levine, D. S.et al.The Open Molecules 2025 (OMol25) Dataset, Evaluations, and Models, 2025.

(38) 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.

(39) Blum, L. C.; Reymond, J.-L. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J. Am. Chem. Soc.* **2009**, *131*, 8732−8733.

(40) Ruddigkeit, L.; Van Deursen, R.; Blum, L. C.; Reymond, J.-L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.* **2012**, *52*, 2864−2875.

(41) Korth, M.; Grimme, S. Mindless DFT benchmarking. *J. Chem. Theor. Comp.* **2009**, *5*, 993−1003.

(42) Khan, D.; Benali, A.; Kim, S. Y. H.; von Rudorff, G. F.; von Lilienfeld, O. A. *Quantum mechanical dataset of 836k neutral closed shell molecules with upto 5 heavy atoms from CNOFSiPSClBr*, 2024.

(43) Ganscha, S.; Unke, O. T.; Ahlin, D.; Maennel, H.; Kashubin, S.; Müller, K.-R. The QCML dataset, Quantum chemistry reference data from 33.5M DFT and 14.7B semi-empirical calculations. *Sci. Data* **2025**, *12*, 406.

(44) Gould, T.; Chan, B.; Dale, S. G.; Vuckovic, S. Identifying and embedding transferability in data-driven representations of chemical space. *Chem. Sci.* **2024**, *15*, 11122−11133.

(45) Gould, T.; Vuckovic, S. "Slim" benchmark sets for faster method development. *J. Chem. Theory Comput.* **2025**, *21*, 6517.

(46) Arús-Pous, J.; Blaschke, T.; Ulander, S.; Reymond, J. L.; Chen, H.; Engkvist, O. Exploring the GDB-13 chemical space using deep generative models. *J. Cheminform.* **2019**, *11*, 20.

(47) Türk, H.; Landini, E.; Kunkel, C.; Margraf, J. T.; Reuter, K. Assessing Deep Generative Models in Chemical Composition Space. *Chem. Mater.* **2022**, *34*, 9455−9467.

(48) Anstine, D. M.; Isayev, O. Generative Models as an Emerging Paradigm in the Chemical Sciences. *J. Am. Chem. Soc.* **2023**, *145*, 8736−8750.

(49) 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.

(50) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652−1671.

(51) 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.* **2021**, *11*, No. e1493.

(52) Katbashev, A.; Stahn, M.; Rose, T.; Alizadeh, V.; Friede, M.; Plett, C.; Steinbach, P.; Ehlert, S. Overview on Building Blocks and Applications of Efficient and Robust Extended Tight Binding. *J. Phys. Chem. A* **2025**, *129*, 2667−2682.

(53) Pyykkö, P.; Atsumi, M. Molecular Single-Bond Covalent Radii for Elements 1−118. *Chem.—Eur. J.* **2009**, *15*, 186−197.

(54) TURBOMOLE GmbH, *TURBOMOLE V7.9, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007*, 2025; https://www.turbomole.org/.

(55) Neese, F. Software Update: The ORCA Program System—Version 6.0. *WIREs Comput. Mol. Sci.* **2025**, *15*, No. e70019.

(56) Shelby, J.; Noonan, J. Formation and properties of sodium indium silicate glasses. *Phys. Chem. Glas.* **1998**, *39*, 36−40.

(57) Buchholz, D. B.; Ma, Q.; Alducin, D.; Ponce, A.; Jose-Yacaman, M.; Khanal, R.; Medvedeva, J. E.; Chang, R. P. H. The Structure and Properties of Amorphous Indium Oxide. *Chem. Mater.* **2014**, *26*, 5401−5411.

(58) Grimme, S.; Hansen, A. A practicable real-space measure and visualization of static electron-correlation effects. *Angew. Chem. - Int. Ed.* **2015**, *54*, 12308−12313.

(59) Bauer, C. A.; Hansen, A.; Grimme, S. The fractional occupation number weighted density as a versatile analysis tool for molecules with a complicated electronic structure. *Chem.—Eur. J.* **2017**, *23*, 6150−6164.

(60) Ramos-Cordoba, E.; Salvador, P.; Matito, E. Separation of dynamic and nondynamic correlation. *Phys. Chem. Chem. Phys.* **2016**, *18*, 24015−24023.

(61) Bauernschmitt, R.; Ahlrichs, R. Stability analysis for solutions of the closed shell Kohn−Sham equation. *J. Chem. Phys.* **1996**, *104*, 9047−9052.

(62) https://www.faccts.de/docs/orca/6.0/manual/contents/detailed/stabilityanalysis.html, [Accessed 06-08-2025].

(63) Neugebauer, H.; Vuong, H. T.; Weber, J. L.; Friesner, R. A.; Shee, J.; Hansen, A. Toward Benchmark-Quality Ab Initio Predictions for 3d Transition Metal Electrocatalysts: A Comparison of CCSD(T) and ph-AFQMC. *J. Chem. Theory Comput.* **2023**, *19*, 6208−6225.

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

(65) Caldeweyher, E.; Bannwarth, C.; Grimme, S. Extension of the D3 dispersion coefficient model. *J. Chem. Phys.* **2017**, *147*, No. 034112.

(66) 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.

(67) Werner, H.-J.; Knizia, G.; Krause, C.; Schwilk, M.; Dornbach, M. Scalable electron correlation methods I.: PNO-LMP2 with linear scaling in the molecular size and near-inverse-linear scaling in the number of processors. *J. Chem. Theory Comput.* **2015**, *11*, 484−507.

(68) Ma, Q.; Werner, H.-J. Scalable electron correlation methods. 2. Parallel PNO-LMP2-F12 with near linear scaling in the molecular size. *J. Chem. Theory Comput.* **2015**, *11*, 5291−5304.

(69) Schwilk, M.; Ma, Q.; Köppl, C.; Werner, H.-J. Scalable electron correlation methods. 3. Efficient and accurate parallel local coupled cluster with pair natural orbitals (PNO-LCCSD). *J. Chem. Theory Comput.* **2017**, *13*, 3650−3675.

(70) Ma, Q.; Schwilk, M.; Köppl, C.; Werner, H.-J. Scalable electron correlation methods. 4. Parallel explicitly correlated local coupled

cluster with pair natural orbitals (PNO-LCCSD-F12). *J. Chem. Theory Comput.* **2017**, *13*, 4871−4896.

(71) Ma, Q.; Werner, H.-J. Scalable electron correlation methods. 5. Parallel perturbative triples correction for explicitly correlated local coupled cluster with pair natural orbitals. *J. Chem. Theory Comput.* **2018**, *14*, 198−215.

(72) Ma, Q.; Werner, H.-J. Scalable electron correlation methods. 7. Local open-shell coupled-cluster methods using pair natural orbitals: PNO-RCCSD and PNO-UCCSD. *J. Chem. Theory Comput.* **2020**, *16*, 3135−3151.

(73) Ma, Q.; Werner, H.-J. Scalable electron correlation methods. 8. Explicitly correlated open-shell coupled-cluster with pair natural orbitals PNO-RCCSD (T)-F12 and PNO-UCCSD (T)-F12. *J. Chem. Theory Comput.* **2021**, *17*, 902−926.

(74) Werner, H.-J.et al.MOLPRO, version 2025.1, a package of ab initio programs. 2025; see https://www.molpro.net.

(75) Werner, H. J.; Knowles, P. J.; Manby, F. R.; Black, J. A.; Doll, K.; Heßelmann, A.; Kats, D.; Köhn, A.; Korona, T.; Kreplin, D. A.; Ma, Q.; Miller, T. F.; Mitrushchenkov, A.; Peterson, K. A.; Polyak, I.; Rauhut, G.; Sibaev, M. The Molpro quantum chemistry package. *J. Chem. Phys.* **2020**, *152*, 144107.

(76) Gasevic, T.; Bursch, M.; Ma, Q.; Grimme, S.; Werner, H.-J.; Hansen, A. The p-block challenge: assessing quantum chemistry methods for inorganic heterocycle dimerizations. *Phys. Chem. Chem. Phys.* **2024**, *26*, 13884−13908.

(77) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. Molpro: a general-purpose quantum chemistry program package. *WIREs Comput. Mol. Sci.* **2012**, *2*, 242−253.

(78) Werner, H.-J.; Hansen, A. Accurate Calculation of Isomerization and Conformational Energies of Larger Molecules Using Explicitly Correlated Local Coupled Cluster Methods in Molpro and ORCA. *J. Chem. Theory Comput.* **2023**, *19*, 7007−7030.

(79) Plett, C.; Grimme, S.; Hansen, A. Toward Reliable Conformational Energies of Amino Acids and Dipeptides—The DipCONFS Benchmark and DipCONL Datasets. *J. Chem. Theory Comput.* **2024**, *20*, 8329−8339.

(80) Plett, C.; Grimme, S.; Hansen, A. Conformational energies of biomolecules in solution: Extending the MPCONF196 benchmark with explicit water molecules. *J. Comput. Chem.* **2024**, *45*, 419−429.

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

(82) 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.

(83) Brahmkshatriya, P. S.; Dobes, P.; Fanfrlik, J.; Rezac, J.; Paruch, K.; Bronowska, A.; Lepsík, 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.

(84) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods VI: More Modifications to the NDDO Approximations and Re-optimization of Parameters. *J. Mol. Model.* **2013**, *19*, 1−32.

(85) Pracht, P.; Caldeweyher, E.; Ehlert, S.; Grimme, S.; A robust non-self-consistent tight-binding quantum chemistry method for large molecules, **2019**,

(86) Zhang, Y.; Xiao, J.; Wang, S.; Zhu, T.; Zhang, J. Z. H. The Atomic Density-Based Tight-Binding (aTB) Model: A Robust and Accurate Semiempirical Method Parametrized for H−Ra; Applied to Structures, Vibrational Frequencies, Noncovalent Interactions, and Excited States. *J. Chem. Theory Comput.* **2025**, *21*, 3410−3425.

(87) 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*, 74111.

(88) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098−3100.

(89) Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33*, 8822−8824.

(90) 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.

(91) Yu, H. S.; He, X.; Truhlar, D. G. MN15-L: A New Local Exchange-Correlation Functional for Kohn-Sham Density Functional Theory with Broad Accuracy for Atoms, Molecules, and Solids. *J. Chem. Theory Comput.* **2016**, *12*, 1280−1293.

(92) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.

(93) Furness, J. W.; Kaplan, A. D.; Ning, J.; Perdew, J. P.; Sun, J. Accurate and Numerically Efficient r$^2$SCAN Meta-Generalized Gradient Approximation. *J. Phys. Chem. Lett.* **2020**, *11*, 8208−8215.

(94) Furness, J. W.; Kaplan, A. D.; Ning, J.; Perdew, J. P.; Sun, J. Correction to: 'Accurate and Numerically Efficient r$^2$SCAN Meta-Generalized Gradient Approximation' (J. Phys. Chem. Lett. (2020) 11:19 (8208−8215) DOI: 10.1021/acs.jpclett.0c02405). *J. Phys. Chem. Lett.* **2020**, *11*, 9248.

(95) Ehlert, S.; Huniar, U.; Ning, J.; Furness, J. W.; Sun, J.; Kaplan, A. D.; Perdew, J. P.; Brandenburg, J. G. r$^2$SCAN-D4: Dispersion corrected meta-generalized gradient approximation for general chemical applications. *J. Chem. Phys.* **2021**, *154*, No. 061101.

(96) 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*, No. 146401.

(97) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5656.

(98) 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. Chem.* **1994**, *98*, 11623−11627.

(99) 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 function. *Theor. Chem. Acc.* **2008**, *120*, 215−241.

(100) Yu, H. S.; He, X.; Li, S. L.; Truhlar, D. G. MN15: A Kohn-Sham global-hybrid exchange-correlation density functional with broad accuracy for multi-reference and single-reference systems and non-covalent interactions. *Chem. Sci.* **2016**, *7*, 5032−5051.

(101) 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.

(102) Bursch, M.; Neugebauer, H.; Ehlert, S.; Grimme, S. Dispersion Corrected r$^2$SCAN Based Global Hybrid Functionals: r$^2$SCANh, r$^2$SCAN0, and r$^2$SCAN50. *J. Chem. Phys.* **2022**, *156*, 10−12.

(103) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J. Chem. Phys.* **2003**, *119*, 12129−12137.

(104) Mardirossian, N.; Head-Gordon, M. $\omega$B97X-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.

(105) Mardirossian, N.; Head-Gordon, M. $\omega$B97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **2016**, *144*, No. 214110.

(106) Friede, M.; Ehlert, S.; Grimme, S.; Mewes, J.-M. Do Optimally Tuned Range-Separated Hybrid Functionals Require a Reparametrization of the Dispersion Correction? *It Depends. J. Chem. Theory Comput.* **2023**, *19*, 8097−8107.

(107) Wittmann, L.; Neugebauer, H.; Grimme, S.; Bursch, M. Dispersion-corrected r$^2$SCAN based double-hybrid functionals. *J. Chem. Phys.* **2023**, *159*, 22413.

(108) Yu, F. Double-Hybrid Density Functionals Free of Dispersion and Counterpoise Corrections for Non-Covalent Interactions. *J. Phys. Chem. A* **2014**, *118*, 3175−3182.

(109) 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.

(110) Santra, G.; Cho, M.; Martin, J. M. Exploring Avenues beyond Revised DSD Functionals: I. Range Separation, with xDSD as a Special Case. *J. Phys. Chem. A* **2021**, *125*, 4614−4627.

(111) Mardirossian, N.; Head-Gordon, M. Survival of the most transferable at the top of Jacob's ladder: Defining and testing the *ω*B97M(2) double hybrid density functional. *J. Chem. Phys.* **2018**, *148*, 241736.

(112) Chai, J. D.; Head-Gordon, M. Long-range corrected double-hybrid density functionals. *J. Chem. Phys.* **2009**, *131*, 174105.

(113) Sure, R.; Grimme, S. Corrected Small Basis Set Hartree-Fock Method for Large Systems. *J. Comput. Chem.* **2013**, *34*, 1672−1685.

(114) 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*, No. 064104.

(115) Grimme, S.; Hansen, A.; Ehlert, S.; Mewes, J. M. r$^2$SCAN-3c: A "swiss army knife" composite electronic-structure method. *J. Chem. Phys.* **2021**, *154*, No. 064103.

(116) 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*, No. 054107.

(117) Müller, M.; Hansen, A.; Grimme, S. *ω*B97X-3c: A composite range-separated hybrid DFT method with a molecule-optimized polarized valence double-*ζ* basis set. *J. Chem. Phys.* **2023**, *158*, 14103.

(118) Froitzheim, T.; Müller, M.; Hansen, A.; Grimme, S. g-xTB: A General-Purpose Extended Tight-Binding Electronic Structure Method For the Elements H to Lr (Z=1−103). *ChemRxiv.* **2025**, DOI: 10.26434/chemrxiv-2025-bjxvt.

(119) Anstine, D. M.; Zubatyuk, R.; Isayev, O. AIMNet2: A Neural Network Potential to Meet your Neutral, Charged, Organic, and Elemental-Organic Needs. *Chem. Sci.* **2025**, *16*, 10228−10244.

(120) Wood, B. M.et al.UMA: A Family of Universal Models for Atoms, 2025.

(121) Metz, B.; Schweizer, M.; Stoll, H.; Dolg, M.; Liu, W. A small-core multiconfiguration Dirac−Hartree−Fock-adjusted pseudopotential for Tl−application to Tl X (X= F, Cl, Br, I). *Theor. Chem. Acc.* **2000**, *104*, 22−28.

(122) Metz, B.; Stoll, H.; Dolg, M. Small-core multiconfiguration-Dirac−Hartree−Fock-adjusted pseudopotentials for post-d main group elements: Application to PbH and PbO. *J. Chem. Phys.* **2000**, *113*, 2563−2569.

(123) Peterson, K. A.; Figgen, D.; Goll, E.; Stoll, H.; Dolg, M. Systematically convergent basis sets with relativistic pseudopotentials. II. Small-core pseudopotentials and correlation consistent basis sets for the post-d group 16−18 elements. *J. Chem. Phys.* **2003**, *119*, 11113−11123.

(124) 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.

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

(126) 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.

(127) 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**, *22*, 8499−8512.

(128) Wittmann, L.; Gordiy, I.; Friede, M.; Helmich-Paris, B.; Grimme, S.; Hansen, A.; Bursch, M. Extension of the D3 and D4 London dispersion corrections to the full actinides series. *Phys. Chem. Chem. Phys.* **2024**, *26*, 21379−21394.

(129) Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals density functional made simple. *Phys. Rev. Lett.* **2009**, *103*, No. 063004.

(130) Hujo, W.; Grimme, S. Performance of the van der waals density functional VV10 and (hybrid)GGA variants for thermochemistry and noncovalent interactions. *J. Chem. Theory Comput.* **2011**, *7*, 3866−3871.

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

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

(133) GitHub—facebookresearch/fairchem: FAIR Chemistry's library of machine learning methods for chemistry—github.com. https://github.com/facebookresearch/fairchem, [Accessed 04-06-2025].

(134) s-DFT-D3 at GitHub, https://github.com/dftd3/simple-dftd3 [Accessed 06-02-2025].

(135) Ehlert, S. Simple DFT-D3: Library first implementation of the D3 dispersion correction. *J. Open Source Softw.* **2024**, *9*, 7169.

(136) DFT-D4 at GitHub, https://github.com/dftd4/dftd4 [Accessed 06-02-2025].

(137) Perdew, J. P.; Schmidt, K. Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conf. Proc.* **2001**, *577*, 1−20.

(138) GitHub—isayevlab/AIMNet2—github.com. https://github.com/isayevlab/AIMNet2, [Accessed 04-06-2025].

(139) Kutzelnigg, W.; Liu, W. Quasirelativistic theory equivalent to fully relativistic theory. *J. Chem. Phys.* **2005**, *123*, 241102.

(140) Liu, W.; Peng, D. Infinite-order quasirelativistic density functional method based on the exact matrix quasirelativistic theory. *J. Chem. Phys.* **2006**, *125*, No. 044102.

(141) Iliaš, M.; Saue, T. An infinite-order two-component relativistic Hamiltonian by a simple one-step transformation. *J. Chem. Phys.* **2007**, *126*, No. 064102.

(142) Franzke, Y. J.; Spiske, L.; Pollak, P.; Weigend, F. Segmented Contracted Error-Consistent Basis Sets of Quadruple-*ζ* Valence Quality for One- and Two-Component Relativistic All-Electron Calculations. *J. Chem. Theory Comput.* **2020**, *16*, 5658−5674.

(143) Surzhikova, E.; Proppe, J. regAL: Python Package for Active Learning of Regression Problems, 2024. 10.48550/arXiv.2410.17917

# APPENDIX E

# Confined Lewis Pairs: Investigation of the $X^{-}\rightarrow Si_{20}$ Interaction in Halogen-Encapsulating Silafulleranes

Thomas Gasevic,[†] Marcel Bamberg,[‡] Julius Wicke,[‡] Michael Bolte,[‡] Alexander Virovets,[‡] Hans-Wolfram Lerner,[‡] Stefan Grimme,[†] Andreas Hansen,[†] Matthias Wagner,[‡] and Markus Bursch,[†]

*Received: September 27, 2023*
*First published: December 7, 2023*

Reprinted (adapted) with permission from:
T. Gasevic, M. Bamberg, J. Wicke, M. Bolte, A. Virovets, H.-W. Lerner, S. Grimme, A. Hansen, M. Wagner, and M. Bursch, *Confined Lewis Pairs: Investigation of the $X^{-}\rightarrow Si_{20}$ Interaction in Halogen-Encapsulating Silafulleranes*, Angew. Chem., Int. Ed. **63**.6 (2024) e202314238, DOI: `10.1002/anie.202314238`

**Own contributions**

- Conceptualization
- Formal Analysis
- Investigation – Performing DFT and LED calculations
- Writing & Visualization – Original draft preparation, review & editing

---

[†] Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany

[‡] Institut für Anorganische und Analytische Chemie, Goethe-Universität Frankfurt am Main, 60438 Frankfurt am Main, Germany

How to cite: *Angew. Chem. Int. Ed.* **2024**, *63*, e202314238
doi.org/10.1002/anie.202314238

Theoretical Chemistry

# Confined Lewis Pairs: Investigation of the $X^{-} \rightarrow Si_{20}$ Interaction in Halogen-Encapsulating Silafulleranes

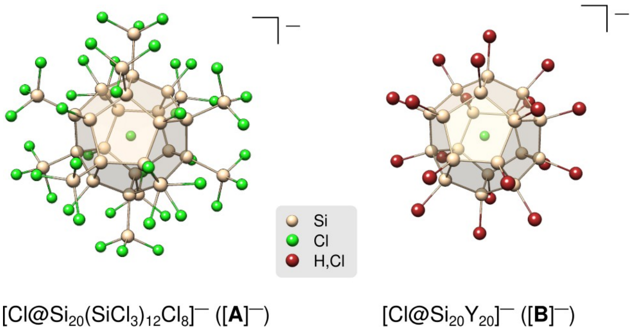
*Thomas Gasevic[+], Marcel Bamberg[+], Julius Wicke, Michael Bolte, Alexander Virovets, Hans-Wolfram Lerner, Stefan Grimme, Andreas Hansen,* Matthias Wagner,* and Markus Bursch**

**Abstract:** A joint theoretical and experimental study on 32 endohedral silafullerane derivatives $[X@Si_{20}Y_{20}]^{-}$ (X=F-I; Y=F-I, H, Me, Et) and $T_h$-$[Cl@Si_{20}H_{12}Y_{8}]^{-}$ (Y=F-I) is presented. First, we evaluated the structure-determining template effect of $Cl^{-}$ in a systematic series of concave silapolyquinane model systems. Second, we investigated the $X^{-} \rightarrow Si_{20}$ interaction energy ($E_{int}$) as a function of $X^{-}$ and Y and found the largest $E_{int}$ values for electron-withdrawing exohedral substituents Y. Given that $X^{-}$ ions can be considered as Lewis bases and empty $Si_{20}Y_{20}$ clusters as Lewis acids, we classify our inseparable host–guest complexes $[X@Si_{20}Y_{20}]^{-}$ as "confined Lewis pairs". Third, $^{35}Cl$ NMR spectroscopy proved to be highly diagnostic for an experimental assessment of the $Cl^{-} \rightarrow Si_{20}$ interaction as the paramagnetic shielding and, in turn, $\delta(^{35}Cl)$ of the endohedral $Cl^{-}$ ion correlate inversely with $E_{int}$. Finally, we disclose the synthesis of $[PPN][Cl@Si_{20}Y_{20}]$ (Y=Me, Et, Br) and provide a thorough characterization of these new silafulleranes.

## Introduction

Mapping the similarities and differences between carbon and its heavier homolog silicon is a central theme in main group chemistry.[1] Therefore, mimicking structural motifs with silicon that are known from carbon-containing molecules has become a fundamental motivation for the synthesis of new silicon compounds.[2–4] Shortly after their discovery, fullerenes ($C_{20}$-$C_{70}$; Kroto, Smalley, Krätschmer, Prinzbach *et al.*, 1985)[5–9] and dodecahedrane ($C_{20}H_{20}$; Paquette *et al.*, 1982)[10] were considered milestones in the field of nanoscience and sparked a growing interest also in spherical molecules made of silicon atoms. However, this compound class remained synthetically inaccessible for a long time. A special interest arose for the smallest possible silafullerene $I_h$-$Si_{20}$ after Prinzbach *et al.* were able to generate the carbonaceous congener $C_{20}$ in the gas phase.[8] Since calculations suggested that the dodecahedral $Si_{20}$ cage is not stable but should collapse into more compact structures,[11–16] the following strategies for its stabilization were postulated by theory: (i) encapsulation of formally neutral or charged guests $X^{n+/n-}$ to obtain endohedral complexes $[X@Si_{20}]^{n+/n-}$ (e.g., $X^{n+/n-}=Ba^{0}, Zr^{0}, U^{6-}$),[17–22] (ii) exohedral saturation by switching from the silafullerene $Si_{20}$ to a silafullerane $Si_{20}Y_{20}$ (e.g., Y=H),[23–36] or (iii) the combination of both approaches leading to endohedral silafulleranes $[X@Si_{20}Y_{20}]^{n+/n-}$ (e.g., $X^{n+/n-}=Li^{+}, Cl^{-}, O^{2-}$; Y=H, F).[37–42]

In 2015, the Wagner group finally discovered a one-step protocol for the synthesis of the [20]silafullerane $T_h$-$[Cl@Si_{20}(SiCl_3)_{12}Cl_8]^{-}$, which was isolated as the tetraalkylammonium salt $[R_4N][\mathbf{A}]$ (R=Et, *n*Bu; Figure 1).



$[Cl@Si_{20}(SiCl_3)_{12}Cl_8]^{-}$ ($[\mathbf{A}]^{-}$)    $[Cl@Si_{20}Y_{20}]^{-}$ ($[\mathbf{B}]^{-}$)

***Figure 1.*** Molecular structures of the [20]silafulleranes $[\mathbf{A}]^{-}$ and $[\mathbf{B}]^{-}$.

---

[*] T. Gasevic,[+] Prof. Dr. S. Grimme, Dr. A. Hansen
Mulliken Center for Theoretical Chemistry,
University of Bonn
Beringstraße 4, 53115 Bonn (Germany)
E-mail: hansen@thch.uni-bonn.de

Dr. M. Bamberg,[+] J. Wicke, Dr. M. Bolte, Dr. A. Virovets,
Dr. H.-W. Lerner, Prof. Dr. M. Wagner
Institut für Anorganische und Analytische Chemie,
Goethe-Universität Frankfurt am Main
Max-von-Laue-Straße 7, 60438 Frankfurt am Main (Germany)
E-mail: matthias.wagner@chemie.uni-frankfurt.de

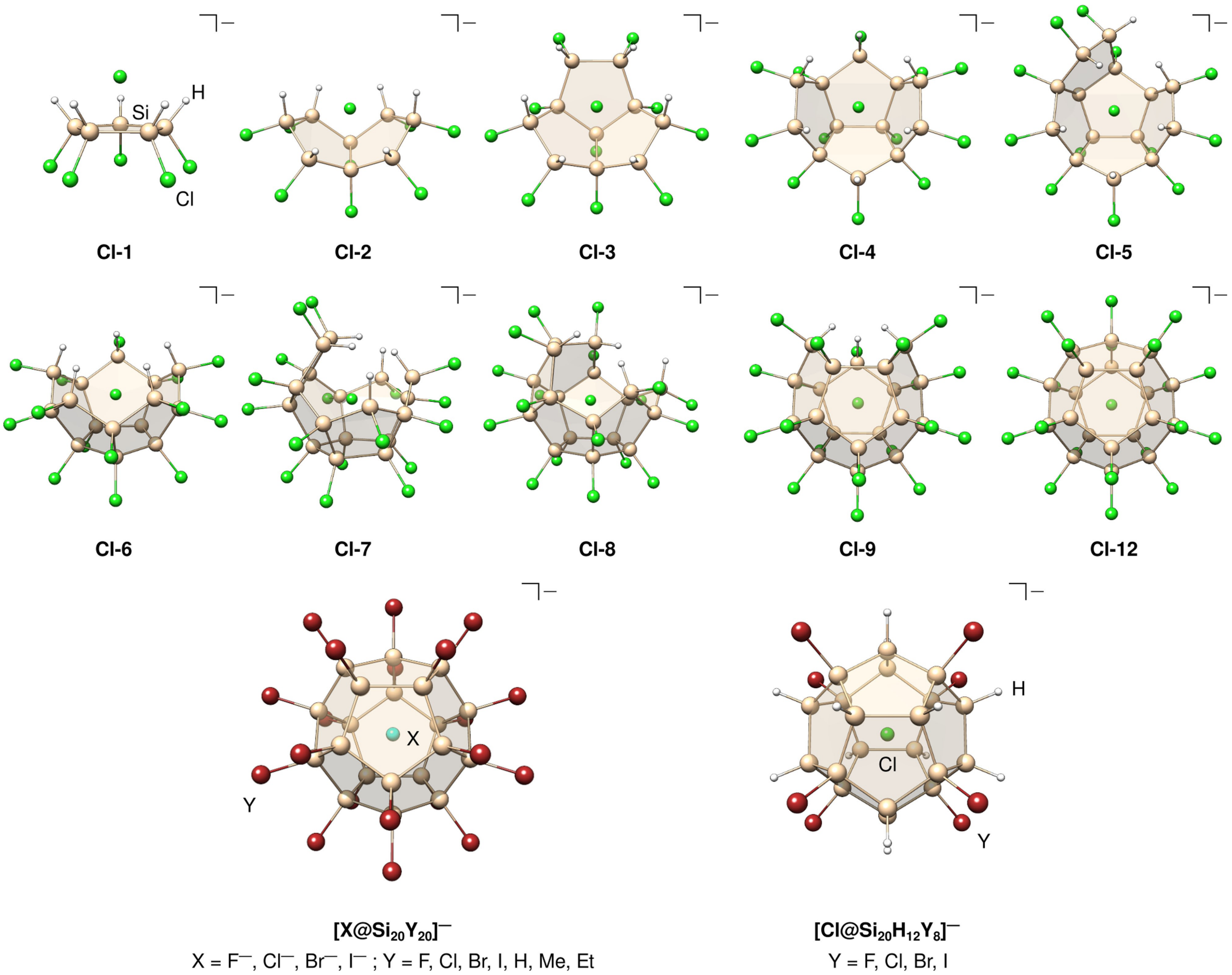
Dr. M. Bursch
Max-Planck-Institut für Kohlenforschung
Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr (Germany)
E-mail: bursch@kofo.mpg.de

[+] These authors contributed equally.

[**A**]$^-$ contains a dodecahedral $Si_{20}$ core with endohedral $Cl^-$ guest and is exohedrally saturated by 12 $SiCl_3$ and 8 Cl substituents, hence following design principle (iii).[43] Recently, [**A**]$^-$ was converted into the $T_h$-symmetric, mixed-substituted $[Cl@Si_{20}H_{12}Cl_8]^-$ and into a variety of $[Cl@Si_{20}Y_{20}]^-$ derivatives ([$n$Bu$_4$N][**B**]; Y = H, Cl; Figure 1).[44] These experimental results proved that $Si_{20}Y_{20}$ cages, which according to calculations of Goedecker *et al.* (Y = H) should be hardly accessible,[45] can be made accessible by $Cl^-$ encapsulation. This result immediately raises the question of whether the $Si_{20}$ dodecahedron is assembled around the $Cl^-$ ion and to what extent this endohedral guest exerts a structure-determining template effect. In this context, the nature and the degree of the $Cl^- \rightarrow Si_n$ interaction are of prime importance. Theoretical investigations by Holthausen *et al.* revealed that the formation of [**A**]$^-$ by formal incorporation of $Cl^-$ into the hypothetical empty cluster is exergonic by $>100$ kcal mol$^{-1}$.[43] Ponce-Vargas and Muñoz-Castro contributed an energy decomposition analysis (EDA) of the $Cl^- \rightarrow Si_{20}$ interaction in [**A**]$^-$ that revealed a mainly electrostatic character, which we also observed in our previous publication.[44,46]

Herein, we present a combined theoretical and experimental investigation of the cooperative effects between the endohedral ion and the exohedral decoration in [20]silafulleranes $[X@Si_{20}Y_{20}]^-$ (Figure 2). To ensure reliable results from our quantum mechanical calculations,[47–52] we assessed various density functional approximations on high-level reference data beforehand. From a theoretical perspective, we evaluate the central interaction between the endohedral ion $X^-$ and the empty host $Si_{20}Y_{20}$. This is done for (i) small cluster fragments composed of annulated five-membered rings ("silapolyquinanes") and (ii) for $[X@Si_{20}Y_{20}]^-$ clusters with different endohedral ions ($X^-$) and exohedral substituents (Y). These calculations not only shed light on the role of the $Cl^-$ ion as a structure-determining template during cluster buildup but also provide a first insight into the thermodynamics of cluster degradation.

Further, we present a thorough NMR study based on theoretical predictions of $^{35}$Cl chemical shifts. We investigate parallels to the well-known Gutmann–Beckett method,



**Cl-1** **Cl-2** **Cl-3** **Cl-4** **Cl-5**

**Cl-6** **Cl-7** **Cl-8** **Cl-9** **Cl-12**

**$[X@Si_{20}Y_{20}]^-$**
X = F$^-$, Cl$^-$, Br$^-$, I$^-$ ; Y = F, Cl, Br, I, H, Me, Et

**$[Cl@Si_{20}H_{12}Y_8]^-$**
Y = F, Cl, Br, I

***Figure 2.*** Computationally optimized structures (PBEh-3c(SMD($CH_2Cl_2$))) of the systems relevant for the discussion.

which is used to evaluate the effective Lewis acidity of compounds by the $^{31}$P NMR chemical shift of a coordinated $Et_3PO$ sensor ligand.[53,54] For this, we compare different properties, such as the interaction energy or the donor-acceptor gap, with the respective $^{35}$Cl NMR chemical shift of the silafullerane-encapsulated $Cl^-$ ion. On the experimental side, we expand the range of synthetically available derivatives by providing the syntheses of $[Cl@Si_{20}Y_{20}]^-$ (Y = Me, Et, Br) to ultimately include systems that distribute the negative charge over an even larger volume than $[Cl@Si_{20}Cl_{20}]^-$ and those that have a lipophilic shell.

## Results and Discussion

### Template Effect in the Cluster Assembly

Perhalogenated cyclopentasilanes $Si_5X_{10}$ and cyclohexasilanes $Si_6X_{12}$ are known to be ditopic Lewis acids (LA) that coordinate Lewis bases (LB) such as nitriles (RCN) or halide ions in a $\mu_5$ ($Si_5X_{10}\cdot 2LB$) or $\mu_6$ ($Si_6X_{12}\cdot 2LB$) mode on both sides of the respective ring to form inverse sandwich complexes.[55–64] A similar interaction is found for the halide encapsulating $[X@Si_{20}Y_{20}]^-$ clusters, where the dissociation of the encapsulated ion is inhibited by the surrounding $Si_{20}$ cage. Accordingly, we classify our silafulleranes as "confined Lewis pairs" with the $Si_{20}$ cage being the Lewis acid and the endohedral ion ($X^-$) the Lewis base. The confinement of a Lewis base inside a molecular cage has been applied in frustrated Lewis pair (FLP) chemistry, but in those cases, the cage itself did not act as a Lewis acid.[65,66] The classification as "confined Lewis pair" is helpful to distinguish our endohedral silafulleranes from non-confined Lewis pairs such as cyclosilane-base adducts. In the following, we will investigate the intrinsic interactions and NMR properties of this newly introduced compound class of confined Lewis pairs.

The mechanism underlying the one-step synthesis of $[Cl@Si_{20}(SiCl_3)_{12}Cl_8]^-$ from $Si_2Cl_6$ poses a fundamentally interesting but so far unsolved problem. Since, to our best knowledge, no intermediates of the cluster assembly are known, we apply QM computations to gain some insights into the process of cluster assembly. Our central working hypothesis assigns a structure-directing template effect to the $Cl^-$ ion that is ultimately trapped inside the cluster[43] (cf. Goedecker's conclusion that an empty $Si_{20}H_{20}$ siladodecahedrane will not form spontaneously – despite its predicted thermodynamic stability – due to Levinthal's paradox).[45] The siladodecahedrane consists exclusively of five-membered rings, and it has already been theoretically established that cyclopentasilanes are early intermediates of oligosilane "Aufbau" reactions occurring in $Si_2Cl_6/Cl^-$ mixtures.[55] Thus, we first investigated the energetic landscape of adducts between one $Cl^-$ ion and a successively growing concave framework of $m$ mutually annulated cyclopentasilane rings. For these adducts, we will use the nomenclature "**Cl-*m***" (Figure 2; $m = 1$–12 $Si_5$ rings; note that **Cl-10** and **Cl-11** do not exist due to the simultaneous formation of three five-membered rings in the final step leading from **Cl-9** directly to **Cl-12**). Geometries of each adduct have been fully optimized. Generally, we used the PBEh-3c[67]-(SMD($CH_2Cl_2$))[68] level of theory for the optimization of all geometries in this work. At each stage, the emerging silapolyquinane[69] scaffolds represent growing fragments of the $Si_{20}Y_{20}$ cluster (Kyushin's permethylated decasilahexahydrotriquinacene[70] is the only experimentally accessible bowl-shaped oligosilane yet known). Specifically, we began with the $[Si_5H_5Cl_5\cdot Cl]^-$ adduct **Cl-1**, in which the $Cl^-$ anion is placed above the perhydrogenated face of the ring. For two reasons, the positions serving as anchor points for the subsequent annulation steps were saturated by H atoms: (i) H substituents have the smallest possible steric demand and (ii) the difference in electronegativity is smaller between H and Si than between Cl and Si. Although different reaction mechanisms could be possible, we focus on one of the structurally most related mechanisms. Nevertheless, it should not be taken as a proposed reaction mechanism.

For each silapolyquinane (**Cl-*m***) we modeled for the cluster assembly series, we performed a local energy decomposition (LED)[71–73] analysis at the DLPNO-CCSD(T)[74–78]/*TightPNO*/def2-TZVPP level of theory to understand the changes in the interaction energy composition upon subsequent annulation (cf. Figure 3). The LED gives rise to different contributions of the total $Cl^-\rightarrow Si_n$ interaction energy ($E_{int}$) and contains, e.g., the electronic preparation energy at the Hartree–Fock (HF) level ($E_{HF-elprep}$) that describes the repulsive part of the exchange interaction and can therefore be identified as "Pauli repulsion". Further components describe the electrostatic interaction ($E_{elstat}$), attractive exchange energy ($E_{exch}$), orbital-relaxation ($E_{orb-relax}$), London dispersion ($E_{disp}$) and non-dispersion ($E_{non-disp}$) effects. The $E_{non-disp}$ term is added as a correction to errors in the permanent electrostatic interactions inherent to the HF method. The total interaction energy is decomposed according to Eq. (1). Each contribution is computed as the difference (Δ) between the relaxed complex and the sum of the unrelaxed $Si_n$ fragment and the endohedral ion $X^-$.

$$\begin{aligned} E_{int} &= \Delta E_{HF-elprep} + \Delta E_{elstat} \\ &\quad + \Delta E_{exch} + \Delta E_{orb-relax} + \Delta E_{disp} \\ &\quad + \Delta E_{non-disp} \end{aligned} \tag{1}$$

Overall, the total $Cl^-\rightarrow Si_n$ interaction steadily becomes stronger as the silapolyquinanes grow in size, which underlines the postulated template effect of the $Cl^-$ ion (cf. Figure 3a). The respective interaction energies range from −78.3 (**Cl-1**) to −146.6 kcal mol$^{-1}$ (**Cl-12**). Electron correlation contributes significantly to the interaction energy and is therefore indispensable for reliable results (see the difference between DLPNO-CCSD(T) and HF in Figure 3a). The electron correlation describes the sum of the non-dispersion and London dispersion terms found in the LED. As the London dispersion term is larger, it has a bigger influence on the overall interaction (see Supporting Information, Figure S1). The average distance between the endohedral

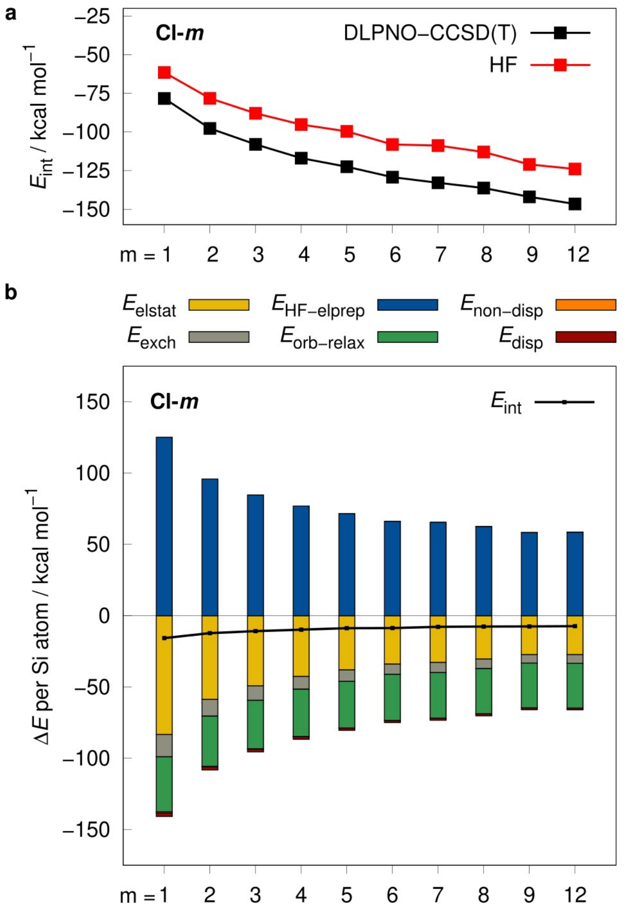


***Figure 3.*** (a) Computed Cl→$\text{Si}_n$ interaction energy applying DLPNO-CCSD(T)/*TightPNO*/def2-TZVPP and HF/def2-TZVPP. (b) LED per Si atom for the cluster assembly series of $[\text{Cl@Si}_{20}\text{Cl}_{20}]^-$ (**Cl-*m***) computed with DLPNO-CCSD(T)/*TightPNO*/def2-TZVPP. Contributions of $E_{\text{non-disp}}$ are in a range between −0.83 and 0.16 kcal mol$^{-1}$ and therefore not visible.

$\text{Cl}^-$ ion and the centers of the five-membered Si rings increases with the size of the silapolyquinanes and it therefore correlates with the interaction strength (Table 1). In the same series, the bond lengths between the Si atoms and the exohedral Cl substituents become subsequently smaller, and the positive partial charge of the Si atoms is reduced. As the silapolyquinanes grow in size, there are more available interaction sites, which leads to an increased interaction strength. A more detailed overview of the interaction can be obtained by analyzing each contribution with respect to the number of Si atoms $n$ in the molecule (cf. Figure 3b). With increasing cluster size, each contribution to $E_{\text{int}}$ per Si atom is decreasing as the $\text{Cl}^-\rightarrow\text{Si}_n$ interaction is distributed among more atom pairs. This indicates the importance of many-body effects for the overall interaction. London dispersion and non-dispersion effects are very small compared to the remaining contributions and are almost negligible. Contributions from the attractive exchange interaction are slightly larger but still small compared to the remaining contributions. During the cluster assembly, the orbital-relaxation per Si atom is the only contribution that is almost constant as each silapolyquinane has a similar orbital overlap between the endohedral Lewis base and the Si center due to high symmetry. The main repulsive interaction is the Pauli repulsion, while the main attractive interactions are orbital-relaxation effects and electrostatics. This is in line with the previous theoretical study on $[\text{X@Si}_{32}\text{Cl}_{44}]^-$ by Muñoz-Castro *et al.*, which points out the dominant electrostatic character of the $\text{Cl}^-\rightarrow\text{Si}_{20}$ interaction.[46]

### *Interplay of Endohedral Ion and Exohedral Substitution*

In the next sections, we will discuss the influence of different endohedral ions and exohedral substituents on the electronic $\text{X}^-\rightarrow\text{Si}_{20}$ interaction. For this, we modeled $[\text{X@Si}_{20}\text{Y}_{20}]^-$ clusters with the endohedral ions X=F, Cl, Br, I and the exohedral substituents Y=F, Cl, Br, I, H, Me, Et (Figure 2). Additionally, we included the partially substituted $[\text{Cl@Si}_{20}\text{H}_{12}\text{Y}_8]^-$, Y=F, Cl, Br, I clusters.

The $\text{X}^-\rightarrow\text{Si}_{20}$ interaction is the most important one in our tested systems. It is given as the electronic energy difference between the $[\text{X@Si}_{20}\text{Y}_{20}]^-$ cluster and the sum of the unrelaxed fragments $\text{X}^-$ and $\text{Si}_{20}\text{Y}_{20}$. We computed the corresponding interaction energies at the most accurate level of theory, which was still affordable given the complexity

***Table 1:*** Average distances between the endohedral $\text{Cl}^-$ ion and the centers of the five-membered Si rings (c($\text{Si}_5$)) as well as the average distances between the Si atoms and the exohedral Cl substituents, and the average of natural charges in the molecules **Cl-*m***.

| | av. distances/Å | | av. natural charges/e$^-$ | | |
|---|---|---|---|---|---|
| | $d(\text{Cl}^-\cdots\text{c}(\text{Si}_5))$ | $d(\text{Si–Cl})$ | $\text{Cl}^-$ | Si | Cl |
| **Cl-1** | 2.1473 | 2.1097 | −0.531 | 0.467 | −0.416 |
| **Cl-2** | 2.2740 | 2.0955 | −0.497 | 0.432 | −0.387 |
| **Cl-3** | 2.3383 | 2.0895 | −0.486 | 0.409 | −0.373 |
| **Cl-4** | 2.3914 | 2.0859 | −0.479 | 0.392 | −0.364 |
| **Cl-5** | 2.4532 | 2.0360 | −0.472 | 0.381 | −0.357 |
| **Cl-6** | 2.5090 | 2.0808 | −0.484 | 0.365 | −0.351 |
| **Cl-7** | 2.5111 | 2.0800 | −0.461 | 0.358 | −0.347 |
| **Cl-8** | 2.5641 | 2.0784 | −0.468 | 0.346 | −0.343 |
| **Cl-9** | 2.6269 | 2.0764 | −0.486 | 0.335 | −0.339 |
| **Cl-12** | 2.6206 | 2.0742 | −0.468 | 0.307 | −0.333 |

and size of the systems (PNO-LCCSD(T)-F12b[79–81]/AVTZ'/default including scaled triples contributions to approximate PNO-LCCSD(T)-F12b/AVQZ'/tight results).[82–86] More details can be found in section 1.2 of the Supporting Information. Although estimated results with centered, endohedral $F^-$ ions are presented, it is worth mentioning that these compounds have imaginary modes up to 100 cm$^{-1}$ for the endohedral ion movement indicating unstable structures that cannot be isolated experimentally. Since we investigate the interaction between centered endohedral ions and the Si moieties, these structures still yield valuable information for the evaluation of [20]silafulleranes. We included them in our studies, but we will not focus on these compounds in the following.

For the mixed-substituted $[Cl@Si_{20}H_{12}Y_8]^-$ clusters, we observe no significant change in $E_{int}$ with different exohedral halogen substituents Y and will therefore omit them in the following energy discussion (cf. Supporting Information, Figure S2). In the fully substituted derivatives $[X@Si_{20}Y_{20}]^-$, small $E_{int}$ were computed for systems with Y=H, Me, Et, while large $E_{int}$ were computed for Y=Halogen (cf. Figure 4). As described in the previous section, the main attractive contribution to the $X^- \rightarrow Si_{20}$ interaction is of electrostatic origin. Therefore, the largest $E_{int}$ is expected for a system with a hard Lewis base in the endohedral position and an electron-poor $Si_{20}$ cage, which can be achieved, e.g., with inductively electron-withdrawing halogen substituents. Based on our calculations we predict the weakest $E_{int}$ for $[I@Si_{20}Et_{20}]^-$ and the strongest $E_{int}$ for $[F@Si_{20}F_{20}]^-$ within our test set. Since $[F@Si_{20}F_{20}]^-$ is no minimum structure, $[Cl@Si_{20}F_{20}]^-$ can be regarded as having the strongest interaction energy in our test set.

The interaction energy ($E_{int}$) only consists of the electronic energy ($E_{el}$), but experimentally measured properties also include solvation ($\delta G_{solv}$) and thermostatistical contributions ($G_{RRHO}$) which we computed with COSMO-RS[87–90] and the thermo submodule of XTB,[91] respectively. These corrections can be directly added to the electronic energy which leads to the total Gibbs free energy ($G$) according to Eq. (2). Each contribution is calculated as the difference ($\Delta$) between the complex, the $Si_{20}$ fragment, and the endohedral ion. In contrast to the interaction energy, each contribution is computed for fully optimized geometries and therefore it also includes geometrical relaxation effects.

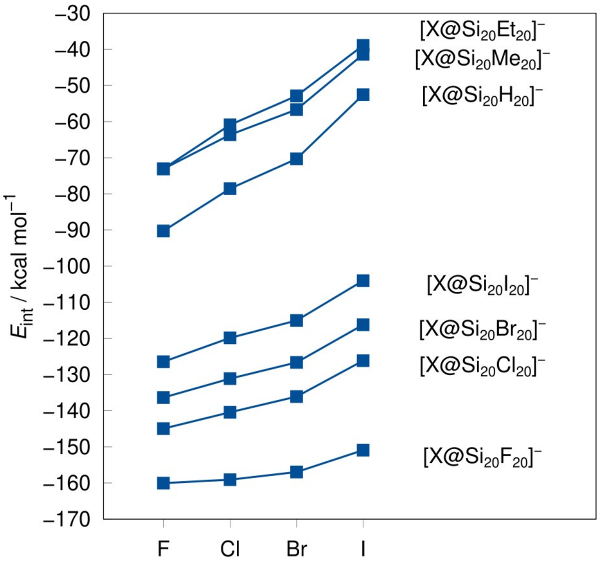


***Figure 4.*** Computed interaction energies ($E_{int}$) for the $X^- \rightarrow Si_{20}$ bonds at PNO-LCCSD(T)-F12b/AVTZ'/default level including scaled triples contributions. $[X@Si_{20}Et_{20}]^-$ was computed with our best-performing density functional approximation (cf. Supporting Information).

$$\Delta G = \Delta E_{el} + \Delta G_{RRHO} + \Delta\delta G_{solv} \tag{2}$$

As different QM methods with varying accuracy can be applied for $E_{el}$, it is important to validate them beforehand. Based on our previously discussed reference values, we assessed different density functional approximations[92–110] and we selected the hybrid density functional r$^2$SCAN0-D4 [94,95,102] in combination with the def2-QZVPPD[111] (on $X^-$), def2-TZVPPD (on Si), and def2-QZVPP[112] (on Y) basis sets for all further discussed results. More details can be found in the Computational Details section of the Supporting Information.

Due to the strong Si–Si cage bonds, it is practically impossible to chemically extract the endohedral ions from the $Si_{20}$ cages, i.e., to separate the confined Lewis pair. Nevertheless, we can theoretically predict association free energies ($\Delta G$) for this process (cf. Figure 5). Since the main difference between $\Delta G$ and $E_{int}$ is just the solvation and thermostatistical contributions, the main trends for both are similar: systems with a hard Lewis base as endohedral ion and electron-withdrawing substituents are generally more stable and systems with positive inductive substituents are less stable. Solvation and thermostatistical contributions counteract the attractive electronic interaction and thus can also cause thermodynamically unstable structures. The most stable structures investigated are $[Br@Si_{20}F_{20}]^-$ ($\Delta G = -91.4$ kcal mol$^{-1}$) and $[Cl@Si_{20}F_{20}]^-$ ($\Delta G = -89.6$ kcal mol$^{-1}$), while the most unstable structure is $[I@Si_{20}Et_{20}]^-$ ($\Delta G = 13.2$ kcal mol$^{-1}$) among the investigated systems. Overall, solvation and thermostatistics have a large impact on the free energy of association in [20]silafulleranes with endohedral ions.

By encapsulation of an endohedral ion, the $Si_{20}Y_{20}$ fragment of the adduct will be distorted compared to an empty $Si_{20}Y_{20}$ cage; we term the corresponding energies $E_{cage,adduct}$ and $E_{cage,relaxed}$, respectively. We computed the geometric change in cluster width as well as cage strain ($E_{strain}$) in each investigated complex (cf. Figure 6) according to Eq. (3).

$$E_{strain} = E_{cage,adduct} - E_{cage,relaxed} \tag{3}$$

In most systems, the $Si_{20}$ cage is contracted upon encapsulation of an endohedral ion due to the attractive $X^- \rightarrow Si_{20}$ interaction.

Only for some combinations of the larger endohedral ions $Br^-$ and $I^-$ with less electron-withdrawing cluster

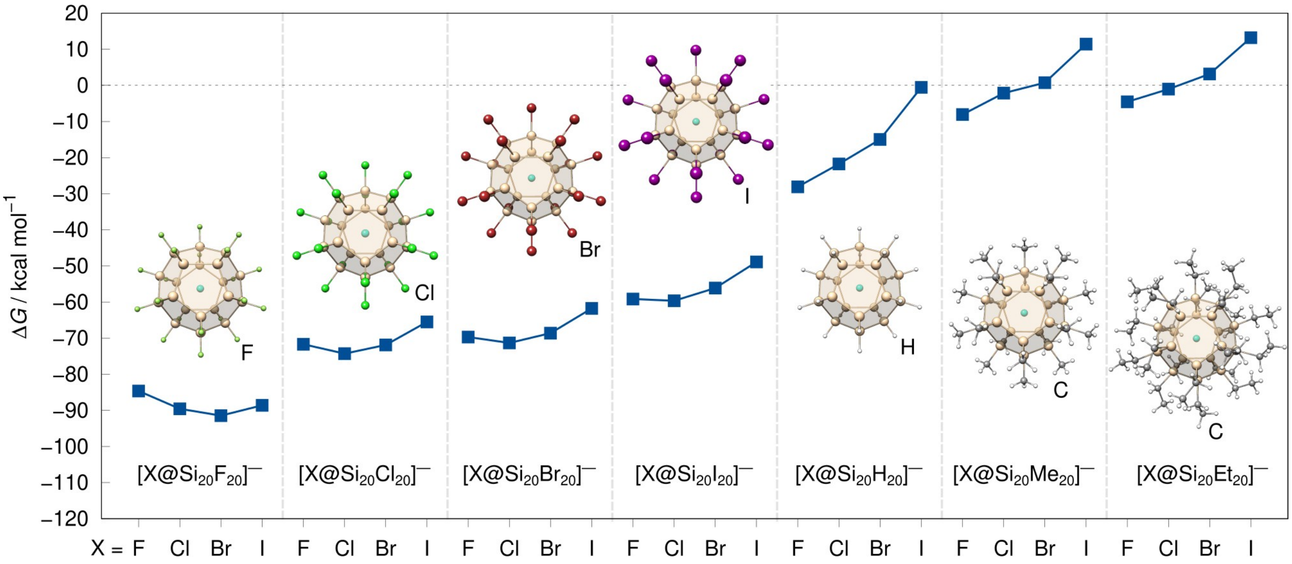


***Figure 5.*** Gibbs free energies ($\Delta G$) for the encapsulation of the respective endohedral ion in $Si_{20}$ cages with different substituents computed with our best-performing DFT method (see Supporting Information).

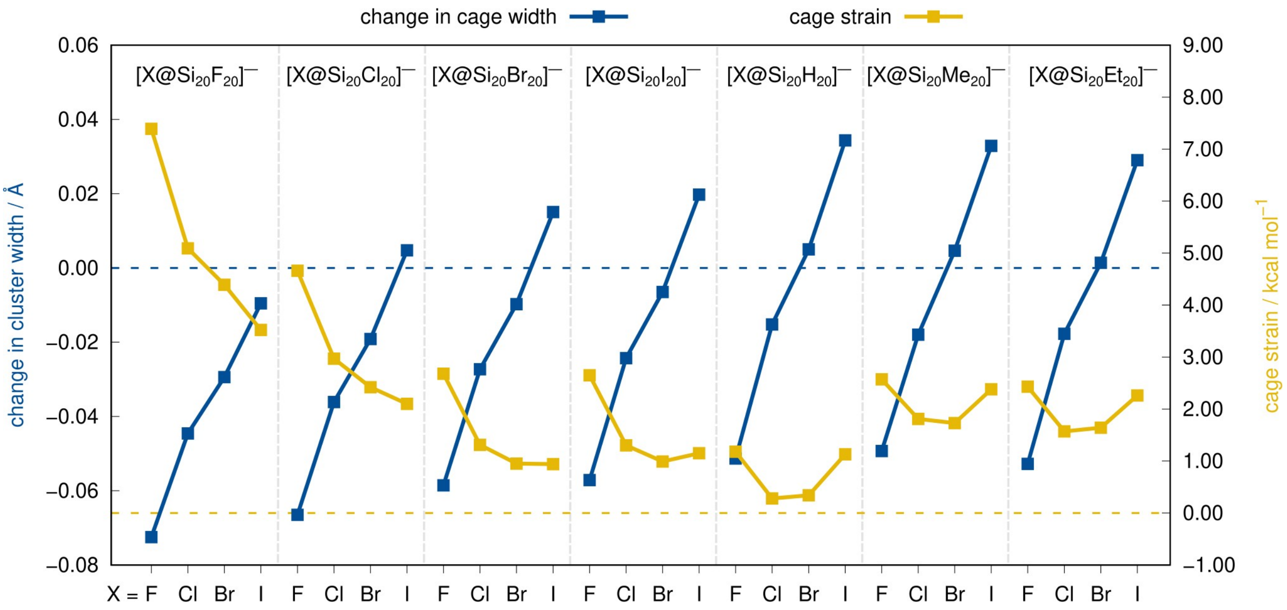


***Figure 6.*** Change in cage width between the $Si_{20}$ cluster fragment in $[X@Si_{20}Y_{20}]^-$ and the empty cage $Si_{20}Y_{20}$ (blue) and the corresponding cage strain (yellow) computed with our best-performing DFT method (see Supporting Information).

substituents, we observe an expansion of the cage (cf. $[I@Si_{20}Me_{20}]^-$). Generally, the cage strain correlates inversely with cage contractions and directly with cage expansions. The most contracted system is $[Cl@Si_{20}F_{20}]^-$ with a cage strain of 5.1 kcal mol$^{-1}$ and the most expanded system is $[I@Si_{20}H_{20}]^-$ with a cage strain of only 1.1 kcal mol$^{-1}$.

To further investigate the effect of steric bulk on the $X^- \rightarrow Si_{20}$ interaction in $[X@Si_{20}Y_{20}]^-$, we conducted an energy decomposition analysis (EDA) at the B3LYP–D4/def2-QZVPPD level of theory for selected clusters (cf. Figure 7).[113] The EDA is similar to the previously discussed LED analysis but is computationally less demanding and can therefore also be applied to larger structures. The EDA is based on density functional theory instead of wavefunction theory, which is used in the LED. This results in slightly different contributions to the total interaction energy. The EDA includes a repulsion ($E_{rep}$), electrostatic ($E_{elstat}$), exchange ($E_{exch}$), orbital-relaxation ($E_{orb-relax}$), electron correlation ($E_{correl}$), and dispersion ($E_{disp}$) contributions to the total energy.

The influence of the exohedral substituent on the $Cl^- \rightarrow Si_{20}$ interaction is significantly smaller than that of the endohedral ion. The less electron-withdrawing exohedral substituents (Y = H, Me, Et) cause a stronger repulsion and

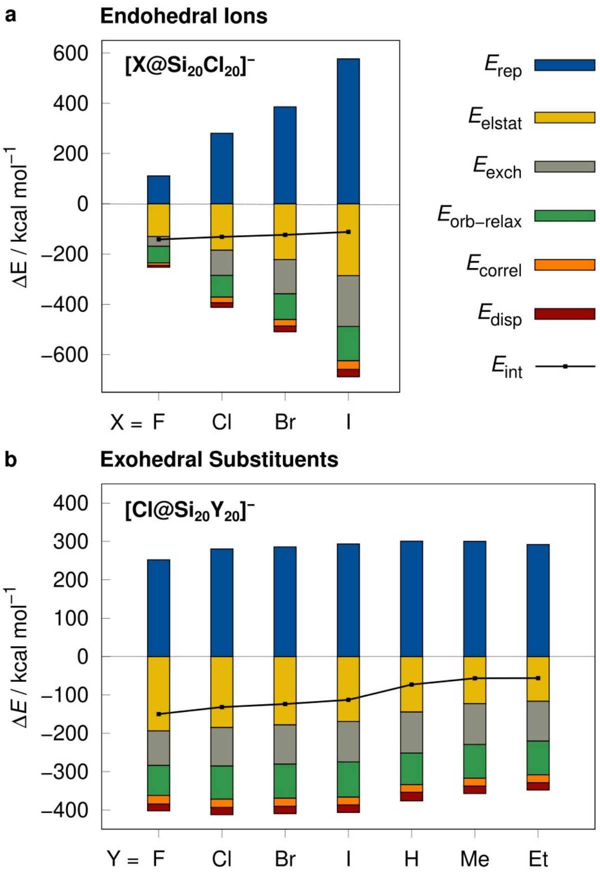


***Figure 7.*** EDA for complexes with (a) different endohedral ions $[X@Si_{20}Cl_{20}]^-$ and (b) exohedral substituents $[Cl@Si_{20}Y_{20}]^-$.

a less attractive electrostatic interaction leading to an overall lower total interaction energy. The repulsion is slightly lower for the electron-withdrawing substituents (Y=F, Cl, Br, I) leading to an overall more attractive interaction.

As the endohedral ion is directly involved in the investigated interaction, the influence of the nature of $X^-$ on the individual contributions to $E_{int}$ is much larger than that of Y. For example, going from $F^-$ ($E_{rep}=111.0$ kcal mol$^{-1}$) to $Cl^-$ ($E_{rep}=280.6$ kcal mol$^{-1}$), the repulsion energy is more than twice as large. All EDA contributions are increasing as the endohedral ion becomes larger, whereas the total interaction energy decreases. Since the increase in repulsion is larger than the gain in attractive contributions, there is a net unfavorable effect of larger endohedral ions on $E_{int}$.

In order to analyze the effect of charge transfer in our tested systems, we computed deformation density differences from Natural Orbitals for Chemical Valence (NOCV)[114,115] and applied the Extended Transition State (ETS)[116] method at the same level of theory as applied in our electronic energy calculations to obtain corresponding energy contributions ($\Delta E_k$, cf. Figure 8).[117] Since the nature of the donor and acceptor orbitals is similar for all systems investigated, we only discuss the results for the $[Cl@Si_{20}Cl_{20}]^-$ cluster exemplarily. Charge transfer mainly takes place by electron-density donation from the $p_x$, $p_y$, and $p_z$ orbitals of the endohedral $Cl^-$ to the $\sigma^*$(Si–Cl) orbitals of the exohedral Cl substituents which are located in the same axis. The charge transfer contribution of each $p$ orbital is $-19.1$ kcal mol$^{-1}$ ($\Delta\rho1$, $\Delta\rho2$, $\Delta\rho3$). We also observe a fourth contribution coming from the $s$ orbital of the endohedral $Cl^-$, which donates electron density to the $\sigma^*$ orbitals of the exohedrally located chlorine atoms, but this contribution is significantly smaller ($\Delta\rho4$, $\Delta E_k=-5.1$ kcal mol$^{-1}$). These findings are in line with prior studies on [20]silafulleranes.[118]

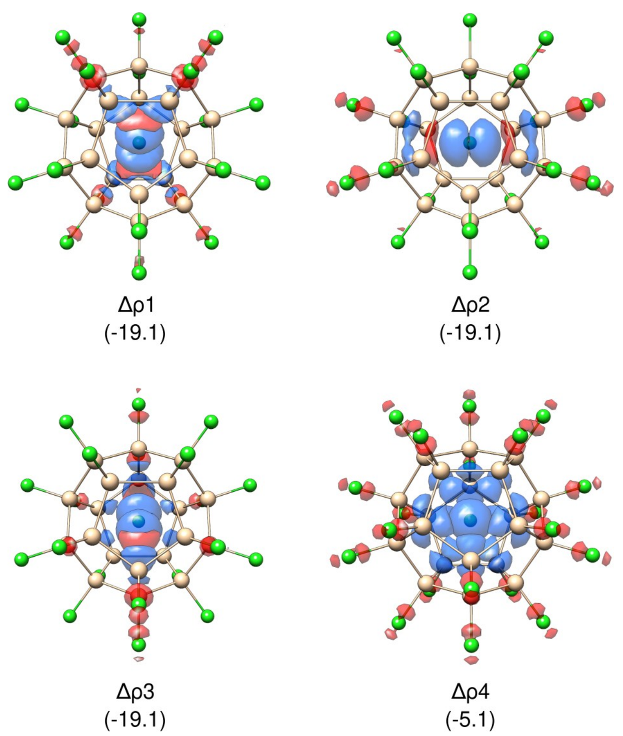

Δρ1
(-19.1)
Δρ2
(-19.1)
Δρ3
(-19.1)
Δρ4
(-5.1)

***Figure 8.*** Deformation density difference ($\Delta\rho$) of the main contributions to the charge transfer found in $[Cl@Si_{20}Cl_{20}]^-$ and the respective energy contributions ($\Delta E_k$) in parenthesis [kcal mol$^{-1}$]. Blue represents depletion and red represents expansion of density. An isosurface value of 0.0015 a.u. was applied.

### NMR Chemical Shifts

In our previous publication on silafulleranes $[Cl@Si_{20}Y_{20}]^-$,[44] we have already found some evidence that experimentally determined $^{35}$Cl NMR chemical shift values are diagnostic for a number of important properties, such as the $Cl^- \rightarrow Si_{20}$ interaction energies or the main donor-acceptor orbital gaps. Back then, however, these studies had to be limited to the three derivatives that were synthetically accessible at the time. Since a much wider palette of silafulleranes is now available, we revisit the topic in the following and confirm the correlations between $\delta(^{35}Cl)$ and key silafullerane features on a broader basis. A new aspect arises from our classification of endohedral silafulleranes $[Cl@Si_{20}Y_{20}]^-$ as confined Lewis pairs: if we take the $Cl^- \rightarrow Si_{20}$ interactions as a measure of the Lewis acidities of different empty siladodecahedranes $Si_{20}Y_{20}$, their quantification by $^{35}$Cl NMR spectroscopy is reminiscent of the well-established method of Gutmann and Beckett.[53,54] They used the difference $\Delta\delta(^{31}P)$ between $\delta(^{31}P)$ of the free sensor Lewis base $Et_3PO$ and $\delta(^{31}P)$ of the adduct $Et_3PO\rightarrow$LA to assess the acidity of a given Lewis acid LA. Generally speaking, the computed chemical shift $\delta$ of a given nucleus (here: $^{35}$Cl) is the difference between the total isotropic shielding of this nucleus in a reference compound (here: $CH_2Cl_2$; $\sigma_{tot,ref}$) and in the compound under study ($\sigma_{tot,comp}$; Eq. (4)). Each total shielding ($\sigma_{tot}$) is the sum of diamagnetic ($\sigma_{dia}$), paramagnetic ($\sigma_{para}$), and spin-orbit contributions ($\sigma_{so}$) (Eq. (5)).[119,120]

$$\delta = \sigma_{tot,ref} - \sigma_{tot,comp} \qquad (4)$$

$$\sigma_{tot} = \sigma_{dia} + \sigma_{para} + \sigma_{so} \qquad (5)$$

The diamagnetic contributions are derived from the unperturbed electron density and thus mainly depend on the ground state density. The paramagnetic shielding constant is related to frontier orbital transitions which renders it highly responsive to the chemical environment, including bonding and other interactions.[121] For the computation of the $^{35}$Cl NMR chemical shieldings we applied the SO-ZORA[122–124] -PBE0/TZP[125] level of theory and applied the COSMO solvation model for $CH_2Cl_2$.[126] First, we emphasize that the theoretically and experimentally obtained $\delta(^{35}Cl)$ values of all our silafulleranes $[Cl@Si_{20}Y_{20}]^-$ are in agreement with each other; the subsequent discussion refers exclusively to calculated chemical shift data. As shown in Figure 9, the $\sigma_{tot}$ values (and hence $\delta(^{35}Cl)$) of $[Cl@Si_{20}Y_{20}]^-$ are mainly determined by $\sigma_{para}$, which is much more influenced by the nature of Y than $\sigma_{dia}$ and $\sigma_{so}$. A recent theoretical analysis of the Gutmann–Beckett method by Greb *et al.* revealed that the same is true for $\delta(^{31}P)$ in $Et_3PO\rightarrow$LA.[127] Nevertheless, the trends in $\delta(^{31}P)$ and $\delta(^{35}Cl)$ go in opposite directions with increasing strength of the Lewis acid: While stronger LAs cause larger *downfield* shifts of the $^{31}$P resonances of corresponding $Et_3PO\rightarrow$LA adducts, the $^{35}$Cl resonances of silafulleranes $[Cl@Si_{20}Y_{20}]^-$ become more *upfield* shifted the more electron-withdrawing the Y substituents are and the stronger the $Cl^-\rightarrow Si_{20}$ interaction energy $E_{int}$ becomes (cf. Figure 10). We note in passing that this inverse correlation between $\delta(^{35}Cl)$ and $E_{int}$ is most pronounced for Y=H, F-I, but also holds for the subsets $[Cl@Si_{20}H_{20}]^-\rightarrow[Cl@Si_{20}H_{12}Y_8]^-\rightarrow[Cl@Si_{20}Y_{20}]^-$ (Y=F-I). In contrast, the less electron-withdrawing substituents Y=Me, Et do not yet lead to a clear trend and require more data points to make robust statements. In addition to the values for the complete silafulleranes $[Cl@Si_{20}Y_{20}]^-$, we have also computed $\delta(^{35}Cl)$ and $E_{int}$ for the non-spherical model systems **Cl-1** to **Cl-9** (Y=Cl; Figure 2) and the analogous subsets with Y=H, F, Br, I. For each subset, we can extrapolate trajectories starting at **Y-1** and approaching the values of the fully assembled $[Cl@Si_{20}Y_{20}]^-$ clusters as the model systems grow bigger (Figure 10). For Y=H, Cl, Br, I, we now see that an increasing $E_{int}$ results in increasingly *deshielded* $^{35}$Cl nuclei – analogous to the Gutmann–Beckett scale and inverse to the trend observed for $[Cl@Si_{20}Y_{20}]^-$.[128] Taken together, these results lead to the important conclusion that there does not seem to be a generally valid correlation between $\delta(^{35}Cl)$ and $E_{int}$ for all pairs of Lewis acidic oligosilanes and $Cl^-$ ions. However, for well-defined subsets, such as either $[Cl@Si_{20}Y_{20}]^-$ or **Cl-1** to **Cl-12**, corresponding correlations do exist and are of considerable diagnostic and predictive value.

According to the Ramsey equation, $\sigma_{para}$ is proportional to the negative inverse of the energy gaps in pairs of magnetically coupled occupied and vacant orbitals.[129,130] For the silafulleranes $[Cl@Si_{20}Y_{20}]^-$ (Y=H, Me, Et, F-I), we conducted a Ditchfield decomposition analysis to reveal the

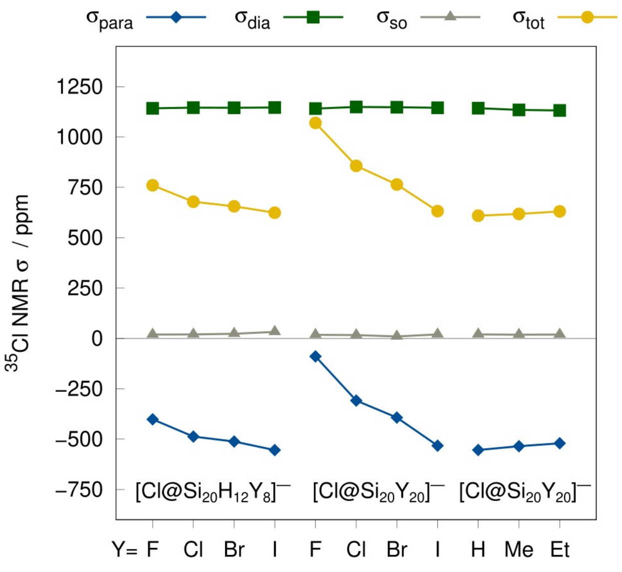


***Figure 9.*** Paramagnetic ($\sigma_{para}$), diamagnetic ($\sigma_{dia}$), and spin-orbit contributions ($\sigma_{so}$) to the total isotropic shielding $\sigma_{tot}$ of the $^{35}$Cl nucleus inside the respective $Si_{20}$ cluster.

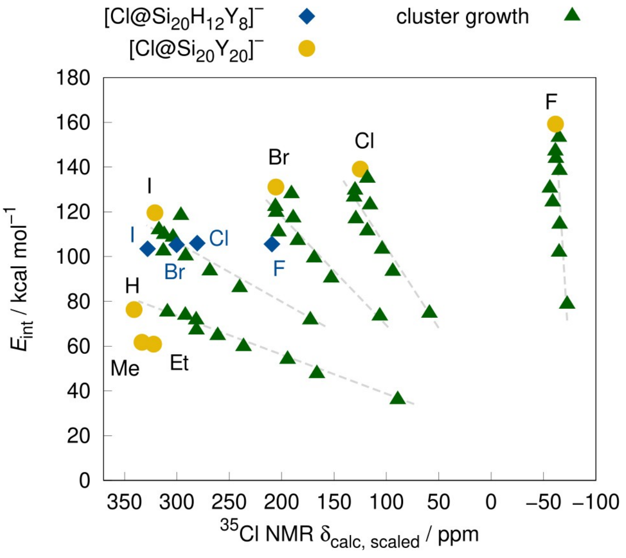


***Figure 10.*** Correlation between the interaction energy ($E_{int}$) and the calculated and linearly scaled $^{35}$Cl NMR chemical shift $\delta$.

contribution of each molecular orbital pair to the paramagnetic shielding.[131–133] In all cases, the main contributions to $\sigma_{para}$ come from excitations from the electron lone pairs of the endohedral $Cl^-$ ion into vacant orbitals above the LUMO. The corresponding donor-acceptor orbital gaps ($\Delta E_{DA}$) indeed show a good correlation with the paramagnetic $^{35}Cl$ shielding constants ($\sigma_{para}$; Figure 11). Consequently, and in agreement with the experiment, $\delta(^{35}Cl)$ is small when $\Delta E_{DA}$ is large because electronegative substituents Y lead to high $Cl^- \rightarrow Si_{20}$ interaction energies $E_{int}$. If, in a simplified approach, $\Delta E_{DA}$ is approximated by the HOMO–LUMO gap $\Delta E_{HL}$ of $[Cl@Si_{20}Y_{20}]^-$, the above-mentioned inverse correlation with $\sigma_{para}$ does no longer apply (Table 2). Thus, a reliable analysis of $\delta(^{35}Cl)$ inevitably requires the laborious consideration of all donor-acceptor orbital pairs because excitation from the HOMO to the LUMO is not the main factor. From our density-deformation computations, we also obtained the contributions of each natural orbital of chemical valence (NOCV) to $\delta(^{35}Cl)$. Along the series of exohedral substituents Y, the respective largest contribution ($\Delta E_k$) correlates with $\sigma_{para}$ (Y = F-I; Table 2). No obvious interdependence was observed between $\delta(^{35}Cl)$ and the $Cl^- \rightarrow Si_{20}Y_{20}$ charge transfer (CT) or the largest absolute principal component ($|V_{33}|$) of the electric field gradient (EFG), which governs the linewidth of the NMR resonance (Table 2).[134] In conclusion, $\delta(^{35}Cl)$ of our silafulleranes $[Cl@Si_{20}Y_{20}]^-$ correlates inversely with (i) the $Cl^- \rightarrow Si_{20}$ interaction energy $E_{int}$ and in turn the Lewis acidity of the empty $Si_{20}Y_{20}$ cage, and (ii) with the donor-acceptor orbital gaps $\Delta E_{DA}$, but shows no correlation with the $Cl^- \rightarrow Si_{20}Y_{20}$ CT.

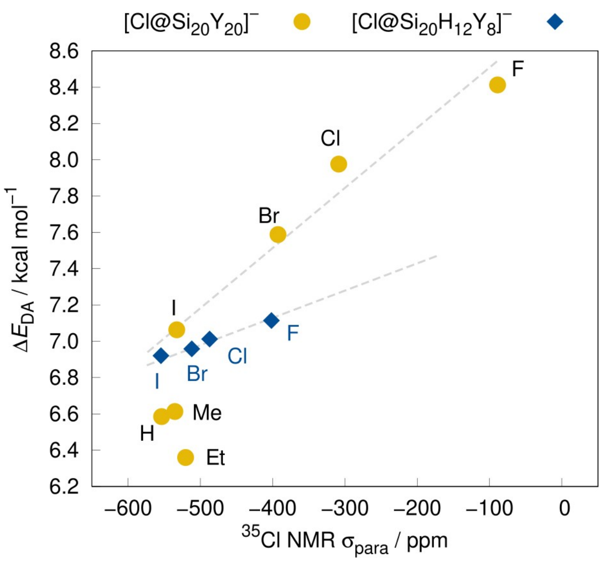


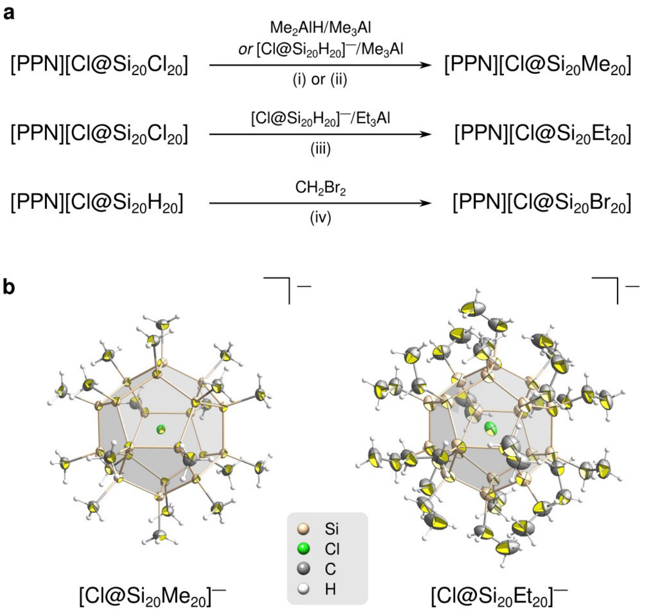
**Figure 11.** Correlation between the donor-acceptor gap ($\Delta E_{DA}$) and the paramagnetic $^{35}Cl$ NMR shielding constant $\sigma_{para}$.

### Introduction of Sterically Demanding Substituents

The primary starting material [$n$Bu$_4$N][**A**] is accessible in a one-pot self-assembly process from $Si_2Cl_6$, [$n$Bu$_4$N]Cl, and $n$Bu$_3$N in 27 % yield.[43] Twelvefold desilylation of [$n$Bu$_4$N][**A**] with pinacol, followed by treatment with $i$Bu$_2$AlH, furnishes [$n$Bu$_4$N][$Cl@Si_{20}H_{20}$].[44] Addition of [PPN]Cl to the reaction mixture gives [PPN][$Cl@Si_{20}H_{20}$] after crystallization ($[PPN]^+ = [(Ph_3P)_2N]^+$). Heating [PPN][$Cl@Si_{20}H_{20}$] in $CHCl_3$ affords [PPN][$Cl@Si_{20}Cl_{20}$]. For the methylation of [PPN][$Cl@Si_{20}Cl_{20}$], we adapted the recently described synthesis of [$n$Bu$_4$N][$Cl@Si_{20}(SiH_3)_{12}Me_8$],[134] which builds on regioselective Cl/H and Cl/Me exchange in a three-component reaction of [$n$Bu$_4$N][**A**] with $Me_2AlH$[135] and $Me_3Al$. The $Me_2AlH/Me_3Al$ system is capable of hydrogenating only the silyl groups of [**A**]$^-$, while the $Si_{20}$ core is selectively methylated.[134] In a similar vein, treatment of [PPN][$Cl@Si_{20}Cl_{20}$] with $Me_2AlH/Me_3Al$ (8/40 equiv.) resulted in the exhaustive methylation of the $Si_{20}$ cluster (Scheme 1a). [PPN][$Cl@Si_{20}Me_{20}$] was isolated by crystallization from THF in 24 % yield.[136] In the absence of $Me_2AlH$, a mixture of [PPN][$Cl@Si_{20}Cl_{20}$] and $Me_3Al$ (40 equiv.) remained unchanged at room temperature for at least 1 d. This result suggests that hydridic H substituents on Al or Si promote core methylation. This conclusion is supported by the fact that [PPN][$Cl@Si_{20}Me_{20}$] was also formed from a mixture of the chlorosilane [PPN][$Cl@Si_{20}Cl_{20}$] (1.0 equiv.), the hydridosilane [PPN][$Cl@Si_{20}H_{20}$] (0.50 equiv.), and $Me_3Al$ (60 equiv.). Following this approach, we also achieved perethylation of a [PPN][$Cl@Si_{20}Cl_{20}$]/[PPN][$Cl@Si_{20}H_{20}$] mixture (1.0/0.50 equiv.) with $Et_3Al$ (20 equiv.; Scheme 1a).

**Table 2:** Different calculated properties for the investigated systems.

| | $\delta^{35Cl}_{calc,scaled}$/ppm | $\sigma^{35Cl}_{para}$/ppm | EFG $\lvert V_{33} \rvert$/a.u. | $\Delta E_{HL}$/eV | $\Delta E_{DA}$/eV | $E_{int}$/kcal mol$^{-1}$ | CT/e$^{-1}$ | $\Delta E_k$/kcal mol$^{-1}$ |
|---|---|---|---|---|---|---|---|---|
| $[Cl@Si_{20}H_{12}F_8]^-$ | 209.3 | −401.7 | 0.00106 | 4.5 | 7.1 | 105.6 | 0.513 | −18.4 |
| $[Cl@Si_{20}H_{12}Cl_8]^-$ | 280.6 | −487.1 | 0.00037 | 4.6 | 7.0 | 106.0 | 0.492 | −18.9 |
| $[Cl@Si_{20}H_{12}Br_8]^-$ | 300.3 | −511.9 | 0.00115 | 4.4 | 7.0 | 105.3 | 0.489 | −19.1 |
| $[Cl@Si_{20}H_{12}I_8]^-$ | 328.0 | −554.4 | 0.00069 | 4.1 | 6.9 | 103.5 | 0.503 | −19.7 |
| $[Cl@Si_{20}F_{20}]^-$ | −61.4 | −88.9 | 0.00048 | 3.0 | 8.4 | 159.2 | 0.508 | −17.6 |
| $[Cl@Si_{20}Cl_{20}]^-$ | 124.9 | −308.5 | 0.00035 | 3.5 | 8.0 | 139.1 | 0.452 | −19.1 |
| $[Cl@Si_{20}Br_{20}]^-$ | 205.4 | −392.6 | 0.00054 | 3.5 | 7.6 | 131.1 | 0.438 | −19.6 |
| $[Cl@Si_{20}I_{20}]^-$ | 320.9 | −532.6 | 0.00013 | 3.4 | 7.1 | 119.6 | 0.477 | −20.4 |
| $[Cl@Si_{20}H_{20}]^-$ | 340.9 | −553.6 | 0.00013 | 5.4 | 6.6 | 76.3 | 0.513 | −18.2 |
| $[Cl@Si_{20}Me_{20}]^-$ | 333.3 | −535.2 | 0.00008 | 4.7 | 6.6 | 61.7 | 0.504 | −19.2 |
| $[Cl@Si_{20}Et_{20}]^-$ | 322.3 | −520.3 | 0.01025 | 4.7 | 6.4 | 60.9 | 0.522 | −19.0 |



***Scheme 1.*** (a) Syntheses of the siladodecahedrane salts [PPN][Cl@$Si_{20}Y_{20}$] (Y = Me, Et, Br). Reaction conditions: (i) 8 equiv. $Me_2AlH$, 40 equiv. $Me_3Al$, *o*DFB/toluene, rt, 15 h. Yield: 24%. (ii) 0.50 equiv. [PPN][Cl@$Si_{20}H_{20}$], 60 equiv. $Me_3Al$, *o*DFB/toluene, rt, 3 d. Yield: 10%. (iii) 0.50 equiv. [PPN][Cl@$Si_{20}H_{20}$], 20 equiv. $Et_3Al$, *o*DFB/hexanes, rt, 5 d. Yield: 30%. (iv) $CH_2Br_2/C_6D_6$ 5:1 (380 equiv. $CH_2Br_2$), 65 C, 48 h. Yield: 29%. *o*DFB = *ortho*-difluorobenzene. (b) Crystallographically determined structures of the siladodecahedrane salts [PPN][Cl@$Si_{20}Me_{20}$]×THF and [PPN][Cl@$Si_{20}Et_{20}$] in the solid state (the cations and co-crystallized THF are omitted for clarity). Displacement ellipsoids are drawn at the 50% probability level.

[PPN][Cl@$Si_{20}Et_{20}$] was isolated by crystallization from THF in 30% yield.[136]

The conversion of [PPN][Cl@$Si_{20}H_{20}$] to [PPN][Cl@$Si_{20}Cl_{20}$][44] by heating the hydridosilane in $CHCl_3$ served as the role model for the synthesis of [PPN][Cl@$Si_{20}Br_{20}$]: H/Br exchange on [PPN][Cl@$Si_{20}H_{20}$] was achieved by keeping its $CH_2Br_2/C_6D_6$ (5:1) solution in an NMR tube at 65 C for 48 h (Scheme 1a). At the endpoint of the reaction, a SiH signal was no longer detectable in the $^1$H NMR spectrum, and the resonance of the $CH_3Br$ byproduct showed an integral value indicating a 20:1 stoichiometry relative to $[PPN]^+$. After purification by precipitation from *o*DFB with *n*-hexane, [PPN][Cl@$Si_{20}Br_{20}$] was isolated as a yellow solid in 29% yield (Scheme 1a).[137]

In a first investigation of the stability of the siladodecahedrane salts [PPN][Cl@$Si_{20}Y_{20}$] toward Lewis bases, we found that $[Cl@Si_{20}Cl_{20}]^-$ and $[Cl@Si_{20}Br_{20}]^-$ decompose rapidly in THF ($[Cl@Si_{20}Cl_{20}]^-$ is stable in dry MeCN for extended periods of time). In stark contrast, $[Cl@Si_{20}Me_{20}]^-$ is not only inert to THF but persists for days even in the presence of the strong $F^-$ donor $[S(NMe_2)_3][Me_3SiF_2]$ (TASF). Apparently, the relatively low computed affinity of the permethylated siladodecahedrane cage for its endohedral $Cl^-$ guest finds its counterpart in a comparably low affinity for attacking exohedral Lewis bases. In contrast, halogenated $Si_{20}$ cages appear to be much more prone to both endohedral and exohedral interactions with Lewis bases. Attempts at the synthesis of [PPN][Cl@$Si_{20}F_{20}$] and [PPN][Cl@$Si_{20}I_{20}$] were of limited success, although we were able to detect the molecular-ion peak of $[Cl@Si_{20}I_{20}]^-$ in the mass spectrum ($m/z = 3134.66$, calcd.: 3134.59; cf. the Supporting Information for more details).

Crystals of [PPN][Cl@$Si_{20}Me_{20}$]×THF suitable for X-ray crystallography grew during the slow evaporation of the above-mentioned [PPN][Cl@$Si_{20}Me_{20}$]/TASF mixture in THF. [PPN][Cl@$Si_{20}Et_{20}$] crystallized from an *ortho*-difluorobenzene/hexanes mixture upon evaporation. The proposed molecular structures of the anions $[Cl@Si_{20}Me_{20}]^-$ and $[Cl@Si_{20}Et_{20}]^-$ were confirmed (Scheme 1b); however, the structure determination of [PPN][Cl@$Si_{20}Et_{20}$] was challenging due to the pseudo-merohedral twinning of three components[138] and required high-quality data obtained with synchrotron radiation (cf. the Supporting Information for details). The endohedral Cl···Si distances and Si–Si bond lengths in $[Cl@Si_{20}Me_{20}]^-$ and $[Cl@Si_{20}Et_{20}]^-$ are very similar to the corresponding published values[44] of $[Cl@Si_{20}H_{20}]^-$ and $[Cl@Si_{20}Cl_{20}]^-$ (Table 3). $[Cl@Si_{20}Me_{20}]^-$, $[Cl@Si_{20}Et_{20}]^-$, and $[Cl@Si_{20}(SiH_3)_{12}Me_8]^-$[134] have essentially identical Si–C bond lengths (Table 3).

The molecular-ion peaks of $[Cl@Si_{20}Y_{20}]^-$ (Y = Me, Et, Br, I) with matching isotope patterns were detected by LDI-MS(−). In contrast to $[Cl@Si_{20}Me_{20}]^-$ and $[Cl@Si_{20}Et_{20}]^-$, which do not appear to undergo significant fragmentation under the applied measurement conditions, $[Cl@Si_{20}Br_{20}]^-$ and $[Cl@Si_{20}I_{20}]^-$ show cluster fragmentation in the mass spectrometer. In both cases, peaks assignable to smaller clusters $[Si_aY_bCl]^-$ (Y = Br, I; $a,b \leq 20$) were found (cf. the Supporting Information for details).

The NMR spectra of [PPN][Cl@$Si_{20}Me_{20}$]/[PPN][Cl@$Si_{20}Et_{20}$] and [PPN][Cl@$Si_{20}Br_{20}$] were recorded in [$D_8$]THF and $CH_2Br_2/C_6D_6$ (5:1), respectively (Table 3). $[Cl@Si_{20}Me_{20}]^-$ gives singlets at 0.12 ppm and −7.6 ppm (Table 3) in the $^1$H and $^{13}C\{^1H\}$ NMR spectrum, respectively. The protons of $[Cl@Si_{20}Et_{20}]^-$ come to resonance at 1.20–1.11 ppm; two $^{13}$C signals were found at 12.8 ppm ($CH_3$) and 8.5 ppm ($CH_2$). In accordance with an average $I_h$ symmetry on the NMR time scale, the $^{29}$Si NMR spectra showed one resonance for each of the anions $[Cl@Si_{20}Me_{20}]^-$ ($\delta = -30.6$ ppm, calcd: −29.7 ppm; Table 3), $[Cl@Si_{20}Et_{20}]^-$ ($\delta = -18.4$ ppm, calcd: −13.0 ppm), and $[Cl@Si_{20}Br_{20}]^-$ ($\delta = -22.1$ ppm, calcd: −22.3 ppm). In the case of $[Cl@Si_{20}Me_{20}]^-$, the $J$(Si,Cl) coupling constant of 3.4 Hz was determined using a $^{29}$Si DEPT-3P experiment (calcd: 3.4 Hz; cf. $[Cl@Si_{20}H_{20}]^-$: 3.3 Hz, $[Cl@Si_{20}Cl_{20}]^-$: 4.7 Hz).[44] The $^{35}$Cl nuclei in $[Cl@Si_{20}Me_{20}]^-$ ($\delta = 332.7$ ppm, scaled calcd. value: 333.4 ppm) and $[Cl@Si_{20}Et_{20}]^-$ ($\delta = 326.3$, scaled calcd. value: 322.3 ppm) possess chemical shift values similar to that in $[Cl@Si_{20}H_{20}]^-$ ($\delta = 344.9$ ppm); significantly stronger $^{35}$Cl shielding was observed for $[Cl@Si_{20}Br_{20}]^-$ ($\delta = 210.9$ ppm, scaled calcd. value: 205.4 ppm).

*Table 3:* Selected experimentally determined (calculated) crystallographic and NMR-spectroscopic parameters[a] of the silafulleranes presented herein. Calculated NMR shifts have been computed at the SO-ZORA-PBE0[101] (COSMO($CH_2Cl_2$))/ZORA/TZP level of theory.

| Compound | $d$(Cl···Si)/Å | $d$(Si–Si)/Å | $d$(Si–C)/Å | $\delta(^1H)$ | $\delta(^{13}C)$ | $\delta(^{29}Si)$ | $\delta(^{35}Cl)$ |
|---|---|---|---|---|---|---|---|
| [PPN][Cl@$Si_{20}Me_{20}$] | 3.2967(11)–3.3161(8) | 2.3522(17)–2.3685(16) | 1.896(5)–1.920(3) | 0.12 | −7.6 (−4.8) | −30.6 (−29.7) | 332.7 (333.4)[b] |
| [PPN][Cl@$Si_{20}Et_{20}$] | 3.293(2)–3.3290(18) | 2.348(2)–2.381(3) | 1.903(5)–1.933(6) | 1.20–1.11 | 12.8 (16.3, $CH_3$)<br>8.5 (8.3, $CH_2$) | −18.4 (−13.0) | 326.3 (322.3)[b] |
| [PPN][Cl@$Si_{20}Br_{20}$] | – | – | – | – | – | −22.1 (−22.3) | 210.9 (205.4)[b] |
| [PPN][Cl@$Si_{20}H_{20}$][44] | 3.268(3)–3.314(4)[c] | 2.338(5)–2.363(5)[c] | – | 3.99 | – | −54.7 | 344.9 (340.9)[b] |
| [PPN][Cl@$Si_{20}Cl_{20}$][44] | 3.293(2)–3.324(2) | 2.350(2)–2.368(2) | – | – | – | −21.8[d] | 126.0[d] (124.9)[b] |
| [$n$$Bu_4N$][Cl@$Si_{20}(SiH_3)_{12}Me_8$][134] | 3.2541(3)–3.3040(3) (Cl···$Si^0$)<br>3.3191(3)–3.3433(3) (Cl···SiMe) | 2.3200(5)–2.3377(4) ($Si^0$–$SiH_3$)<br>2.3485(4)–2.3611(4) ($Si^0$–SiMe)<br>2.3453(4)–2.3544(4) ($Si^0$–$Si^0$) | 1.8974(12)–1.9189(13) | 0.51 ($CH_3$)<br>3.29 ($SiH_3$) | −2.4 | 19.1 (SiMe)<br>−67.2 ($Si^0$)<br>−100.6 ($SiH_3$) | 457.1 (456.2)[b] |

[a] NMR spectra were recorded in [$D_8$]THF ([PPN][Cl@$Si_{20}Me_{20}$], [PPN][Cl@$Si_{20}Et_{20}$], [PPN][Cl@$Si_{20}H_{20}$], [$n$$Bu_4N$][Cl@$Si_{20}(SiH_3)_{12}Me_8$]), $CH_2Br_2/C_6D_6$ (5:1; [PPN][Cl@$Si_{20}Br_{20}$]), or $CD_2Cl_2$ ([PPN][Cl@$Si_{20}Cl_{20}$]). [b] These values were obtained after scaling according to the following linear equation: $\delta(^{35}Cl, \text{scaled}) = 0.8728 \cdot \delta(^{35}Cl, \text{calcd}) - 7.3179$.[139] [c] These ranges cover the values determined for both the [$n$$Bu_4N$]$^+$ and the [PPN]$^+$ salt of [Cl@$Si_{20}H_{20}$]$^-$. [d] These values were determined for the [$n$$Bu_4N$]$^+$ salt.

## Conclusion

We conducted comprehensive quantum chemical and experimental studies to evaluate different compelling properties of [X@$Si_{20}Y_{20}$]$^-$ clusters and introduced the syntheses of the Y = Me, Et, Br derivatives. To find a reasonable level of theory for our computations, we evaluated the performance of different density functional approximations. The hybrid density functional r$^2$SCAN0-D4 in combination with a large and diffuse basis set yields the most accurate results for the tested systems. We elucidated the template effect of endohedral $Cl^-$ ions in the cluster assembly of [20]silafulleranes and investigated the underlying $X^- \rightarrow Si_{20}$ interaction as well as NMR properties of differently substituted [X@$Si_{20}Y_{20}$]$^-$ clusters. To further broaden the experimental basis of our theoretical considerations, the syntheses of compounds [PPN][Cl@$Si_{20}Y_{20}$] (Y = Me, Et, Br) are also described. Through our combined experimental and theoretical efforts, we are becoming increasingly able to synthesize novel silafulleranes in a targeted manner; at the same time, we have developed robust analytical tools to characterize their key properties.

Since intermediates of the cluster assembly are unknown, we studied a hypothetical series of silapolyquinane model systems and their propensity to coordinate a $Cl^-$ ion on their concave side. Our computations revealed that the interaction between the coordinated $Cl^-$ ion and the respective silicon frameworks becomes stronger as the clusters grow, which supports our previous postulate of a structure-determining template effect of the $Cl^-$ ion.

The main contributions to the attractive $X^- \rightarrow Si_{20}$ interaction energy ($E_{\text{int}}$) in [X@$Si_{20}Y_{20}$]$^-$ clusters stem from electrostatic and orbital-interaction effects. $E_{\text{int}}$ is influenced by a cooperative effect between the endohedral ion $X^-$ and the exohedral substituent Y. It is maximized if $X^-$ has the proper size and Y is strongly electron-withdrawing. We investigated different steric effects and found that endohedral ions bigger than $Cl^-$ lead to a larger cage strain and eventually to an expansion of the cage. As it is not feasible to extract the endohedral ion from the cage, we consider [X@$Si_{20}Y_{20}$]$^-$ clusters as confined Lewis pairs.

Further, we verified that the $^{35}$Cl NMR shift of an endohedral $Cl^-$ ion is a probe for $E_{\text{int}}$ in halogenated $Si_{20}$ cages and we could draw similarities and differences to the prominent Gutmann–Beckett method. The main contribution to the $^{35}$Cl NMR chemical shift is the paramagnetic shielding constant $\sigma_{\text{para}}$. Therefore, $\delta(^{35}Cl)$ is smaller for large donor-acceptor orbital gaps, which occur when the cage bears electronegative substituents.

With the newly found properties, the differently substituted [20]silafulleranes could, e.g., be used as weakly coordinating anions (WCA) or as building blocks for reticular assemblies in future applications.

## Acknowledgements

M.Ba. wishes to thank the Fonds der Chemischen Industrie (FCI) for a Kekulé Ph.D. grant. S.G. and M.Bu. gratefully acknowledge financial support from the Max Planck Society through the Max Planck Fellow program. M.W. thanks the Deutsche Forschungsgemeinschaft for financial support (DFG grant no. 506550642). The authors are grateful to Evonik Operations GmbH, Rheinfelden (Germany), for the generous donation of $Si_2Cl_6$. Parts of this research (project I-20220822) were carried out at PETRA III at DESY, a member of the Helmholtz Association (HGF). A.V. and M.Ba. thank Dr. Leila Noohinejad, Dr. Martin Tolkiehn, and Dr. Eugenia Peresypkina for their assistance regarding the use of the beamline P24. A.V. thanks Dr. Matthias Meyer (Rigaku Oxford Diffraction) for his precious help with the implementation of the CrysAlisPro software for the synchrotron and STOE IPDS II diffraction data. T.G. thanks Christoph Plett for fruitful discussions. M.Ba. thanks Prof. Dr. Lutz Greb for helpful discussions. Open Access funding enabled and organized by Projekt DEAL.

## Conflict of Interest

The authors declare no competing financial interest.

## Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

**Keywords:** $^{35}$Cl NMR · Confined Lewis Pairs · Coupled-Cluster · DFT · Silafulleranes

---

[1] C. Marschner, T. D. Tilley, *Dalton Trans.* **2017**, *46*, 8699.
[2] N. Wiberg, C. M. M. Finger, K. Polborn, *Angew. Chem. Int. Ed.* **1993**, *32*, 1054.
[3] H. Matsumoto, K. Higuchi, Y. Hoshino, H. Koike, Y. Naoi, Y. Nagai, *J. Chem. Soc. Chem. Commun.* **1988**, *3*, 1083.
[4] J. Fischer, J. Baumgartner, C. Marschner, *Science* **2005**, *310*, 825.
[5] H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, R. E. Smalley, *Nature* **1985**, *318*, 162.
[6] R. F. Curl, R. E. Smalley, *Science* **1988**, *242*, 1017.
[7] W. Krätschmer, L. D. Lamb, K. Fostiropoulos, D. R. Huffman, *Nature* **1990**, *347*, 354.
[8] H. Prinzbach, A. Weiler, P. Landenberger, F. Wahl, J. Wörth, L. T. Scott, M. Gelmont, D. Olevano, B. v. Issendorff, *Nature* **2000**, *407*, 60.
[9] W. Krätschmer, *Nanoscale* **2011**, *3*, 2485.
[10] R. J. Ternansky, D. W. Balogh, L. A. Paquette, *J. Am. Chem. Soc.* **1982**, *104*, 4503.
[11] J. C. Grossman, L. Mitáš, *Phys. Rev. Lett.* **1995**, *74*, 1323.
[12] K.-M. Ho, A. A. Shvartsburg, B. Pan, Z.-Y. Lu, C.-Z. Wang, J. G. Wacker, J. L. Fye, M. F. Jarrold, *Nature* **1998**, *392*, 582.
[13] B.-x. Li, P.-l. Cao, *Phys. Rev. A* **2000**, *62*, 023201.
[14] I. Rata, A. A. Shvartsburg, M. Horoi, T. Frauenheim, K. W. M. Siu, K. A. Jackson, *Phys. Rev. Lett.* **2000**, *85*, 546.
[15] B.-x. Li, P.-l. Cao, *J. Phys. Condens. Matter* **2001**, *13*, 10865.
[16] Unsaturated silicon clusters containing non-substituted cluster vertices are mainly known in the form of Zintl ions and siliconoids: a) Y. Heider, D. Scheschkewitz, *Chem. Rev.* **2021**, *121*, 9674; b) D. Scheschkewitz, *Angew. Chem. Int. Ed.* **2005**, *44*, 2954; c) J. Keuter, C. Schwermann, A. Hepp, K. Bergander, J. Droste, M. R. Hansen, N. L. Doltsinis, C. Mück-Lichtenfeld, F. Lips, *Chem. Sci.* **2020**, *11*, 5895; d) S. Scharfe, F. Kraus, S. Stegmaier, A. Schier, T. F. Fässler, *Angew. Chem. Int. Ed.* **2011**, *50*, 3630; e) S. Joseph, M. Hamberger, F. Mutzbauer, O. Härtl, M. Meier, N. Korber, *Angew. Chem. Int. Ed.* **2009**, *48*, 8770.
[17] K. Jackson, B. Nellermoe, *Chem. Phys. Lett.* **1996**, *254*, 249.
[18] Q. Sun, Q. Wang, T. M. Briere, V. Kumar, Y. Kawazoe, P. Jena, *Phys. Rev. B: Condens. Matter Mater. Phys.* **2002**, *65*, 235417.
[19] T. Nagano, K. Tsumuraya, H. Eguchi, D. J. Singh, *Phys. Rev. B: Condens. Matter Mater. Phys.* **2001**, *64*, 155403.
[20] A. K. Singh, V. Kumar, Y. Kawazoe, *Phys. Rev. B: Condens. Matter Mater. Phys.* **2005**, *71*, 115429.
[21] J.-P. Dognon, C. Clavaguéra, P. Pyykkö, *Chem. Sci.* **2012**, *3*, 2843.
[22] This formal charge assignment does not claim to reflect the actual charge distribution in the molecule.
[23] S. Nagase, *Acc. Chem. Res.* **1995**, *28*, 469.
[24] L. Silaghi-Dumitrescu, A. Kun, I. Haiduc, *Fullerene Sci. Technol.* **1999**, *7*, 841.
[25] F. Pichierri, V. Kumar, *J. Mol. Struct.: THEOCHEM* **2009**, *900*, 71.
[26] A. D. Zdetsis, *Phys. Rev. B: Condens. Matter Mater. Phys.* **2009**, *80*, 195417.
[27] M. Anafcheh, R. Ghafouri, N. L. Hadipour, *Phys. E* **2012**, *44*, 2099.
[28] E. N. Koukaras, A. D. Zdetsis, P. Karamanis, C. Pouchan, A. Avramopoulos, M. G. Papadopoulos, *J. Comput. Chem.* **2012**, *33*, 1068.
[29] F. Marsusi, M. Qasemnazhand, *Nanotechnology* **2016**, *27*, 275704.
[30] C. W. Earley, *J. Phys. Chem. A* **2000**, *104*, 6622.
[31] V. Kumar, Y. Kawazoe, *Phys. Rev. Lett.* **2003**, *90*, 055502.
[32] F. Pichierri, V. Kumar, Y. Kawazoe, *Chem. Phys. Lett.* **2004**, *383*, 544.
[33] G. Ramachandran, S. Manogaran, *J. Mol. Struct.: THEOCHEM* **2005**, *730*, 171.
[34] A. J. Karttunen, M. Linnolahti, T. A. Pakkanen, *J. Phys. Chem. C* **2007**, *111*, 2545.
[35] Y. Pei, Y. Gao, X. C. Zeng, *J. Chem. Phys.* **2007**, *127*, 044704.
[36] B.-C. Wang, Y.-M. Chou, J.-P. Deng, Y.-T. Dung, *J. Phys. Chem. A* **2008**, *112*, 6351.
[37] H. Wang, L. Wu, *Chin. J. Chem.* **2011**, *29*, 2063.
[38] D. Palagin, K. Reuter, *ACS Nano* **2013**, *7*, 1763.
[39] F. Pichierri, V. Kumar, Y. Kawazoe, *Chem. Phys. Lett.* **2005**, *406*, 341.
[40] D. Palagin, K. Reuter, *Phys. Rev. B: Condens. Matter Mater. Phys.* **2012**, *86*, 045416.
[41] C.-Y. Zhang, H.-S. Wu, *J. Mol. Struct.: THEOCHEM* **2006**, *770*, 145.
[42] C.-Y. Zhang, H.-S. Wu, H. Jiao, *Chem. Phys. Lett.* **2005**, *410*, 457.
[43] J. Tillmann, J. H. Wender, U. Bahr, M. Bolte, H.-W. Lerner, M. C. Holthausen, M. Wagner, *Angew. Chem. Int. Ed.* **2015**, *54*, 5429.
[44] M. Bamberg, M. Bursch, A. Hansen, M. Brandl, G. Sentis, L. Kunze, M. Bolte, H.-W. Lerner, S. Grimme, M. Wagner, *J. Am. Chem. Soc.* **2021**, *143*, 10865.
[45] D. S. De, B. Schaefer, B. v. Issendorff, S. Goedecker, *Phys. Rev. B: Condens. Matter Mater. Phys.* **2020**, *101*, 214303.
[46] M. Ponce-Vargas, A. Muñoz-Castro, *J. Phys. Chem. C* **2018**, *122*, 12551.
[47] F. Neese, F. Wennmohs, U. Becker, C. Riplinger, *J. Chem. Phys.* **2020**, *152*, 224108.
[48] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, S. Grimme, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2021**, *11*, e1493.
[49] TURBOMOLE V7.3 2018, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–

2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.

[50] H.-J. Werner, P. J. Knowles, F. R. Manby, J. A. Black, K. Doll, A. Heßelmann, D. Kats, A. Köhn, T. Korona, D. A. Kreplin, Q. Ma, T. F. Miller, A. Mitrushchenkov, K. A. Peterson, I. Polyak, G. Rauhut, M. Sibaev, *J. Chem. Phys.* **2020**, *152*, 144107.

[51] H.-J. Werner, P. J. Knowles, P. Celani, W. Györffy, A. Hesselmann, D. Kats, G. Knizia, A. Köhn, T. Korona, D. Kreplin, R. Lindh, Q. Ma, F. R. Manby, A. Mitrushenkov, G. Rauhut, M. Schütz, K. R. Shamasundar, T. B. Adler, R. D. Amos, S. J. Bennie, A. Bernhardsson, A. Berning, J. A. Black, P. J. Bygrave, R. Cimiraglia, D. L. Cooper, D. Coughtrie, M. J. O. Deegan, A. J. Dobbyn, K. Doll, M. Dornbach, F. Eckert, S. Erfort, E. Goll, C. Hampel, G. Hetzer, J. G. Hill, M. Hodges, T. Hrenar, G. Jansen, C. Köppl, C. Kollmar, S. J. R. Lee, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, B. Mussard, S. J. Mc-Nicholas, W. Meyer, T. F. Miller III, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. A. Peterson, K. Pflüger, R. Pitzer, I. Polyak, M. Reiher, J. O. Richardson, J. B. Robinson, B. Schröder, M. Schwilk, T. Shiozaki, M. Sibaev, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, J. Toulouse, M. Wang, M. Welborn, B. Ziegler, MOLPRO, 2022.3, a package of ab initio programs.

[52] G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, T. Ziegler, *J. Comput. Chem.* **2001**, *22*, 931.

[53] U. Mayer, V. Gutmann, W. Gerger, *Monatsh. Chem.* **1975**, *106*, 1235.

[54] M. A. Beckett, G. C. Strickland, J. R. Holland, K. S. Varma, *Polymer* **1996**, *37*, 4629.

[55] J. Tillmann, L. Meyer, J. I. Schweizer, M. Bolte, H.-W. Lerner, M. Wagner, M. C. Holthausen, *Chem. Eur. J.* **2014**, *20*, 9234.

[56] S.-B. Choi, B.-K. Kim, P. Boudjouk, D. G. Grier, *J. Am. Chem. Soc.* **2001**, *123*, 8117.

[57] X. Dai, D. L. Schulz, C. W. Braun, A. Ugrinov, P. Boudjouk, *Organometallics* **2010**, *29*, 2203.

[58] X. Dai, K. J. Anderson, D. L. Schulz, P. Boudjouk, *Dalton Trans.* **2010**, *39*, 11188.

[59] X. Dai, S.-B. Choi, C. W. Braun, P. Vaidya, S. Kilina, A. Ugrinov, D. L. Schulz, P. Boudjouk, *Inorg. Chem.* **2011**, *50*, 4047.

[60] K. Pokhodnya, C. Olson, X. Dai, D. L. Schulz, P. Boudjouk, A. P. Sergeeva, A. I. Boldyrev, *J. Chem. Phys.* **2011**, *134*, 014105.

[61] J. Tillmann, F. Meyer-Wegner, A. Nadj, J. Becker-Baldus, T. Sinke, M. Bolte, M. C. Holthausen, M. Wagner, H.-W. Lerner, *Inorg. Chem.* **2012**, *51*, 8599.

[62] A. Robertazzi, J. A. Platts, P. Gamez, *ChemPhysChem* **2014**, *15*, 912.

[63] K. Pokhodnya, K. Anderson, S. Kilina, N. Dandu, P. Boudjouk, *J. Phys. Chem. A* **2018**, *122*, 4067.

[64] J. Teichmann, B. Köstler, J. Tillmann, M. Moxter, R. Kupec, M. Bolte, H.-W. Lerner, M. Wagner, *Z. Anorg. Allg. Chem.* **2018**, *644*, 956.

[65] J. Yang, B. Chatelet, V. Dufaud, D. Hérault, S. Michaud-Chevallier, V. Robert, J.-P. Dutasta, A. Martinez, *Angew. Chem. Int. Ed.* **2018**, *57*, 14212.

[66] C. Li, A.-D. Manick, J.-P. Dutasta, X. Bugaut, B. Chatelet, A. Martinez, *Org. Chem. Front.* **2022**, *9*, 1826.

[67] S. Grimme, J. G. Brandenburg, C. Bannwarth, A. Hansen, *J. Chem. Phys.* **2015**, *143*, 054107.

[68] A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **2009**, *113*, 6378.

[69] For an overview of organic polyquinane chemistry, see: L. A. Paquette, The development of polyquinane chemistry, in *Organic Chemistry*, Springer Berlin Heidelberg, Berlin, Heidelberg **1979**, pages 41–165.

[70] A. Tsurusaki, Y. Koyama, S. Kyushin, *J. Am. Chem. Soc.* **2017**, *139*, 3982.

[71] W. B. Schneider, G. Bistoni, M. Sparta, M. Saitow, C. Riplinger, A. A. Auer, F. Neese, *J. Chem. Theory Comput.* **2016**, *12*, 4778.

[72] G. Bistoni, *WIREs Comput. Mol. Sci.* **2020**, *10*, e1442.

[73] A. Altun, M. Saitow, F. Neese, G. Bistoni, *J. Chem. Theory Comput.* **2019**, *15*, 1616.

[74] C. Riplinger, F. Neese, *J. Chem. Phys.* **2013**, *138*, 034106.

[75] C. Riplinger, B. Sandhoefer, A. Hansen, F. Neese, *J. Chem. Phys.* **2013**, *139*, 134101.

[76] C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, F. Neese, *J. Chem. Phys.* **2016**, *144*, 024109.

[77] M. Saitow, U. Becker, C. Riplinger, E. F. Valeev, F. Neese, *J. Chem. Phys.* **2017**, *146*, 164105.

[78] Y. Guo, C. Riplinger, U. Becker, D. G. Liakos, Y. Minenkov, L. Cavallo, F. Neese, *J. Chem. Phys.* **2018**, *148*, 011101.

[79] Q. Ma, H.-J. Werner, *J. Chem. Theory Comput.* **2021**, *17*, 902.

[80] Q. Ma, H.-J. Werner, *WIREs Comput. Mol. Sci.* **2018**, *8*, e1371.

[81] Q. Ma, M. Schwilk, C. Köppl, H.-J. Werner, *J. Chem. Theory Comput.* **2017**, *13*, 4871.

[82] To generate reliable reference interaction energies ($E_{\text{int}}$), we applied a high-level, state-of-the-art local coupled cluster method including explicit correlation (PNO-LCCSD(T)-F12b employed with tight domain settings) with a modified aug-cc-pVQZ basis set (aug-cc-pVQZ-PP for the heavier elements, cc-pVQZ for H; denoted AVQZ' in the following). Since this setup is computationally too expensive for the larger structures (and those with many heavier elements), we applied a slightly less accurate but clearly faster reference protocol (PNO-LCCSD(T)-F12b/AVTZ' with default domain settings) for the latter systems. The respective additional error could be significantly reduced by applying a scaling factor to the triples contributions (see Computational Details in Supporting Information). The, in this way, obtained interaction energies for the subset comprised of the smaller systems are virtually identical to the respective PNO-LCCSD(T)-F12b/AVQZ'/tight results with a mean unsigned error of only 0.05 kcal mol$^{-1}$ (see Supporting Information, Table S1), so that comparably accurate reference values could also be generated for the larger subset. We conservatively estimate the residual error of the generated reference values to be 0.5–1.0 kcal mol$^{-1}$ for $E_{\text{int}}$, thus allowing for a statistical discriminability of about 0.1 kcal mol$^{-1}$ of the mean errors for the evaluated DFT methods.

[83] H.-J. Werner, A. Hansen, *J. Chem. Theory Comput.* **2023**, *19*, 7007.

[84] N. B. Balabanov, K. A. Peterson, *J. Chem. Phys.* **2005**, *123*, 064107.

[85] K. A. Peterson, D. Figgen, M. Dolg, H. Stoll, *J. Chem. Phys.* **2007**, *126*, 124101.

[86] T. H. Dunning, *J. Chem. Phys.* **1989**, *90*, 1007.

[87] A. Klamt, G. Schüürmann, *J. Chem. Soc. Perkin Trans. 2* **1993**, 799.

[88] M. Cossi, N. Rega, G. Scalmani, V. Barone, *J. Comput. Chem.* **2003**, *24*, 669.

[89] A. Klamt, *J. Phys. Chem.* **1995**, *99*, 2224.

[90] COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design, in Klamt (Editor), *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*, pages 1–234, Elsevier, Amsterdam **2005**.

[91] S. Grimme, *Chem. Eur. J.* **2012**, *18*, 9955.

[92] We tested the following density functional approximations in combination the the ma-def2-QZVPP basis set: PBE-D4, TPSS-D4, PW6B95-D4, B3LYP-D4, PBE0-D4, r$^2$SCAN0-D4, $\omega$B97X-D4, $\omega$B97X-V, $\omega$B97M-D4, $\omega$B97M-V, $\omega$B97X-2-D4, B2PLYP-D4 and PWPB95-D4. Further, the composite r$^2$SCAN-3c DFT method was tested with the def2-mTZVPP basis set.

[93] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865.

[94] E. Caldeweyher, C. Bannwarth, S. Grimme, *J. Chem. Phys.* **2017**, *147*, 034112.

[95] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, S. Grimme, *J. Chem. Phys.* **2019**, *150*, 154122.

[96] J. Tao, J. P. Perdew, V. N. Staroverov, G. E. Scuseria, *Phys. Rev. Lett.* **2003**, *91*, 146401.

[97] Y. Zhao, D. G. Truhlar, *J. Phys. Chem. A* **2005**, *109*, 5656.

[98] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785.

[99] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.

[100] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623.

[101] C. Adamo, V. Barone, *J. Chem. Phys.* **1999**, *110*, 6158.

[102] M. Bursch, H. Neugebauer, S. Ehlert, S. Grimme, *J. Chem. Phys.* **2022**, *156*, 134105.

[103] A. Najibi, L. Goerigk, *J. Comput. Chem.* **2020**, *41*, 2562.

[104] N. Mardirossian, M. Head-Gordon, *Phys. Chem. Chem. Phys.* **2014**, *16*, 9904.

[105] N. Mardirossian, M. Head-Gordon, *J. Chem. Phys.* **2016**, *144*, 214110.

[106] J.-D. Chai, M. Head-Gordon, *J. Chem. Phys.* **2009**, *131*, 174105.

[107] S. Grimme, *J. Chem. Phys.* **2006**, *124*, 034108.

[108] L. Goerigk, S. Grimme, *J. Chem. Theory Comput.* **2011**, *7*, 291.

[109] J. Zheng, X. Xu, D. G. Truhlar, *Theor. Chem. Acc.* **2011**, *128*, 295.

[110] S. Grimme, A. Hansen, S. Ehlert, J. M. Mewes, *J. Chem. Phys.* **2021**, *154*, 064103.

[111] D. Rappoport, F. Furche, *J. Chem. Phys.* **2010**, *133*, 134105.

[112] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.

[113] P. Su, H. Li, *J. Chem. Phys.* **2009**, *131*, 014102.

[114] M. Mitoraj, A. Michalak, *J. Mol. Model.* **2007**, *13*, 347.

[115] A. Michalak, M. Mitoraj, T. Ziegler, *J. Phys. Chem. A* **2008**, *112*, 1933.

[116] T. Ziegler, A. Rauk, *Theor. Chim. Acta* **1977**, *46*, 1.

[117] M. P. Mitoraj, A. Michalak, T. Ziegler, *J. Chem. Theory Comput.* **2009**, *5*, 962.

[118] D. MacLeod-Carey, P. L. Rodríguez-Kessler, A. Muñoz-Castro, *Phys. Chem. Chem. Phys.* **2023**, *25*, 19845.

[119] S. K. Wolff, T. Ziegler, *J. Chem. Phys.* **1998**, *109*, 895.

[120] G. Schreckenbach, T. Ziegler, *J. Phys. Chem.* **1995**, *99*, 606.

[121] H. Fukui, *Prog. Nucl. Magn. Reson. Spectrosc.* **1997**, *31*, 317.

[122] C. Chang, M. Pelissier, P. Durand, *Phys. Scr.* **1986**, *34*, 394.

[123] E. van Lenthe, E. J. Baerends, J. G. Snijders, *J. Chem. Phys.* **1993**, *99*, 4597.

[124] E. van Lenthe, E. J. Baerends, J. G. Snijders, *J. Chem. Phys.* **1994**, *101*, 9783.

[125] E. van Lenthe, E. J. Baerends, *J. Comput. Chem.* **2003**, *24*, 1142.

[126] C. C. Pye, T. Ziegler, E. van Lenthe, J. N. Louwen, *Can. J. Chem.* **2009**, *87*, 790.

[127] P. Erdmann, L. Greb, *Angew. Chem. Int. Ed.* **2022**, *61*, e202114550.

[128] For Y=F, we find no such correlation between $E_{\text{int}}$ and $\delta(^{35}\text{Cl})$. However, in this case $\triangle\delta(^{35}\text{Cl})$ is rather small between the different cluster sizes, and the same is true for the absolute $\delta(^{35}\text{Cl})$ values, so that any systematic trends can easily be obscured by inaccuracies inherent to the theoretical method.

[129] J. C. Facelli, *Concepts Magn. Reson.* **2004**, *20A*, 42.

[130] J. Gauss, U. Schneider, R. Ahlrichs, C. Dohmeier, H. Schnöckel, *J. Am. Chem. Soc.* **1993**, *115*, 2402.

[131] J. Autschbach, *J. Chem. Phys.* **2008**, *128*, 164112.

[132] R. Ditchfield, *Mol. Phys.* **1974**, *27*, 789.

[133] K. Wolinski, J. F. Hinton, P. Pulay, *J. Am. Chem. Soc.* **1990**, *112*, 8251.

[134] M. Bamberg, T. Gasevic, M. Bolte, A. Virovets, H.-W. Lerner, S. Grimme, M. Bursch, M. Wagner, *J. Am. Chem. Soc.* **2023**, *145*, 11440.

[135] As reported previously,[134] our synthesis leads to a product with the composition $Me_{2.2}AlH_{0.8}$. For simplicity, we refer to the idealized formula $Me_2AlH$ in the main text, but we use the exact stoichiometry in the Supporting Information.

[136] CIF files containing the crystallographic information were deposited in the Cambridge Crystallographic Data Centre under the deposition codes CCDC 2294476 ([PPN][Cl@$Si_{20}Me_{20}$]×THF), CCDC 2294477 ([PPN][Cl@$Si_{20}Et_{20}$]), and CCDC 2294478 ([$n$$Bu_4N$]$_2$[$Si_6I_{10.85}Cl_{1.15}$·2I]; cf. the Supporting Information for details) and can be obtained free of charge via www.ccdc.cam.ac.uk/data_request/cif.

[137] The actual conversion to [PPN][Cl@$Si_{20}Br_{20}$] is about 90%; the purification step served to remove traces of unknown impurities.

[138] R. Herbst-Irmer, G. M. Sheldrick, *Acta Crystallogr. Sect. B* **1998**, *54*, 443.

[139] M. Bamberg, T. Gasevic, M. Bolte, A. Virovets, H.-W. Lerner, S. Grimme, M. Bursch, M. Wagner, *Chem. Commun.* **2023**, *59*, 7459.

Manuscript received: September 27, 2023
Accepted manuscript online: December 7, 2023
Version of record online: January 2, 2024

# APPENDIX F

# Regioselective Derivatization of Silylated [20]Silafulleranes

Marcel Bamberg,[†] Thomas Gasevic,[‡] Michael Bolte,[†] Alexander Virovets,[†] Hans-Wolfram Lerner,[†] Stefan Grimme,[‡] Markus Bursch,[‡] and Matthias Wagner[†]

*Received: March 29, 2023*
*First published: May 12, 2023*

Repinted (adapted) with permission from:
M. Bamberg, T. Gasevic, M. Bolte, A. Virovets, H.-W. Lerner, S. Grimme, M. Bursch, and M. Wagner, *Regioselective Derivatization of Silylated [20]Silafulleranes*, J. Am. Chem. Soc. **145**.20 (2023) 11440, DOI: `10.1021/jacs.3c03270`

**Own contributions**

- Formal Analysis
- Investigation – Performing QC calculations
- Writing – Review & Editing

---

[†] Institut für Anorganische und Analytische Chemie, Goethe-Universität Frankfurt am Main, 60438 Frankfurt am Main, Germany

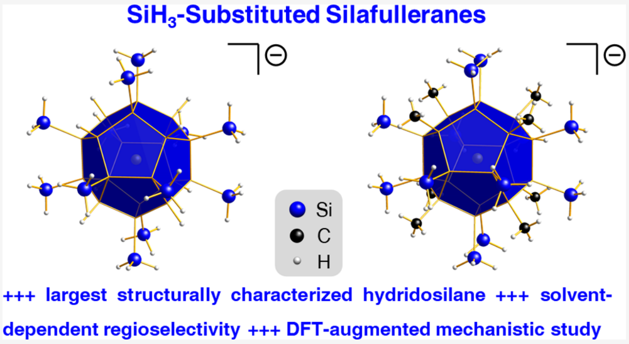
[‡] Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany

J A C S
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS Article

# Regioselective Derivatization of Silylated [20]Silafulleranes

Marcel Bamberg, Thomas Gasevic, Michael Bolte, Alexander Virovets, Hans-Wolfram Lerner, Stefan Grimme, Markus Bursch,* and Matthias Wagner*

Cite This: J. Am. Chem. Soc. 2023, 145, 11440−11448



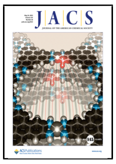
**ABSTRACT:** Silafulleranes with endohedral $Cl^-$ ions are a unique, scarcely explored class of structurally well-defined silicon clusters and host-guest complexes. Herein, we report regioselective derivatization reactions on the siladodecahedrane $[nBu_4N][Cl@Si_{20}(SiCl_3)_{12}Cl_8]$ ($[nBu_4N][\mathbf{1}]$), which has its cluster surface decorated with 12 $SiCl_3$ and 8 Cl substituents in perfect $T_h$ symmetry. The room-temperature reaction of $[nBu_4N][\mathbf{1}]$ with excess $iBu_2AlH$ in *ortho*-difluorobenzene (*o*DFB) furnishes perhydrogenated $[nBu_4N][Cl@Si_{20}(SiH_3)_{12}H_8]$ ($[nBu_4N][\mathbf{2}]$) in 50% yield; the non-pyrophoric $[\mathbf{2}]^-$ is the largest structurally authenticated (by X-ray diffraction) hydridosilane known to date. A simple switch from pure *o*DFB to an *o*DFB/$Et_2O$ solvent mixture suppresses core hydrogenation and results in the formation of $[nBu_4N][Cl@Si_{20}(SiH_3)_{12}Cl_8]$ ($[nBu_4N][\mathbf{3}]$). In addition to the exhaustive Cl/H exchange at all 44 Si−Cl bonds of $[\mathbf{1}]^-$ and the regioselective 36-fold silyl group hydrogenation, we achieved the simultaneous introduction of Me substituents at all 8 SiCl vertices along with the conversion of all 12 $SiCl_3$ to $SiH_3$ groups by treating $[nBu_4N][\mathbf{1}]$ with $Me_2AlH/Me_3Al$ in *o*DFB ($[nBu_4N][Cl@Si_{20}(SiH_3)_{12}Me_8]$, $[nBu_4N][\mathbf{4}]$; 73%). Quantum-chemical free-energy calculations find an $S_N2$-Si-type hydrogenation of the exohedral $SiCl_3$ moieties in $[\mathbf{1}]^-$ (trigonal-bipyramidal intermediate) slightly preferred over metathesis-like $S_Ni$-Si substitutions (four-membered transition state). Cage hydrogenation likely occurs via $S_Ni$-Si processes. The experimentally demonstrated influence of an $Et_2O$ co-solvent, which drastically increases the respective reaction barriers, is attributed to the increased stability of the resulting $iBu_2AlH$-$OEt_2$ adduct and its higher steric bulk compared to free $iBu_2AlH$.

## ■ INTRODUCTION

The discovery of the siladodecahedrane $[nBu_4N][Cl@Si_{20}(SiCl_3)_{12}Cl_8]$ ($[nBu_4N][\mathbf{1}]$; Figure 1) in 2015 provided the first preparative access to the class of compounds known as silafulleranes,[1] which had already received much attention from the theoretical chemistry community.[2−5] $[nBu_4N][\mathbf{1}]$ is straightforwardly available from $Si_2Cl_6$, $[nBu_4N]Cl$, and $nBu_3N$ in one step and in about 30% yield.[1,6] The anion $[\mathbf{1}]^-$ consists of a dodecahedral $Si_{20}$ core carrying 12 $SiCl_3$ groups and 8 Cl atoms in a perfectly $T_h$-symmetric pattern and hosting a $Cl^-$ ion as endohedral guest. This $Cl^-$ ion likely plays a crucial role as a structure-directing template during the assembly of the silafullerane cluster before it is eventually encapsulated into it. Furthermore, it imparts a negative charge to the cluster, which facilitates the isolation of the salt-like product by crystallization and its analysis by laser desorption ionization-mass spectrometry (LDI-MS) in the negative-ion mode.[1] The presence of Si−Cl bonds in the substituent sphere of $[\mathbf{1}]^-$ is uncommon in the field of molecular Si and Ge clusters,[7] since most of the saturated clusters reported so far, such as tetrasilatetrahedranes,[8,9] octasilacubanes,[10,11] or decasilaadamantane[12,13] and its isomers,[12,14] are decorated with rather inert alkyl, aryl, or organosilyl groups. Nevertheless, various routes of derivatization have been developed, including reductive[9] or nucleophilic[13] removal of substituents, oxidative cage opening,[15,16] and one-electron reduction of the cluster core.[17] Furthermore, a rich chemistry of unsaturated, so-called "metalloid"[18] Si and Ge clusters[19−24] and completely ligand-free Zintl ions has evolved.[25−31]

Coming back to the silafullerane $[nBu_4N][\mathbf{1}]$, its follow-up chemistry can be developed along two paths (Figure 1): (1) stripping of the 12 $SiCl_3$ groups to enter the area of $Si_{20}$ clusters and (2) derivatization of $[\mathbf{1}]^-$ while preserving its $Si_{32}$ framework. Following path 1, our groups have recently synthesized and characterized the parent siladodecahedrane with an endohedral $Cl^-$ ion, $[Cl@Si_{20}H_{20}]^-$ (Figure 1),[32] and put an end to previous, exclusively theory-based, discussions on the existence of the $I_h$-symmetric $Si_{20}H_{20}$ cage.[3,5] Our synthesis builds upon the protodesilylation of $[\mathbf{1}]^-$ with

**Received:** March 29, 2023
**Published:** May 12, 2023



https://doi.org/10.1021/jacs.3c03270
J. Am. Chem. Soc. 2023, 145, 11440−11448

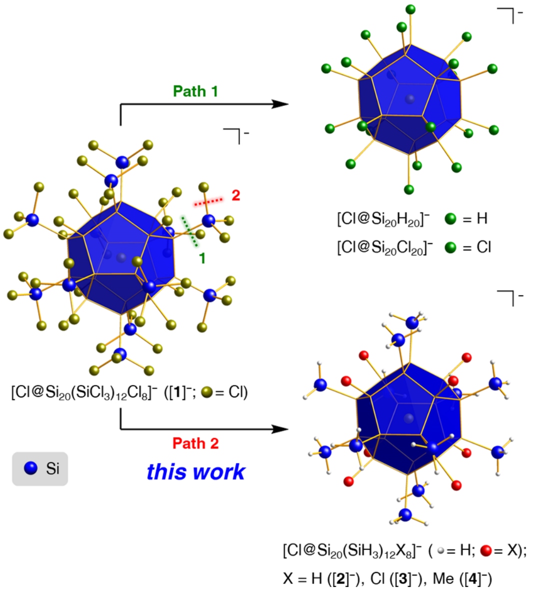


**Figure 1.** Fundamental approaches to the derivatization of $[\mathbf{1}]^-$ (paths 1 and 2). Path 1: Desilylation allowing the synthesis of $Si_{20}$ dodecahedranes. Path 2: Derivatization preserving the original $Si_{32}$ framework.

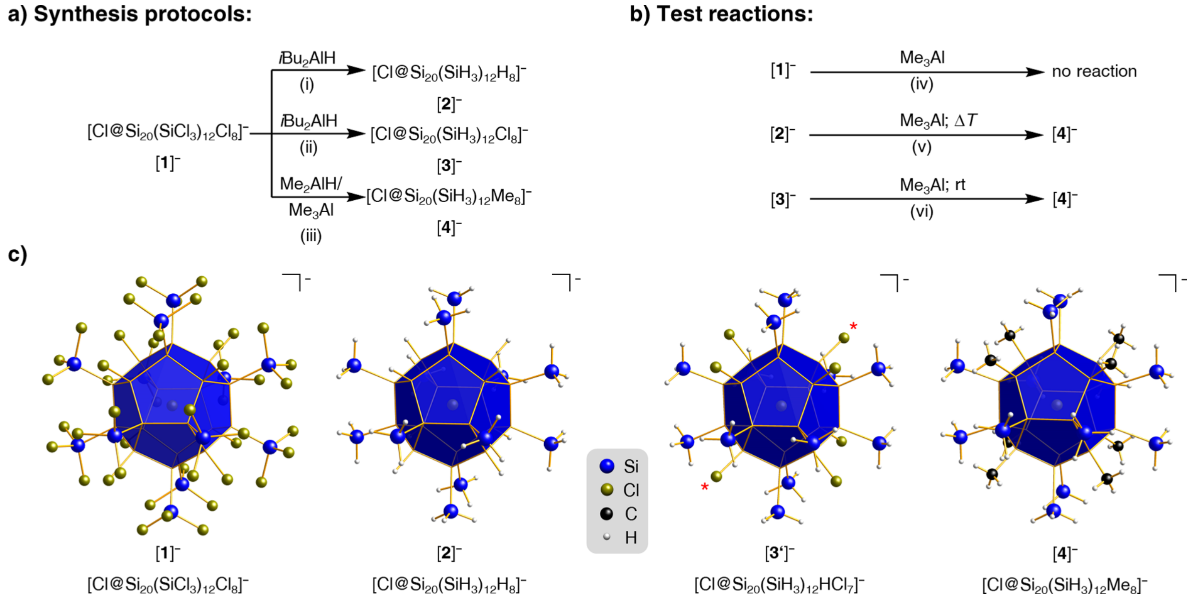
pinacol and a subsequent Cl/H exchange on the primary product $[Cl@Si_{20}H_{12}Cl_8]^-$ with diisobutylaluminum hydride ($i$$Bu_2AlH$). Perchlorination of $[Cl@Si_{20}H_{20}]^-$ to afford $[Cl@Si_{20}Cl_{20}]^-$ was achieved by treatment with chloromethanes. Due to their highly symmetric environments in the centers of the silafullerane $[\mathbf{1}]^-$, $[Cl@Si_{20}H_{12}Cl_8]^-$, $[Cl@Si_{20}H_{20}]^-$, and $[Cl@Si_{20}Cl_{20}]^-$, the endohedral $Cl^-$ ions give rise to unusually sharp $^{35}Cl$ NMR signals and observable $J(H,Cl)$ and $J(Si,Cl)$ couplings in the respective NMR spectra. After a thorough quantum-chemical analysis of the shielding components, we found that the chemical shift value $\delta(^{35}Cl)$ provides a useful analytical probe for evaluating the degree of endohedral $Cl^- \rightarrow Si_{20}$ interaction.[32]

Herein, we are now focusing on path 2 (Scheme 1a) and disclose (i) the perhydrogenation of $[\mathbf{1}]^-$ to furnish $[Cl@Si_{20}(SiH_3)_{12}H_8]^-$ ($[\mathbf{2}]^-$), (ii) a regioselective partial Cl/H exchange on $[\mathbf{1}]^-$ to give $[Cl@Si_{20}(SiH_3)_{12}Cl_8]^-$ ($[\mathbf{3}]^-$), and (iii) a regioselective three-component reaction to afford $[Cl@Si_{20}(SiH_3)_{12}Me_8]^-$ ($[\mathbf{4}]^-$). The new syntheses are complemented by theoretical studies of the underlying reaction mechanisms, the importance of which goes far beyond our specific cluster chemistry, as they improve the understanding of chlorosilane hydrogenation in general.

## ■ RESULTS AND DISCUSSION

Hydridosilanes, along with halosilanes and organosilanes, are among the most versatile classes of silicon compounds, but examples of large, non-polymeric representatives are still rare.[33,34] The same is true for selectively mixed-substituted oligosilanes. The synthesis of hydrogenated silafulleranes with molecular weights of around 1000 g mol⁻¹ is therefore an important objective to get access to the largest structurally well-defined hydridosilanes known to date. The high degree of functionalization of $[\mathbf{1}]^-$ is both a blessing and a calamity. To cope with this ambivalent situation, we set out to develop not only reactions that allow quantitative derivatization of all 44 Si−Cl bonds ($[\mathbf{2}]^-$) but also protocols that enable substitution of a specific number of Cl atoms by either functional groups of orthogonal reactivity ($[\mathbf{3}]^-$) or by inert substituents ($[\mathbf{4}]^-$).

**Scheme 1. (a) Syntheses of $[nBu_4N][\mathbf{2}]$−$[nBu_4N][\mathbf{4}]$, Starting from $[nBu_4N][\mathbf{1}]$; (b) Test Reactions to Shed Light on the Substituent-Exchange Mechanisms; (c) Crystallographically Determined Structures of the Silafullerane Salts $[Et_4N][\mathbf{1}]$,[1] $[nBu_4N][\mathbf{2}]$, $[nBu_4N][\mathbf{3'}]$, and $[nBu_4N][\mathbf{4}]$ in the Solid State (Cations are Omitted for Clarity; in the Structure of $[\mathbf{3'}]^-$, H and Cl are Equally Disordered over the Two Sites Marked With Red Asterisks)[a]**



[a] Reagents and conditions: (i) 100 equiv $iBu_2AlH$, $o$DFB, rt, 1 d. Yield: 50%; (ii) 40 equiv $iBu_2AlH$, $o$DFB/$Et_2O$ (1:1), rt, 18 h. Yield: 40%. (iii) 44 equiv $Me_2AlH$, 30 equiv $Me_3Al$, $o$DFB/toluene, rt, 3 d. Yield: 73%. (iv) 120 equiv $Me_3Al$, $o$DFB/toluene, rt, 1 d. (v) 80 equiv $Me_3Al$, $o$DFB/toluene, 100 °C, 21 h. (vi) 80 equiv $Me_3Al$, $o$DFB/toluene, rt, <1 h.

https://doi.org/10.1021/jacs.3c03270
J. Am. Chem. Soc. 2023, 145, 11440−11448

**Syntheses.** In a previous report, we have already disclosed that [**2**]$^-$ was formed from [$n$Bu$_4$N][**1**] and Li[AlH$_4$] in small amounts, which were, however, sufficient to detect [**2**]$^-$ by $^{29}$Si and $^1$H$^{29}$Si-HMBC NMR spectroscopy as well as LDI-MS.[1] This protocol suffers from a partial degradation of the silafullerane with concomitant formation of pyrophoric SiH$_4$, a known problem in the hydrogenation of chlorinated oligosilanes:[33,35] Li[AlH$_4$] appears to be nucleophilic enough to cleave not only Si−Cl- but also Si−Si bonds. Moreover, the Li[AlX$_4$] (X = H, Cl) byproducts are tedious to remove and therefore impede the purification of [$n$Bu$_4$N][**2**]. We now report that both obstacles can be overcome by using excess diisobutylaluminum hydride ($i$Bu$_2$AlH; 100 equiv) as a Lewis acidic and neutral H$^-$ source that substitutes all 44 exohedral Cl atoms of [$n$Bu$_4$N][**1**] with H atoms in the non-donor solvent *ortho*-difluorobenzene (*o*DFB; Scheme 1a). In this way, non-pyrophoric [$n$Bu$_4$N][**2**] was obtained in a typical yield of 50% and in the form of single crystals suitable for X-ray crystallography. The compound is significantly less soluble in *o*DFB than [$n$Bu$_4$N][**1**] but is relatively well soluble in THF without undergoing decomposition. The suitability of *o*DFB for reactions with [$n$Bu$_4$N][**1**] is particularly noteworthy: due to its combination of high dipole moment, non-basicity, high stability toward Lewis acids, and moderate boiling point,[36] it became the solvent of choice for all subsequent transformations.

We next explored the possibility of selectively replacing only the 36 Cl atoms on the sterically exposed SiCl$_3$ groups of [**1**]$^-$ with H atoms to generate the silafullerane [$n$Bu$_4$N][**3**]. An initial reaction of [$n$Bu$_4$N][**1**] with 36 equiv of $i$Bu$_2$AlH in *o*DFB again gave solely crystals of [$n$Bu$_4$N][**2**]. The selective synthesis of [$n$Bu$_4$N][**3**] was finally achieved by simply switching from pure *o*DFB as the solvent to an equimolar mixture of *o*DFB and Et$_2$O. After precipitation by addition of *n*-hexane and subsequent filtration, [$n$Bu$_4$N][**3**] was isolated as a colorless powder in 40% yield. During the optimization of the synthesis protocol, we obtained X-ray quality crystals of [$n$Bu$_4$N][Cl@Si$_{20}$(SiH$_3$)$_{12}$HCl$_7$] ([$n$Bu$_4$N][**3′**]; cf. the Supporting Information for details). In this compound, one of the eight cluster-bonded Cl atoms is replaced by an H atom. As long as single crystals of [$n$Bu$_4$N][**3**] are not available, the structure analysis of [$n$Bu$_4$N][**3′**] will serve as a substitute to discuss the main geometric parameters of mixed H/Cl-silafulleranes.

The synthesis protocol for [$n$Bu$_4$N][**4**] was inspired by a serendipitous discovery: Intending to replace $i$Bu$_2$AlH with a more easily removable aluminum hydride in the synthesis of [$n$Bu$_4$N][**2**], we treated [$n$Bu$_4$N][**1**] with excess Me$_2$AlH. Indeed, we observed the desired quantitative Cl/H exchange at the 12 silyl groups of [**1**]$^-$; at its eight SiCl moieties, however, not only Cl/H but also Cl/Me exchange took place without a clear preference. This reaction can be developed into an efficient route to [$n$Bu$_4$N][**4**] if more Me groups are supplied by using 44 equiv of Me$_2$AlH along with 30 equiv of Me$_3$Al. The product readily crystallizes from the *o*DFB/toluene solvent mixture in 73% yield.[37,38]

Having optimized the synthesis of [$n$Bu$_4$N][**4**], we subsequently performed several test reactions to shed some light on critical aspects of the underlying three-component reaction. The following results were obtained (Scheme 1b): (1) [$n$Bu$_4$N][**1**] is inert toward Me$_3$Al in the absence of Me$_2$AlH (room temperature). (2) [$n$Bu$_4$N][**2**] undergoes eightfold H/Me exchange at its core when treated with Me$_3$Al in the absence of Me$_2$AlH (100 °C).[39] (3) [$n$Bu$_4$N][**3**] undergoes eightfold Cl/Me exchange at its core when treated with Me$_3$Al in the absence of Me$_2$AlH (room temperature). From these results, the following picture emerges: The aluminum hydride not only serves as the H$^-$ source for the silyl groups but also helps to activate the system for Me transfer to the cluster core. A direct mode of action would be the formation of reactive (Me$_2$AlH)$_x$(Me$_3$Al)$_y$ aggregates. A more indirect influence could be primarily based on the initial conversion of SiCl$_3$ to SiH$_3$ because (i) precoordination of Me$_3$Al to a silyl-H atom would render Me transfer an intramolecular reaction[40] and (ii) the lower steric demand and group electronegativity of SiH$_3$ vs SiCl$_3$ should facilitate Cl$^-$ abstraction from the Si$_{20}$ core. Moreover, the higher temperature required for reaction (2) as opposed to (3) indicates that core methylation most likely occurs at Si−Cl rather than intermediately formed Si−H bonds. It is probably not a "self-healing" process that corrects "false" core hydrogenation.

To validate the experimental findings and to gain deeper insights into the SiCl$_3$ vs core-SiCl hydrogenation mechanisms, quantum-chemical calculations were conducted. Theoretical considerations regarding the three-component reaction for regioselective hydrogenation/methylation are not discussed as the nature of the actual active H/Me transfer reagent(s) is unknown and the number and sizes of potential candidates are prohibitively high.

**Quantum-Chemical Calculations on the Cl/H Exchange Mechanisms.** Free-energy calculations were performed at the r$^2$SCAN0-D4[41−43]/ma-def2-QZVPP[44,45]+COSMO-RS(CH$_2$Cl$_2$)[46,47]+G$_{\text{mRRHO}}$[48,49]//PBEh-3c[50](SMD(CH$_2$Cl$_2$))[51] level using the xtb 6.5.1,[52,53] ORCA 5.0.3,[54−56] and TURBOMOLE 7.6.0[57] program packages (cf. the Supporting Information for further details). To keep the computational effort within reasonable limits, we studied substitution reactions at the exohedral SiCl$_3$ groups and at the SiCl cluster vertices by using the model systems [**M1**]$^-$ and [**M2**]$^-$, respectively (Figure 2).

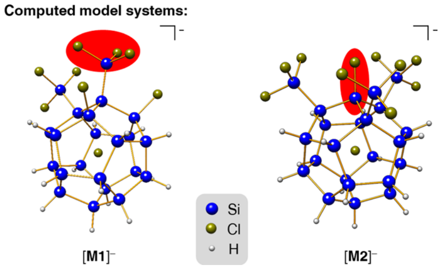
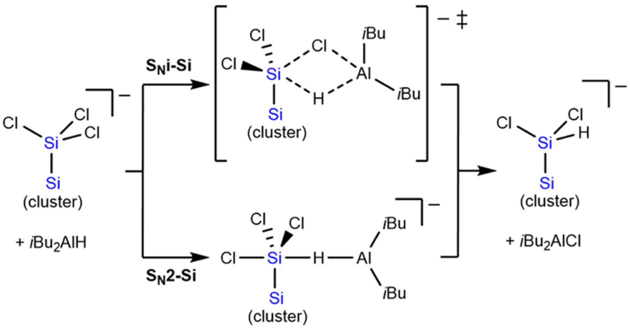


**Figure 2.** Model systems used for the computational studies of substitution reactions at the exohedral SiCl$_3$ groups ([**M1**]$^-$) and at the SiCl cluster vertices ([**M2**]$^-$). The functional groups under investigation are highlighted in red.

In [**M1**]$^-$/[**M2**]$^-$, we consider one Si−SiCl$_3$/Si−Cl moiety and treat only their three adjacent vertices explicitly, while the rest of the cluster is simplified to 16 Si−H vertices.

Two possible mechanisms for the hydrogenation of chlorosilanes using $i$Bu$_2$AlH are known in the literature.[58−61] In Scheme 2, these are illustrated by the example of an exohedral SiCl$_3$ group of [**1**]$^-$: The so-called "intramolecular nucleophilic substitution mechanism" (S$_\text{N}$i-Si) involves a four-membered transition state, whereas the S$_\text{N}$2-Si mechanism

**Scheme 2. Hydrogenation Following the $S_N$i-Si or $S_N$2-Si Mechanism at an Exohedral $SiCl_3$ Group**



includes a trigonal-bipyramidal structure. The latter resembles the transition state of an $S_N2$ reaction at a carbon center but is a true intermediate because the silicon atom can become hypercoordinated.

Since the potential energy curves around these intermediates are very shallow, $S_N$2-Si transition states connecting them with reactants and products could not be located. However, exactly because the energy differences between the transition states and the intermediates must be small in the present cases, we can use the intermediates' energies to gain insights into the favored reaction mechanisms even without knowledge of the transition states. While the $S_N$i-Si mechanism is geometrically possible at both the $SiCl_3$ and SiCl functionalities, $S_N$2-Si hydrogenation can only take place at $SiCl_3$ because the backside attack at SiCl is blocked by the cluster core.[62] A comparison of the $S_N$i-Si barrier heights already reveals that the barrier for SiCl hydrogenation is almost twice as large as that for $SiCl_3$ hydrogenation [$\Delta G^{\ddagger}$(TS2) = 21.4 kcal mol$^{-1}$, $\Delta G^{\ddagger}$(TS4) = 11.5 kcal mol$^{-1}$ $\Delta\Delta G^{\ddagger}$ = 9.9 kcal mol$^{-1}$; Figure 3]. The free-energy difference $\Delta\Delta G^{\ddagger}$ for Cl/H exchange at the two positions increases even to 12.4 kcal mol$^{-1}$, if we take into account that $S_N$2-Si hydrogenation at $SiCl_3$ is slightly more favorable than $S_N$i-Si hydrogenation [cf. $\Delta G$(INT5) = 9.0 kcal mol$^{-1}$ vs $\Delta G^{\ddagger}$(TS4) = 11.5 kcal mol$^{-1}$)]: in the $S_N2$ intermediate, $iBu_2AlH$ is precoordinated to the neighboring SiCl vertex and is therefore already in close vicinity to the reacting $SiCl_3$ group. Most importantly, the barrier height for SiCl hydrogenation decreases from $\Delta G^{\ddagger}$(TS2) = 21.4 to $\Delta G^{\ddagger}$(TS2.1) = 17.6 kcal mol$^{-1}$ once the $SiCl_3$ groups are fully hydrogenated, which explains why the cluster core is also eventually perhydrogenated to afford [**2**]$^-$. Addition of $Et_2O$ to the reaction mixture leads to the formation of an $iBu_2AlH$-$OEt_2$ adduct, which not only increases the steric demand of the hydrogenation reagent but also stabilizes the adduct by 16.6 kcal mol$^{-1}$ compared to free $iBu_2AlH$. As a result, both the $SiCl_3$ and the SiCl hydrogenations are now associated with a higher energy penalty [$\Delta G$(INT3) = 17.3 kcal mol$^{-1}$, $\Delta G^{\ddagger}$(TS1) = 34.1 kcal mol$^{-1}$, $\Delta\Delta G^{\ddagger}$ = 16.8 kcal mol$^{-1}$], but the $\Delta\Delta G^{\ddagger}$ value further increases by 4.4 kcal mol$^{-1}$. While INT3 is still accessible at room temperature, TS1 lies energetically too high to be overcome under these conditions. Even after the exhaustive conversion of $SiCl_3$ to $SiH_3$, hydrogenation at the SiCl vertices still has a barrier of $\Delta G^{\ddagger}$(TS1.1) = 32.8 kcal mol$^{-1}$, so that this reaction is largely suppressed at room temperature. Taken together, this explains why our synthesis protocol gives essentially pure samples of [**3**]$^-$ with very little contamination from singly core-hydrogenated [**3′**]$^-$ (Scheme 1c).

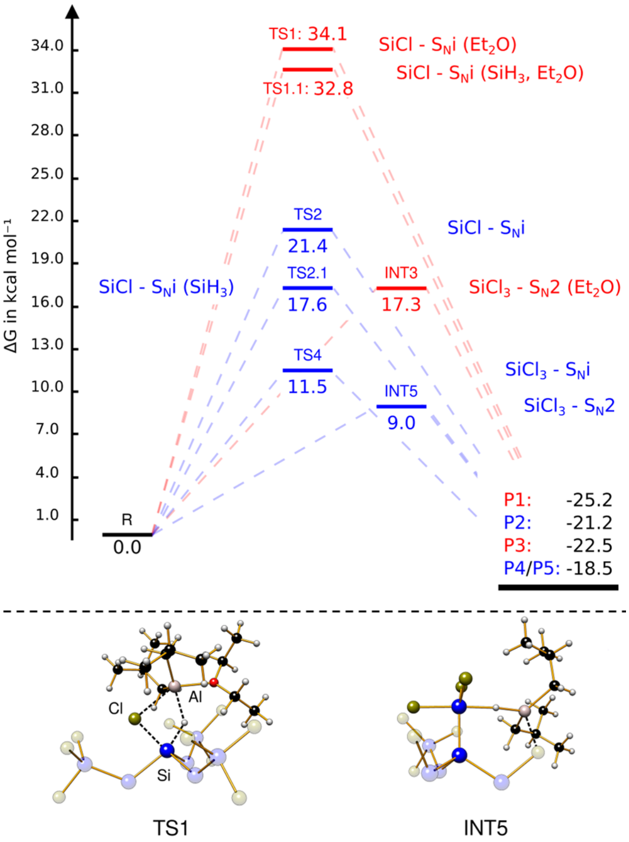


**Figure 3.** Calculated Gibbs free energies $\Delta G$ for the hydrogenation at the r$^2$SCAN0-D4/ma-def2-QZVPP+COSMO-RS-($CH_2Cl_2$)+$G_{mRRHO}$//PBEh-3c(SMD($CH_2Cl_2$)) level of theory. R represents the separated reactants which are [**M1**]$^-$ or [**M2**]$^-$ in combination with the respective nucleophile. P represents the products obtained after a single hydrogenation step via the transition state or intermediates with the same number.

**X-ray Crystal Structure Analyses, IR Spectroscopy, and LDI Mass Spectrometry.** Scheme 1c shows the anion structures of [$nBu_4N$][**2**], [$nBu_4N$][**3′**], and [$nBu_4N$][**4**] in the solid state (the structure of the known [**1**]$^-$ was added for completeness). The proposed substitution patterns are fully confirmed. Careful analysis of the residual electron density in the regions near the SiMe and $SiH_3$ groups of [**4**]$^-$ (based on high-quality diffraction data obtained with synchrotron radiation) provides no evidence for a possible Me/H crystallographic disorder. All anions lie on inversion centers; Si−H/Cl disorder in [**3′**]$^-$ occurs only at two equivalent positions (marked with red asterisks). The Si−Si bond lengths of the three anions fall in the range of 2.3200(5)−2.381(6) Å.[63] For an assessment of the different propensities of differently substituted SiR cluster vertices to accept charge density from the endohedral $Cl^-$ ion, the $Cl^-\cdots Si$ distances within each individual cluster are particularly diagnostic (Figure 4a; blue bars: range of experimental values, red lines: computed values). For [**2**]$^-$, the ranges of Cl···SiH and Cl···$SiSiH_3$ distances are overlapping. For [**3′**]$^-$ and [**4**]$^-$, the Cl···SiCl[64] and Cl···SiMe distances are shorter and longer, respectively, than the corresponding Cl···$SiSiH_3$ distances within the same molecule. This ordering fits to the expected trend in the acceptor-orbital energies according to $\sigma^*$(Si−Me) > $\sigma^*$(Si−$SiH_3$) ≈ $\sigma^*$(Si−H) > $\sigma^*$(Si−Cl).

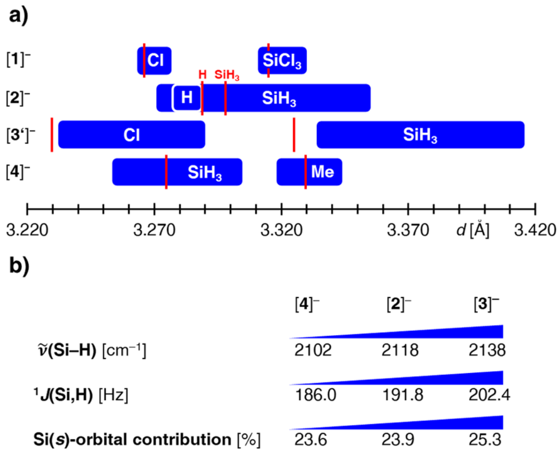


**Figure 4.** (a) Cl···SiR distances between the endohedral $Cl^-$ ions and the vertices of the $Si_{20}$ cages in $[\mathbf{1}]^-$, $[\mathbf{2}]^-$, $[\mathbf{3'}]^-$, and $[\mathbf{4}]^-$ (blue bars: experimental ranges within $3\sigma$ margins, white font: specification of R; red lines: computed values). (b) Correlations between the experimentally obtained stretching frequencies [$\tilde{\nu}$(Si–H)] and NMR coupling constants [$^1J$(Si,H)] with the computed Si($s$) orbital contribution to the $H_2Si$–H bonds in $[\mathbf{4}]^-$, $[\mathbf{2}]^-$, and $[\mathbf{3}]^-$.

In the solid-state IR spectra, bands at $\tilde{\nu}$(Si–H) = 2118 $cm^{-1}$ ([$n$Bu$_4$N][**2**]; calcd 2160 $cm^{-1}$), 2138 $cm^{-1}$ ([$n$Bu$_4$N][**3**]; calcd 2181 $cm^{-1}$), and 2102 $cm^{-1}$ ([$n$Bu$_4$N][**4**]; calcd 2149 $cm^{-1}$) are assigned to the Si–H stretching modes of the $SiH_3$ groups. Only in the case of [$n$Bu$_4$N][**2**], additional, less intense Si–H stretching bands, originating from the SiH vertices, were detected ($\tilde{\nu}$(Si–H) = 2042 $cm^{-1}$ and 2026 $cm^{-1}$, calcd 2101 $cm^{-1}$; cf. [$n$Bu$_4$N][Cl@$Si_{20}H_{20}$]:[32] $\tilde{\nu}$(Si–H) = 2064 $cm^{-1}$). Since IR wavenumbers correlate directly with the force constants and these, in turn, with the $s$ character of the respective bonds, the experimentally observed trend in the $H_2Si$–H stretching modes indicates an increase in the $s$ character in the order $[\mathbf{4}]^- < [\mathbf{2}]^- < [\mathbf{3}]^-$ (Figure 4b). This claim is further supported by a comparison of the corresponding $^1J$(Si,H) NMR coupling constants, which also grow larger along the sequence $[\mathbf{4}]^- < [\mathbf{2}]^- < [\mathbf{3}]^-$, thereby testifying to an increasing Fermi contact interaction. Theoretical validation comes from Bent's rule[65] and a density functional theory (DFT)-based natural localized molecular orbital (NLMO) analysis conducted with the NBO 6.0 program,[66,67] which predicts Si($s$) orbital contributions to the $H_2Si$–H bonds of 23.6% < 23.9% < 25.3%.

LDI-MS(−) measurements on [$n$Bu$_4$N][**2**]–[$n$Bu$_4$N][**4**] gave the molecular ion peaks $[M]^-$ of $[\mathbf{2}]^-$–$[\mathbf{4}]^-$ with fitting isotope patterns. Similar to $[\mathbf{1}]^-$, which undergoes gas-phase elimination of $SiCl_2$ units under LDI-MS conditions,[1] we now see evidence of $SiH_2$ elimination/insertion (−/+ 30 Da).

**NMR Spectroscopy.** NMR spectra of [$n$Bu$_4$N][**2**] and [$n$Bu$_4$N][**4**] were recorded in THF-$d_8$, whereas [$n$Bu$_4$N][**3**] had to be characterized in $o$DFB/$C_6D_{12}$ (5:1 mixture) due to its lability to donor solvents. The Si$H_3$ and Si$H$ moieties of [$n$Bu$_4$N][**2**] give rise to a singlet at 3.47 ppm (36H) and a 1:1:1:1 quartet at 4.31 ppm [8H; $J$(H,Cl) = 2.9 Hz], respectively. Upon $^{35}$Cl decoupling, the quartet collapses into a singlet with a broadened base [due to unresolved coupling to $^{37}$Cl; $S(^{35}Cl) = S(^{37}Cl) = 3/2$, natural abundances: 75.5 and 24.5%].[68,69] The exclusively peripheral hydrogenation in $[\mathbf{3}]^-$ is evident from the absence of core-Si$H$ resonances. The complete core methylation in $[\mathbf{4}]^-$ is reflected by the Si$H_3$ vs core-Si$Me$ integral ratio of 36H vs 24H.

In its $^{29}$Si{$^1$H} NMR spectrum, each of the $T_h$-symmetric silafulleranes $[\mathbf{2}]^-$–$[\mathbf{4}]^-$ shows three resonances for the chemically inequivalent $Si^0$, $Si^I$, and $Si^{III}$ centers. The corresponding chemical shifts agree well with the computed values (Table 1).[70,71]

The endohedral $Cl^-$ ions within the highly symmetric siladodecahedrane cages of $[Cl@Si_{20}R_{20}]^-$ (R = H, Cl) give rise to sharp $^{35}$Cl resonances.[32] Based on quantum-chemical analysis of the shielding components, we have shown that $\delta(^{35}Cl)$ is a useful probe of the degree of the $Cl^- \rightarrow Si_{20}$ host-guest interaction: a downfield shift of the $^{35}$Cl signal is diagnostic for a decreased $Cl^- \rightarrow Si_{20}$ interaction, as illustrated by a comparison of $[Cl@Si_{20}Cl_{20}]^-$ (126.0 ppm; more electronegative substituents) and $[Cl@Si_{20}H_{20}]^-$ (345.0 ppm; less electronegative substituents).[32] In line with that, the presence of even less electronegative $SiH_3$ substituents[75] in $[\mathbf{2}]^-$ is accompanied by a further downfield shift to $\delta(^{35}Cl)$ = 469.0 (Table 1; cf. also $[\mathbf{4}]^-$). $\delta(^{35}Cl)$ = 363.7 of the partly chlorinated $[\mathbf{3}]^-$ lies between the shifts of $[\mathbf{2}]^-$ and exhaustively chlorinated $[\mathbf{1}]^-$ (274.5 ppm).[32] Our measured $\delta(^{35}Cl)$ values were verified by spin–orbit relativistic DFT calculations,[72−74,76−79] which gave an excellent fit after scaling according to the following linear equation: $\delta(^{35}Cl, \text{scaled}) = 0.8739 \cdot \delta(^{35}Cl, \text{calcd}) - 8.3536$ (cf. the Supporting Information for more details).

As discussed above, our early attempts at the synthesis of [$n$Bu$_4$N][**4**] from [$n$Bu$_4$N][**1**] and $Me_2AlH$ (without $Me_3Al$) resulted in a product mixture. In the $^{35}$Cl{$^1$H} NMR spectrum, this mixture gave rise to a multitude of resonances in the range of 469.1–456.8 ppm. Since the two range limits fit to the shift values of [$n$Bu$_4$N][**2**] (8 core-SiH) and [$n$Bu$_4$N][**4**] (8 core-SiMe; Table 1), it is reasonable to assume that the resonances in between belong to mixed-substituted silafulleranes [$n$ core-SiH + (8−$n$) core-SiMe; $n$ = 1−7]. On the other hand, this

**Table 1. Experimental (Calculated) $^{29}$Si and $^{35}$Cl NMR Spectroscopic Parameters of the Silafulleranes [$n$Bu$_4$N][1]–[$n$Bu$_4$N][4]**[a,b]

| | $\delta(^{29}Si)$ | | | $\delta(^{35}Cl)$ |
|---|---|---|---|---|
| compound | $Si^0$ | $Si^I$ | $Si^{III}$ | |
| [$n$Bu$_4$N][**1**] | −60.3 (−63.5) | 31.1 (31.7) | 10.3 (15.9) | 274.5 (277.6)[c] |
| [$n$Bu$_4$N][**2**] | −58.5 (−58.3) | −14.9 (−23.9) | −93.7 (−98.6) | 469.0 (471.8)[c] |
| [$n$Bu$_4$N][**3**] | −73.4 (−72.6) | 51.5 (54.3) | −97.7 (−98.5) | 363.7 (360.0)[c] |
| [$n$Bu$_4$N][**4**] | −67.2 (−67.3) | 19.1 (21.3) | −100.6 (−102.4) | 457.1 (455.8)[c] |

[a] Calculated NMR shifts have been computed at the SO-ZORA-PBE0[72](COSMO($CH_2Cl_2$))[73]/ZORA/TZP[74] level of theory. [b] NMR spectra were recorded in THF-$d_8$ ([$n$Bu$_4$N][**1**],[1,32][$n$Bu$_4$N][**2**], [$n$Bu$_4$N][**4**]), or $o$DFB/$C_6D_{12}$ (5:1; [$n$Bu$_4$N][**3**]). [c] These values were obtained after scaling according to the following linear equation: $\delta(^{35}Cl, \text{scaled}) = 0.8739 \cdot \delta(^{35}Cl, \text{calcd}) - 8.3536$.

conclusion is at odds with previous experience gained from $[Cl@Si_{20}(H/Cl)_{20}]^-$ silafulleranes, where any symmetry breaking had resulted in such severely broadened $^{35}Cl$ signals that they were no longer detectable. One possible solution of this riddle would be that unsymmetrically distributed core-SiH/SiMe leads to a smaller electric field gradient (EFG) than unsymmetrically distributed core-SiH/SiCl. A smaller EFG should accelerate the relaxation of the $^{35}Cl$ quadrupole nucleus to a lesser extent, resulting in a narrower resonance line.

The influence of the substitution pattern on the EFG within a silafullerane cage was investigated by DFT calculations on selected derivatives. To a good approximation, the NMR line width $\Delta\nu$ is proportional to the square of the largest principal component of the EFG tensor, $V_{33}^2$ (cf. the Supporting Information for more details).[80−82] For $T_h$-symmetric $[Cl@Si_{20}H_{12}Cl_8]^-$ and $[\mathbf{4}]^-$, the computed $V_{33}^2$ values for the endohedral $Cl^-$ ion are close to 0[83] (compare the exohedral Cl substituents of $[Cl@Si_{20}H_{12}Cl_8]^-$ with $V_{33}^2$ = 4.44 a.u.). For the less symmetric isomer of $[Cl@Si_{20}H_{12}Cl_8]^-$, in which one H and one Cl atom have changed places, and for $[Cl@Si_{20}H_{11}Cl_9]^-$, in which one H atom is replaced by one Cl atom, $V_{33}^2$ increases to values in the order of $1 \times 10^{-2}$ a.u. In stark contrast, 100-fold smaller values $V_{33}^2 \leq 1 \times 10^{-4}$ a.u. were computed for an isomer of $[\mathbf{4}]^-$, in which one cage-bonded Me group and one silyl-bonded H atom have changed places, and for $[Cl@Si_{20}(SiH_3)_{12}HMe_7]^-$, in which one cage-bonded Me group is replaced by one H atom. Therefore, it should indeed be more likely to detect $^{35}Cl$ NMR signals of $[\mathbf{4}]^-$-type than that of $[Cl@Si_{20}H_{12}Cl_8]^-$-type compounds with slight deviations from the perfect $T_h$-symmetric substitution pattern.

## CONCLUSIONS

Starting from the perchlorinated silafullerane $[nBu_4N][Cl@Si_{20}(SiCl_3)_{12}Cl_8]$, we developed facile synthesis routes to the perhydrogenated $[nBu_4N][Cl@Si_{20}(SiH_3)_{12}H_8]$ ($[nBu_4N][\mathbf{2}]$; $i$Bu$_2$AlH in $o$DFB), the partially hydrogenated $[nBu_4N][Cl@Si_{20}(SiH_3)_{12}Cl_8]$ ($[nBu_4N][\mathbf{3}]$; $i$Bu$_2$AlH in $o$DFB/Et$_2$O), and the regioselectively mixed-substituted $[nBu_4N][Cl@Si_{20}(SiH_3)_{12}Me_8]$ ($[nBu_4N][\mathbf{4}]$; Me$_2$AlH/Me$_3$Al in $o$DFB/toluene, $o$DFB = *ortho*-difluorobenzene). According to quantum-chemical calculations, SiCl$_3$ hydrogenation (via an S$_N$2-Si-type scenario) is kinetically more favorable than core-SiCl hydrogenation (via an S$_N$i-Si-type transition state). This explains why taming the reactivity of $i$Bu$_2$AlH through Et$_2$O adduct formation suppresses core hydrogenation and ultimately furnishes $[\mathbf{3}]^-$ instead of $[\mathbf{2}]^-$. Experimentally obtained structural and spectroscopic parameters of $[\mathbf{2}]^-$−$[\mathbf{4}]^-$ are in very good agreement with computed values. A theoretical assessment of the EFGs experienced by the encapsulated $Cl^-$ ions of selected silafulleranes helps to understand why the corresponding $^{35}Cl$ resonances of $[\mathbf{4}]^-$-type compounds are detectable even if the strict $T_h$ symmetry is violated, whereas these signals are broadened beyond detection in comparable symmetry-broken H/Cl-mixed silafulleranes.

As an outlook, the Si−H bonds of $[\mathbf{2}]^-$−$[\mathbf{4}]^-$ are valuable functional groups for subsequent hydrosilylation[84,85] or dehydrogenative Si−Si coupling reactions.[86,87] Resulting silafullerane oligomers would be nano-sized compounds bridging the gap between small-molecule silanes and extended silicon-based solids.[88−91]

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.3c03270.

Experimental and computational details and characterization data (PDF)

Cartesian coordinates of computed molecular structures (ZIP)

### Accession Codes

CCDC 2250348−2250350 contain the supplementary crystallographic data for this paper. These data can be obtained free of charge via www.ccdc.cam.ac.uk/data_request/cif, or by emailing data_request@ccdc.cam.ac.uk, or by contacting The Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK; fax: +44 1223 336033.

## AUTHOR INFORMATION

### Corresponding Authors

**Markus Bursch** − *Max-Planck-Institut für Kohlenforschung, 45470 Mülheim an der Ruhr, Germany;* orcid.org/0000-0001-6711-5804; Email: bursch@kofo.mpg.de

**Matthias Wagner** − *Institut für Anorganische und Analytische Chemie, Goethe-Universität Frankfurt am Main, 60438 Frankfurt am Main, Germany;* orcid.org/0000-0001-5806-8276; Email: matthias.wagner@chemie.uni-frankfurt.de

### Authors

**Marcel Bamberg** − *Institut für Anorganische und Analytische Chemie, Goethe-Universität Frankfurt am Main, 60438 Frankfurt am Main, Germany;* orcid.org/0000-0001-6044-5077

**Thomas Gasevic** − *Mulliken Center for Theoretical Chemistry, Clausius-Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, 53115 Bonn, Germany;* orcid.org/0000-0003-4864-1758

**Michael Bolte** − *Institut für Anorganische und Analytische Chemie, Goethe-Universität Frankfurt am Main, 60438 Frankfurt am Main, Germany*

**Alexander Virovets** − *Institut für Anorganische und Analytische Chemie, Goethe-Universität Frankfurt am Main, 60438 Frankfurt am Main, Germany;* orcid.org/0000-0002-8843-8503

**Hans-Wolfram Lerner** − *Institut für Anorganische und Analytische Chemie, Goethe-Universität Frankfurt am Main, 60438 Frankfurt am Main, Germany;* orcid.org/0000-0003-1803-7947

**Stefan Grimme** − *Mulliken Center for Theoretical Chemistry, Clausius-Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, 53115 Bonn, Germany;* orcid.org/0000-0002-5844-4371

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.3c03270

### Funding

M.W. thanks the Deutsche Forschungsgemeinschaft for financial support (DFG grant no. 506550642). S.G. and M.Bu. gratefully acknowledge financial support by the Max Planck Society through the Max Planck fellow program. M.Ba. wishes to thank the Fonds der Chemischen Industrie (FCI) for a Kekulé Ph.D. grant.

## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

S.G., M.Bu., and T.G. thank Christoph Plett for fruitful discussion. The authors are grateful to Evonik Operations GmbH, Rheinfelden (Germany), for the generous donation of $Si_2Cl_6$. Parts of this research (project I-20220865) were carried out at PETRA III at DESY, a member of the Helmholtz Association (HGF). A.V. and M.Ba. thank Dr. Leila Noohinejad, Dr. Martin Tolkiehn, and Dr. Eugenia Peresypkina for their assistance regarding the use of the beamline P24. A.V. thanks Dr. Matthias Meyer (Rigaku Oxford Diffraction) for his precious help with the implementation of the CrysAlisPro software for the synchrotron and STOE IPDS II diffraction data. M.Ba. wishes to thank Julius Wicke for preparative support.

## ■ REFERENCES

(1) Tillmann, J.; Wender, J. H.; Bahr, U.; Bolte, M.; Lerner, H.-W.; Holthausen, M. C.; Wagner, M. One-Step Synthesis of a [20]-Silafullerane with an Endohedral Chloride Ion. *Angew. Chem., Int. Ed.* **2015**, *54*, 5429−5433.

(2) Nagase, S. Polyhedral Compounds of the Heavier Group 14 Elements: Silicon, Germanium, Tin, and Lead. *Acc. Chem. Res.* **1995**, *28*, 469−476.

(3) Pichierri, F.; Kumar, V.; Kawazoe, Y. Encapsulation of halide anions in perhydrogenated silicon fullerene: $X^-@Si_{20}H_{20}$ (X = F, Cl, Br, I). *Chem. Phys. Lett.* **2005**, *406*, 341−344.

(4) Marsusi, F.; Qasemnazhand, M. Sila-fulleranes: promising chemically active fullerene analogs. *Nanotechnology* **2016**, *27*, 275704.

(5) De, D. S.; Schaefer, B.; von Issendorff, B.; Goedecker, S. Nonexistence of the decahedral $Si_{20}H_{20}$ cage: Levinthal's paradox revisited. *Phys. Rev. B* **2020**, *101*, 214303.

(6) Our group has developed the $Si_2Cl_6/Cl^-$ system as a versatile trichlorosilylation reagent, which enables the synthesis of a wide range of silicon compounds such as $Cl^-$-coordinated cyclohexasilanes, anionic organosilanes, germanides, and even mixed Si/Ge or Si/Sn heteroadamantanes: (a) Teichmann, J.; Wagner, M. Silicon chemistry in zero to three dimensions: from dichlorosilylene to silafullerane. *Chem. Commun.* **2018**, *54*, 1397−1412. (b) Tillmann, J.; Meyer, L.; Schweizer, J. I.; Bolte, M.; Lerner, H.-W.; Wagner, M.; Holthausen, M. C. Chloride-Induced Aufbau of Perchlorinated Cyclohexasilanes from $Si_2Cl_6$: A Mechanistic Scenario. *Chem.—Eur. J.* **2014**, *20*, 9234−9239. (c) Köstler, B.; Bae, H.; Gilmer, J.; Virovets, A.; Lerner, H.-W.; Albert, P.; Fantuzzi, F.; Wagner, M. Dope it with germanium: selective access to functionalized $Si_5Ge$ heterocycles. *Chem. Commun.* **2023**, *59*, 716−719. (d) Georg, I.; Teichmann, J.; Bursch, M.; Tillmann, J.; Endeward, B.; Bolte, M.; Lerner, H.-W.; Grimme, S.; Wagner, M. Exhaustively Trichlorosilylated $C_1$ and $C_2$ Building Blocks: Beyond the Müller−Rochow Direct Process. *J. Am. Chem. Soc.* **2018**, *140*, 9696−9708. (e) Georg, I.; Bursch, M.; Stückrath, J. B.; Alig, E.; Bolte, M.; Lerner, H.-W.; Grimme, S.; Wagner, M. Building up Strain in One Step: Synthesis of an Edge-Fused Double Silacyclobutene from an Extensively Trichlorosilylated Butadiene Dianion. *Angew. Chem., Int. Ed.* **2020**, *59*, 16181−16187. (f) Georg, I.; Bursch, M.; Endeward, B.; Bolte, M.; Lerner, H.-W.; Grimme, S.; Wagner, M. The power of trichlorosilylation: isolable trisilylated allyl anions, allyl radicals, and allenyl anions. *Chem. Sci.* **2021**, *12*, 12419−12428. (g) Teichmann, J.; Kunkel, C.; Georg, I.; Moxter, M.; Santowski, T.; Bolte, M.; Lerner, H.-W.; Bade, S.; Wagner, M. Tris(trichlorosilyl)tetrelide Anions and a Comparative Study of Their Donor Qualities. *Chem.—Eur. J.* **2019**, *25*, 2740−2744. (h) Köstler, B.; Bolte, M.; Lerner, H.-W.; Wagner, M. Selective One-Pot Syntheses of Mixed Silicon-Germanium Heteroadamantane Clusters. *Chem.—Eur. J.* **2021**, *27*, 14401−14404. (i) Köstler, B.; Gilmer, J.; Bolte, M.; Virovets, A.; Lerner, H.-W.; Albert, P.; Fantuzzi, F.; Wagner, M. Group IV heteroadamantanes: synthesis of $Si_6Sn_4$ and site-selective derivatization of $Si_6Ge_4$. *Chem. Commun.* **2023**, *59*, 2295−2298.

(7) Heider, Y.; Scheschkewitz, D. Molecular Silicon Clusters. *Chem. Rev.* **2021**, *121*, 9674−9718.

(8) Wiberg, N.; Finger, C. M. M.; Polborn, K. Tetrakis(tri-*tert*-butylsilyl)-*tetrahedro*-tetrasilane ($tBu_3Si)_4Si_4$: The First Molecular Silicon Compound with a $Si_4$ Tetrahedron. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1054−1056.

(9) Ichinohe, M.; Toyoshima, M.; Kinjo, R.; Sekiguchi, A. Tetrasilatetrahedranide: A Silicon Cage Anion. *J. Am. Chem. Soc.* **2003**, *125*, 13328−13329.

(10) Matsumoto, H.; Higuchi, K.; Hoshino, Y.; Koike, H.; Naoi, Y.; Nagai, Y. The First Octasilacubane System: Synthesis of Octakis-(t-butyldimethylsilyl)pentacyclo[$4.2.0.0^{2,5}.0^{3,8}.0^{4,7}$]-octasilane. *J. Chem. Soc., Chem. Commun.* **1988**, *3*, 1083−1084.

(11) Furukawa, K.; Fujino, M.; Matsumoto, N. Cubic Silicon Cluster. *Appl. Phys. Lett.* **1992**, *60*, 2744−2745.

(12) Fischer, J.; Baumgartner, J.; Marschner, C. Synthesis and Structure of Sila-Adamantane. *Science* **2005**, *310*, 825.

(13) Siu, T. C.; Imex Aguirre Cardenas, M.; Seo, J.; Boctor, K.; Shimono, M. G.; Tran, I. T.; Carta, V.; Su, T. A. Site-Selective Functionalization of Sila-Adamantane and Its Ensuing Optical Effects. *Angew. Chem., Int. Ed.* **2022**, *61*, e202206877.

(14) Tsurusaki, A.; Koyama, Y.; Kyushin, S. Decasilahexahydrotriquinacene and Decasilaisotwistane: $\sigma$ Conjugation on a Bowl Surface. *J. Am. Chem. Soc.* **2017**, *139*, 3982−3985.

(15) Unno, M.; Higuchi, K.; Ida, M.; Shioyama, H.; Kyushin, S.; Matsumoto, H.; Goto, M. Ring-Opening Reaction of Octakis(1,1,2-trimethylpropyl)octasilacubane. Chlorination with $PCl_5$ Leading to Stereoisomeric 4,8-Dichlorooctakis(1,1,2-trimethylpropyl)-tetracyclo-[$3.3.0.0^{2,7}.0^{3,6}$]octasilanes. *Organometallics* **1994**, *13*, 4633−4640.

(16) Wiberg, N.; Auer, H.; Nöth, H.; Knizek, J.; Polborn, K. Diiodotetrasupersilylcyclotetrasilene $(tBu_3Si)_4Si_4I_2$—A Molecule Containing an Unsaturated $Si_4$ Ring. *Angew. Chem., Int. Ed.* **1998**, *37*, 2869−2872.

(17) Otsuka, K.; Matsumoto, N.; Ishida, S.; Kyushin, S. An Isolable Radical Anion of an Organosilicon Cluster Containing Only $\sigma$ Bonds. *Angew. Chem., Int. Ed.* **2015**, *54*, 7833−7836.

(18) Schnepf, A.; Schnöckel, H. Metalloid Aluminum and Gallium Clusters: Element Modifications on the Molecular Scale? *Angew. Chem., Int. Ed.* **2002**, *41*, 3532−3554.

(19) Scheschkewitz, D. A Molecular Silicon Cluster with a "Naked" Vertex Atom. *Angew. Chem., Int. Ed.* **2005**, *44*, 2954−2956.

(20) Heider, Y.; Poitiers, N. E.; Willmes, P.; Leszczyńska, K.; Huch, V.; Scheschkewitz, D. Site-selective functionalization of $Si_6R_6$ siliconoids. *Chem. Sci.* **2019**, *10*, 4523−4530.

(21) Akasaka, N.; Ishida, S.; Iwamoto, T. Transformative $Si_8R_8$ Siliconoids. *Inorganics* **2018**, *6*, 107.

(22) Iwamoto, T.; Akasaka, N.; Ishida, S. A heavy analogue of the smallest bridgehead alkene stabilized by a base. *Nat. Commun.* **2014**, *5*, 5353.

(23) Keuter, J.; Schwedtmann, K.; Hepp, A.; Bergander, K.; Janka, O.; Doerenkamp, C.; Eckert, H.; Mück-Lichtenfeld, C.; Lips, F. Diradicaloid or Zwitterionic Character: The Non-Tetrahedral Unsaturated Compound $[Si_4\{N(SiMe_3)Dipp\}_4]$ with a Butterfly-Type $Si_4$ Substructure. *Angew. Chem., Int. Ed.* **2017**, *56*, 13866−13871.

(24) Keuter, J.; Schwermann, C.; Hepp, A.; Bergander, K.; Droste, J.; Hansen, M. R.; Doltsinis, N. L.; Mück-Lichtenfeld, C.; Lips, F. A highly unsaturated six-vertex amido-substituted silicon cluster. *Chem. Sci.* **2020**, *11*, 5895−5901.

(25) Scharfe, S.; Kraus, F.; Stegmaier, S.; Schier, A.; Fässler, T. F. Zintl Ions, Cage Compounds, and Intermetalloid Clusters of Group 14 and Group 15 Elements. *Angew. Chem., Int. Ed.* **2011**, *50*, 3630−3670.

(26) Waibel, M.; Kraus, F.; Scharfe, S.; Wahl, B.; Fässler, T. F. $[(MesCu)_2(\eta^3\text{-}Si_4)]^{4-}$: A Mesitylcopper-Stabilized Tetrasilicide Tetraanion. *Angew. Chem., Int. Ed.* **2010**, *49*, 6611−6615.

(27) Schiegerl, L. J.; Karttunen, A. J.; Klein, W.; Fässler, T. F. Silicon clusters with six and seven unsubstituted vertices *via* a two-step reaction from elemental silicon. *Chem. Sci.* **2019**, *10*, 9130−9139.

(28) Goicoechea, J. M.; Sevov, S. C. $[(\mathrm{Ni\text{-}Ni\text{-}Ni})@(\mathrm{Ge}_9)_2]^{4-}$: A Linear Triatomic Nickel Filament Enclosed in a Dimer of Nine-Atom Germanium Clusters. *Angew. Chem., Int. Ed.* **2005**, *44*, 4026−4028.

(29) Goicoechea, J. M.; Sevov, S. C. Deltahedral Germanium Clusters: Insertion of Transition-Metal Atoms and Addition of Organometallic Fragments. *J. Am. Chem. Soc.* **2006**, *128*, 4155−4161.

(30) Joseph, S.; Hamberger, M.; Mutzbauer, F.; Härtl, O.; Meier, M.; Korber, N. Chemistry with Bare Silicon Clusters in Solution: A Transition-Metal Complex of a Polysilicide Anion. *Angew. Chem., Int. Ed.* **2009**, *48*, 8770−8772.

(31) Neumeier, M.; Fendt, F.; Gärtner, S.; Koch, C.; Gärtner, T.; Korber, N.; Gschwind, R. M. Detection of the Elusive Highly Charged Zintl Ions $\mathrm{Si}_4^{4-}$ and $\mathrm{Sn}_4^{4-}$ in Liquid Ammonia by NMR Spectroscopy. *Angew. Chem., Int. Ed.* **2013**, *52*, 4483−4486.

(32) Bamberg, M.; Bursch, M.; Hansen, A.; Brandl, M.; Sentis, G.; Kunze, L.; Bolte, M.; Lerner, H.-W.; Grimme, S.; Wagner, M. $[\mathrm{Cl}@\mathrm{Si}_{20}\mathrm{H}_{20}]^-$: Parent Siladodecahedrane with Endohedral Chloride Ion. *J. Am. Chem. Soc.* **2021**, *143*, 10865−10871.

(33) Gerwig, M.; Böhme, U.; Friebel, M.; Gründler, F.; Franze, G.; Rosenkranz, M.; Schmidt, H.; Kroke, E. Syntheses and Molecular Structures of Liquid Pyrophoric Hydridosilanes. *ChemistryOpen* **2020**, *9*, 762−773.

(34) For examples of structurally authenticated dendritic poly(methylsilane)s, see: (a) Lambert, J. B.; Wu, H. Synthesis and Crystal Structure of a Nanometer-Scale Dendritic Polysilane. *Organometallics* **1998**, *17*, 4904−4909. (b) Krempner, C.; Köckerling, M. Nanoscale Double-Core Oligosilane Dendrimers: Synthesis, Structure, and Electronic Properties. *Organometallics* **2008**, *27*, 346−352.

(35) Höfler, F.; Jannach, R. Zur Kenntnis des Neopentasilans. *Inorg. Nucl. Chem. Lett.* **1973**, *9*, 723−725.

(36) O'Toole, T. R.; Younathan, J. N.; Sullivan, B. P.; Meyer, T. J. 1,2-Difluorobenzene: a relatively inert and noncoordinating solvent for electrochemical studies on transition-metal complexes. *Inorg. Chem.* **1989**, *28*, 3923−3926.

(37) Baldwin, S. M.; Bercaw, J. E.; Henling, L. M.; Day, M. W.; Brintzinger, H. H. Cationic Alkylaluminum-Complexed Zirconocene Hydrides: NMR-Spectroscopic Identification, Crystallographic Structure Determination, and Interconversion with Other Zirconocene Cations. *J. Am. Chem. Soc.* **2011**, *133*, 1805−1813.

(38) $\mathrm{Me_2AlH}$ synthesized according to ref 37 necessarily requires a distillation step in the present case. Thereafter, $^1$H NMR spectroscopy indicated a $\mathrm{Me_{2.2}AlH_{0.8}}$ composition. For simplicity, we nevertheless refer to this compound as "$\mathrm{Me_2AlH}$" throughout this paper but use the precise stoichiometry in the Supporting Information. In principle, $\mathrm{Me_2AlH}$ can be replaced with commercial *i*$\mathrm{Bu_2AlH}$, but this leads to a slightly increased amount of side products.

(39) The corresponding reaction of [*n*$\mathrm{Bu_4N}$][**2**] with the mixture $\mathrm{Me_2AlH/Me_3Al}$ again leads to core methylation; in the case of [*n*$\mathrm{Bu_4N}$][**3**], the cluster is degraded.

(40) The isolable silane-alane complex $[\mathrm{Et_3Si{-}H{\cdots}Al(C_6F_5)_3}]$ has been characterized by X-ray crystallography: Chen, J.; Chen, E. Y. X. Elusive Silane-Alane Complex [Si-H···Al]: Isolation, Characterization, and Multifaceted Frustrated Lewis Pair Type Catalysis. *Angew. Chem., Int. Ed.* **2015**, *54*, 6842−6846. Furthermore, analogous Si−H···B interactions are essential for $\mathrm{B(C_6F_5)_3}$-catalyzed hydrosilylation reactions and for the Piers-Rubinsztajn reaction (a) Parks, D. J.; Piers, W. E. Tris(pentafluorophenyl)boron-Catalyzed Hydrosilation of Aromatic Aldehydes, Ketones, and Esters. *J. Am. Chem. Soc.* **1996**, *118*, 9440−9441. (b) Brook, M. A. New Control Over Silicone Synthesis using SiH Chemistry: The Piers−Rubinsztajn Reaction. *Chem.—Eur. J.* **2018**, *24*, 8458−8469.

(41) Bursch, M.; Neugebauer, H.; Ehlert, S.; Grimme, S. Dispersion corrected r$^2$SCAN based global hybrid functionals: r$^2$SCANh, r$^2$SCAN0, and r$^2$SCAN50. *J. Chem. Phys.* **2022**, *156*, 134105.

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

(43) 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.

(44) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.

(45) Zheng, J.; Xu, X.; Truhlar, D. G. Minimally augmented Karlsruhe basis sets. *Theor. Chem. Acc.* **2011**, *128*, 295−305.

(46) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99*, 2224−2235.

(47) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. Refinement and Parametrization of COSMO-RS. *J. Phys. Chem. A* **1998**, *102*, 5074−5085.

(48) Grimme, S. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chem.—Eur. J.* **2012**, *18*, 9955−9964.

(49) Pracht, P.; Grimme, S. Calculation of absolute molecular entropies and heat capacities made simple. *Chem. Sci.* **2021**, *12*, 6551−6568.

(50) 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.

(51) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378−6396.

(52) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding quantum chemistry methods. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2021**, *11*, e1493.

(53) 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.

(54) Neese, F. The ORCA program system. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 73−78.

(55) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA quantum chemistry program package. *J. Chem. Phys.* **2020**, *152*, 224108.

(56) Neese, F. Software update: The ORCA program system—Version 5.0. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2022**, *12*, e1606.

(57) TURBOMOLE V7.6 2021, a Development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH. 2007, http://www.turbomole.com (accessed April 21, 2023).

(58) Sommer, L. H.; McLick, J.; Golino, C. M. The $\mathrm{S_N}$i-Si Mechanism. Reductive Displacement of Good Leaving Groups with Retention of Configuration by Diisobutylaluminum Hydride. Stereochemical and Mechanistic Crossover with the Etherate Complex of Diisobutylaluminum Hydride. *J. Am. Chem. Soc.* **1972**, *94*, 669−670.

(59) Sommer, L. H.; Golino, C. M.; Roark, D. N.; Bush, R. D. Reduction of silicon halides and alkoxides with diisobutylaluminum hydride. Stereochemistry-rate law correlations for the $\mathrm{S_N}$*i-Si* and $\mathrm{S_N}$*2-Si* mechanisms. *J. Organomet. Chem.* **1973**, *49*, C3−C5.

(60) Corriu, R. J. P.; Guerin, C. Nucleophilic Displacement at Silicon: Recent Developments and Mechanistic Implications. *Adv. Organomet. Chem.* **1982**, *20*, 265−312.

(61) Lainer, T.; Fischer, R.; Leypold, M.; Holthausen, M.; Wunnicke, O.; Haas, M.; Stueger, H. Unusually selective synthesis of chlorohydrooligosilanes. *Chem. Commun.* **2020**, *56*, 13812−13815.

(62) Particularly in the present case, an $\mathrm{S_N}$1-like mechanism for substituent exchange at the SiCl vertices would also be plausible, because the formal $\mathrm{Si}^+$ center in the resulting neutral intermediate could be stabilized through interaction with the endohedral $\mathrm{Cl}^-$ ion.

Our quantum-chemical calculations indicate that such a scenario is indeed reasonable from a thermodynamic point of view. Despite considerable efforts, however, we have not yet succeeded in modeling the entire reaction pathway, so that we are currently unable to draw definite conclusions regarding reaction kinetics.

(63) These values compare nicely to Si−Si = 2.339−2.393 Å within the fused $Si_5$ cycles of crystallographically characterized bicyclo[3.3.0] octasilanes: Kobayashi, H.; Iwamoto, T.; Kira, M. A Stable Fused Bicyclic Disilene as a Model for Silicon Surface. *J. Am. Chem. Soc.* **2005**, *127*, 15376−15377.

(64) Only the 6 SiCl positions that are not affected by H/Cl disorder were considered.

(65) According to Bent's rule, the $H_3Si$ group will direct a hybrid orbital with the highest *p* character toward the cluster with the highest group electronegativity. Since these group electronegativities follow the order [**4**]$^-$ < [**2**]$^-$ < [**3**]$^-$, the *p* character of the $H_3Si$−Si(cluster) bonds and, correspondingly, the *s* character of the $H_2Si$−H bonds should obey the same trend: Bent, H. A. An Appraisal of Valence-bond Structures and Hybridization in Compounds of the First-row Elements. *Chem. Rev.* **1961**, *61*, 275−311.

(66) Glendening, E. D.; Landis, C. R.; Weinhold, F. NBO 6.0: Natural bond orbital analysis program. *J. Comput. Chem.* **2013**, *34*, 1429−1437.

(67) Glendening, E. D.; Landis, C. R.; Weinhold, F. Erratum: NBO 6.0: Natural bond orbital analysis program. *J. Comput. Chem.* **2013**, *34*, 2134.

(68) Akitt, J. W. The Quadrupolar Halides. In *Multinuclear NMR*; Mason, J., Ed.; Springer: Boston, 1987; pp 447−461.

(69) Although *J*(H,Cl) coupling is only resolved for protons directly bonded to the $Si_{20}$ core, we see a weak cross peak between the Si*H*$_3$ and $^{35}$Cl resonances in the $^1$H$^{35}$Cl-HMBC spectrum; a more intense cross peak appears between the Si*H* and $^{35}$Cl resonances.

(70) As a characteristic NMR feature of our silafulleranes, the $^{29}$Si nuclei of the cage vertices ($Si^0$/$Si^I$) are strongly deshielded relative to $Si^0$/$Si^I$ centers of comparable open-chain oligosilanes; signals from $Si^{III}$ in peripheral $SiR_3$ substituents of the same compound classes appear in the same region. Examples: (a) [*n*Bu$_4$N][**2**]: $\delta(^{29}Si)$ = −58.5 ($Si^0$), −14.9 ($Si^I$), −93.7 ($Si^{III}$); $HSi(SiH_3)_3$: $\delta(^{29}Si)$ = −137.4 ($Si^I$), −95.7 ($Si^{III}$); $Si(SiH_3)_4$: $\delta(^{29}Si)$ = −165.7 ($Si^0$), −90.4 ($Si^{III}$).[33] (b) [*n*Bu$_4$N][**1**]: $\delta(^{29}Si)$ = 31.1 ($Si^I$); [*n*Bu$_4$N][**3**]: $\delta(^{29}Si)$ = 51.5 ($Si^I$); $(Me_3Si)_3SiCl$: $\delta(^{29}Si)$ = −13.3 ($Si^I$); [*n*Bu$_4$N][**2**]: $\delta(^{29}Si)$ = −14.9 ($Si^I$); $(Me_3Si)_3SiH$: $\delta(^{29}Si)$ = −115.4 ($Si^I$); [*n*Bu$_4$N][**4**]: $\delta(^{29}Si)$ = 19.1 ($Si^I$); $(Me_3Si)_3SiMe$: $\delta(^{29}Si)$ = −87.9 ($Si^I$). For $\delta(^{29}Si)$ of $(Me_3Si)_3SiX$ (X = Cl, H, Me), see: Marsmann, H. C.; Raml, W.; Hengge, E. $^{29}$Si-Kernresonanzmessungen an Polysilanen 2. Isotetrasilane. *Z. Naturforsch., B: Anorg. Chem., Org. Chem.* **1980**, *35*, 1541−1547.

(71) A previous theoretical study already suggested that the encapsulation of a $Cl^-$ ion has a significant shielding effect on $^{29}$Si nuclei of $Si^I$ centers of $Si_{20}(SiCl_3)_{12}Cl_8$, but a relatively weak deshielding effect on $^{29}$Si nuclei of $Si^0$ and $Si^{III}$ centers: $Si_{20}(SiCl_3)_{12}Cl_8$: $\delta(^{29}Si$, calcd) = −63.1 ($Si^0$), 53.9 ($Si^I$), 7.9 ($Si^{III}$); $[Cl@Si_{20}(SiCl_3)_{12}Cl_8]^-$ ([**1**]$^-$): $\delta(^{29}Si$, calcd) = −59.9 ($Si^0$), 31.6 ($Si^I$), 13.3 ($Si^{III}$): Ponce-Vargas, M.; Muñoz-Castro, A. Stabilizing Role of Halide Ions in Endohedral [20]Silafulleranes: Insights from DFT Calculations toward Silicon Nanocages. *J. Phys. Chem. C* **2018**, *122*, 12551−12558. For the results of our calculations on the derivatives of [**1**]$^-$−[**4**]$^-$ without endohedral $Cl^-$ ions, see Table S5

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

(73) Klamt, A.; Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, *5*, 799−805.

(74) van Lenthe, E.; Baerends, E. J.; Snijders, J. G. Relativistic total energy using regular approximations. *J. Chem. Phys.* **1994**, *101*, 9783−9792.

(75) Mo, Y.; Zhang, Y.; Gao, J. A Simple Electrostatic Model for Trisilylamine: Theoretical Examinations of the n→$\sigma$* Negative Hyperconjugation, $p_\pi$→$d_\pi$ Bonding, and Stereoelectronic Interaction. *J. Am. Chem. Soc.* **1999**, *121*, 5737−5742.

(76) Rüger, R.; Franchini, M.; Trnka, T.; Yakovlev, A.; van Lente, E.; Philipsen, P.; van Vuren, T.; Klumpers, B.; Soini, T. ADF 2022.1, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands. 2022, http://www.scm.com (accessed April 21, 2023).

(77) Schreckenbach, G.; Ziegler, T. Calculation of NMR Shielding Tensors Using Gauge-Including Atomic Orbitals and Modern Density Functional Theory. *J. Phys. Chem.* **1995**, *99*, 606−611.

(78) Pye, C. C.; Ziegler, T.; van Lenthe, E.; Louwen, J. N. An implementation of the conductor-like screening model of solvation within the Amsterdam density functional package—Part II. COSMO for real solvents. *Can. J. Chem.* **2009**, *87*, 790−797.

(79) van Lenthe, E.; Baerends, E. J. Optimized Slater-type basis sets for the elements 1−118. *J. Comput. Chem.* **2003**, *24*, 1142−1156.

(80) Lehmann, J. F.; Schrobilgen, G. J.; Christe, K. O.; Kornath, A.; Suontamo, R. J. X-ray Crystal Structures of $[XF_6][Sb_2F_{11}]$ (X = Cl, Br, I); $^{35,37}$Cl, $^{79,81}$Br, and $^{127}$I NMR Studies and Electronic Structure Calculations of the $XF_6^+$ Cations. *Inorg. Chem.* **2004**, *43*, 6905−6921.

(81) Jaszuński, M.; Mikkelsen, K. V.; Rizzo, A.; Witanowski, M. A Study of the Nitrogen NMR Spectra of Azoles and Their Solvent Dependence. *J. Phys. Chem. A* **2000**, *104*, 1466−1473.

(82) Philips, A.; Marchenko, A.; Ducati, L. C.; Autschbach, J. Quadrupolar $^{14}$N NMR Relaxation from Force-Field and Ab Initio Molecular Dynamics in Different Solvents. *J. Chem. Theory Comput.* **2019**, *15*, 509−519.

(83) Deviations from 0 are due to artificial symmetry breaking inherent to the method applied.

(84) Marciniec, B. Catalysis by transition metal complexes of alkene silylation—recent progress and mechanistic implications. *Coord. Chem. Rev.* **2005**, *249*, 2374−2390.

(85) Troegel, D.; Stohrer, J. Recent advances and actual challenges in late transition metal catalyzed hydrosilylation of olefins from an industrial point of view. *Coord. Chem. Rev.* **2011**, *255*, 1440−1459.

(86) Tilley, T. D. The Coordination Polymerization of Silanes to Polysilanes by a "$\sigma$-Bond Metathesis" Mechanism. Implications for Linear Chain Growth. *Acc. Chem. Res.* **1993**, *26*, 22−29.

(87) Corey, J. Y. Dehydrocoupling of Hydrosilanes to Polysilanes and Silicon Oligomers: A 30 Year Overview. *Adv. Organomet. Chem.* **2004**, *51*, 1−52.

(88) Bock, H.; Ensslin, W. Bond−Bond Interaction in Polysilanes. *Angew. Chem., Int. Ed. Engl.* **1971**, *10*, 404−405.

(89) Miller, R. D.; Michl, J. Polysilane High Polymers. *Chem. Rev.* **1989**, *89*, 1359−1410.

(90) Ishida, S.; Otsuka, K.; Toma, Y.; Kyushin, S. An Organosilicon Cluster with an Octasilacuneane Core: A Missing Silicon Cage Motif. *Angew. Chem., Int. Ed.* **2013**, *52*, 2507−2510.

(91) Jovanovic, M.; Michl, J. Alkanes versus Oligosilanes: Conformational Effects on $\sigma$-Electron Delocalization. *J. Am. Chem. Soc.* **2022**, *144*, 463−477.

# Bibliography

[1] T. Gasevic, J. B. Stückrath, S. Grimme, and M. Bursch, *Optimization of the $r^2$SCAN-3c Composite Electronic-Structure Method for Use with Slater-Type Orbital Basis Sets*, J. Phys. Chem. A **126** (2022) 3826, doi: `10.1021/acs.jpca.2c02951`.

[2] T. Gasevic, J. B. Kleine Büning, S. Grimme, and M. Bursch, *Benchmark Study on the Calculation of $^{207}$Pb NMR Chemical Shifts*, Inorg. Chem. **63** (2024) 5052, doi: `10.1021/acs.inorgchem.3c04539`.

[3] T. Gasevic, M. Bursch, Q. Ma, S. Grimme, H.-J. Werner, and A. Hansen, *The p-block challenge: assessing quantum chemistry methods for inorganic heterocycle dimerizations*, Phys. Chem. Chem. Phys. **26** (2024) 13884, doi: `10.1039/d3cp06217a`.

[4] T. Gasevic, M. Müller, J. Schöps, S. Lanius, J. Hermann, S. Grimme, and A. Hansen, *Chemical Space Exploration with Artificial "Mindless" Molecules*, J. Chem. Inf. Model. **65** (2025) 9576, doi: `10.1021/acs.jcim.5c01364`.

[5] T. Gasevic, M. Bamberg, J. Wicke, M. Bolte, A. Virovets, H.-W. Lerner, S. Grimme, A. Hansen, M. Wagner, and M. Bursch, *Confined Lewis Pairs: Investigation of the $X^{-} \rightarrow Si_{20}$ Interaction in Halogen-Encapsulating Silafulleranes*, Angew. Chem., Int. Ed. **63** (2024) e202314238, doi: `10.1002/anie.202314238`.

[6] T. Gasevic, M. Bamberg, J. Wicke, M. Bolte, A. Virovets, H.-W. Lerner, S. Grimme, A. Hansen, M. Wagner, and M. Bursch, *Confined Lewis Pairs: Investigation of the $X^{-} \rightarrow Si_{20}$ Interaction in Halogen-Encapsulating Silafulleranes*, Angew. Chem. **136** (2024) e202314238, doi: `10.1002/ange.202314238`.

[7] M. Bamberg, T. Gasevic, M. Bolte, A. Virovets, H.-W. Lerner, S. Grimme, M. Bursch, and M. Wagner, *Regioselective Derivatization of Silylated [20]Silafulleranes*, J. Am. Chem. Soc. **145** (2023) 11440, doi: `10.1021/jacs.3c03270`.

[8] M. Bamberg, T. Gasevic, M. Bolte, A. Virovets, H.-W. Lerner, S. Grimme, M. Bursch, and M. Wagner, *Brominated [20]silafulleranes: pushing the limits of steric loading*, Chem. Commun. **59** (2023) 7459, doi: `10.1039/d3cc02142d`.

[9] M. Bamberg, T. Gasevic, T. G. Saint-Denis, J. Martinez Fernandez, S. Grimme, and T. D. Tilley, *Synthesis of Metallostannylenes from Transition Metal Polyhydride Complexes*, J. Am. Chem. Soc. (2025), doi: `10.1021/jacs.5c14064`.

[10] T. Gasevic, C. Plett, L. Wittmann, I. Neira, C. Peinador, M. D. García, and A. Hansen, *Supramolecular Host-Guest Complexation Dynamics by Cost-Efficient Electronic Structure Methods*, ChemRxiv preprint (2025), doi: `10.26434/chemrxiv-2025-3z07r`.

[11] M. Bursch, T. Gasevic, J. B. Stückrath, and S. Grimme, *Comprehensive Benchmark Study on the Calculation of $^{29}$Si NMR Chemical Shifts*, Inorg. Chem. **60** (2021) 272, DOI: `10.1021/acs.inorgchem.0c02907`.

[12] J. B. Stückrath, T. Gasevic, M. Bursch, and S. Grimme, *Benchmark Study on the Calculation of $^{119}$Sn NMR Chemical Shifts*, Inorg. Chem. **61** (2022) 3903, DOI: `10.1021/acs.inorgchem.1c03453`.

[13] L. Fabbrizzi, *Communicating about matter with symbols: Evolving from alchemy to chemistry*, J. Chem. Educ. **85** (2008) 1501, DOI: `10.1021/ed085p1501`.

[14] T. C. Ezike et al., *Advances in drug delivery systems, challenges and future directions*, Heliyon **9** (2023) e17488, DOI: `10.1016/j.heliyon.2023.e17488`.

[15] A. A. Feidenhans'l, Y. N. Regmi, C. Wei, D. Xia, J. Kibsgaard, and L. A. King, *Precious Metal Free Hydrogen Evolution Catalyst Design and Application*, Chem. Rev. **124** (2024) 5617, DOI: `10.1021/acs.chemrev.3c00712`.

[16] T. F. Jaramillo, K. P. Jørgensen, J. Bonde, J. H. Nielsen, S. Horch, and I. Chorkendorff, *Identification of Active Edge Sites for Electrochemical $H_2$ Evolution from $MoS_2$ Nanocatalysts*, Science **317** (2007) 100, DOI: `10.1126/science.1141483`.

[17] B. Qiao, A. Wang, X. Yang, L. F. Allard, Z. Jiang, Y. Cui, J. Liu, J. Li, and T. Zhang, *Single-atom catalysis of CO oxidation using $Pt_1$/FeOx*, Nat. Chem. **3** (2011) 634, DOI: `10.1038/nchem.1095`.

[18] A. C. Anderson, *The Process of Structure-Based Drug Design*, Chem. Biol. **10** (2003) 787, DOI: `10.1016/j.chembiol.2003.09.002`.

[19] C. W. Murray and D. C. Rees, *The rise of fragment-based drug discovery*, Nat. Chem. **1** (2009) 187, DOI: `10.1038/nchem.217`.

[20] D. C. Rees, M. Congreve, C. W. Murray, and R. Carr, *Fragment-based lead discovery*, Nat. Rev. Drug Discov. **3** (2004) 660, DOI: `10.1038/nrd1467`.

[21] H. Furukawa, K. E. Cordova, M. O'Keeffe, and O. M. Yaghi, *The Chemistry and Applications of Metal-Organic Frameworks*, Science **341** (2013), DOI: `10.1126/science.1230444`.

[22] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, *Electric Field Effect in Atomically Thin Carbon Films*, Science **306** (2004) 666, DOI: `10.1126/science.1102896`.

[23] R. A. Mata and M. A. Suhm, *Benchmarking Quantum Chemical Methods: Are We Heading in the Right Direction?* Angew. Chem., Int. Ed. **56** (2017) 11011, DOI: `10.1002/anie.201611308`.

[24] N. Mardirossian and M. Head-Gordon, *Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals*, Mol. Phys. **115** (2017) 2315, DOI: `10.1080/00268976.2017.1333644`.

[25] P. O. Dral, *AI in computational chemistry through the lens of a decade-long journey*, Chem. Commun. **60** (2024) 3240, DOI: `10.1039/D4CC00010B`.

[26] A. D. Becke, *Perspective: Fifty years of density-functional theory in chemical physics*, J. Chem. Phys. **140** (2014), DOI: `10.1063/1.4869598`.

[27] M. Karplus and J. A. McCammon, *Molecular dynamics simulations of biomolecules*, Nat. Struct. Biol. **9** (2002) 646, DOI: `10.1038/nsb0902-646`.

[28] S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, *The high-throughput highway to computational materials design*, Nat. Mater. **12** (2013) 191, DOI: `10.1038/nmat3568`.

[29] E. Schrödinger, *Quantisierung als Eigenwertproblem*, Ann. Phys. **384** (1926) 361, DOI: `10.1002/andp.19263840404`.

[30] C. D. Sherrill, *Frontiers in electronic structure theory*, J. Chem. Phys. **132** (2010), DOI: `10.1063/1.3369628`.

[31] J. P. Perdew and K. Schmidt, "Jacob's ladder of density functional approximations for the exchange-correlation energy," *AIP Conference Proceedings*, vol. 577, 1, American Institute of Physics, AIP, 2001 1, DOI: `10.1063/1.1390175`.

[32] M. Bursch, J.-M. Mewes, A. Hansen, and S. Grimme, *Best-Practice DFT Protocols for Basic Molecular Computational Chemistry*, Angew. Chem., Int. Ed. **61** (2022), DOI: `10.1002/anie.202205735`.

[33] J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller, and A. Tkatchenko, *Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems*, Chem. Rev. **121** (2021) 9816, DOI: `10.1021/acs.chemrev.1c00107`.

[34] A. Katbashev, M. Stahn, T. Rose, V. Alizadeh, M. Friede, C. Plett, P. Steinbach, and S. Ehlert, *Overview on Building Blocks and Applications of Efficient and Robust Extended Tight Binding*, J. Phys. Chem. A **129** (2025) 2667, DOI: `10.1021/acs.jpca.4c08263`.

[35] A. Karton and M. T. de Oliveira, *Good Practices in Database Generation for Benchmarking Density Functional Theory*, WIREs Comput. Mol. Sci. **15** (2025) e1737, DOI: `10.1002/wcms.1737`.

[36] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme, *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. **19** (2017) 32184, DOI: `10.1039/c7cp04913g`.

[37] J. Liang and M. Head-Gordon, *Gold-Standard Chemical Database 138 (GSCDB138): A diverse set of accurate energy differences for assessing and developing density functionals*, arXiv preprint (2025), DOI: `10.48550/arXiv.2508.13468`.

[38] V. K. Prasad, Z. Pei, S. Edelmann, A. Otero-de-la-Roza, and G. A. DiLabio, *BH9, a New Comprehensive Benchmark Data Set for Barrier Heights and Reaction Energies: Assessment of Density Functional Approximations and Basis Set Incompleteness Potentials*, J. Chem. Theory Comput. **18** (2022) 151, DOI: `10.1021/acs.jctc.1c00694`.

[39] C. J. Schattenberg and M. Kaupp, *Extended Benchmark Set of Main-Group Nuclear Shielding Constants and NMR Chemical Shifts and Its Use to Evaluate Modern DFT Methods*, J. Chem. Theory Comput. **17** (2021) 7602, DOI: `10.1021/acs.jctc.1c00919`.

[40] J. Hoja and A. D. Boese, *The V30 benchmark set for anharmonic vibrational frequencies of molecular dimers*, J. Chem. Phys. **161** (2024), DOI: `10.1063/5.0238491`.

[41] I. M. Alecu, J. Zheng, Y. Zhao, and D. G. Truhlar, *Computational Thermochemistry: Scale Factor Databases and Scale Factors for Vibrational Frequencies Obtained from Electronic Model Chemistries*, J. Chem. Theory Comput. **6** (2010) 2872, DOI: `10.1021/ct100326h`.

[42] J. C. Zapata Trujillo and L. K. McKemmish, *VIBFREQ1295: A New Database for Vibrational Frequency Calculations*, J. Phys. Chem. A **126** (2022) 4100, DOI: `10.1021/acs.jpca.2c01438`.

[43] S. Grimme, A. Hansen, S. Ehlert, and J. M. Mewes, *$r^2$SCAN-3c: A "Swiss army knife" composite electronic-structure method*, J. Chem. Phys. **154** (2021) 064103, DOI: `10.1063/5.0040021`.

[44] F. Jensen, *Introduction to Computational Chemistry*, 2nd ed., Chichester, England: John Wiley & Sons, 2006.

[45] D. R. Hartree, *The Wave Mechanics of an Atom with a Non-Coulomb Central Field Part I Theory and Methods*, Math. Proc. Cambridge Philos. Soc. **24** (1928) 89, DOI: `10.1017/S0305004100011919`.

[46] V. Fock, *Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems*, Z. Phys. **61** (1930) 126, DOI: `10.1007/BF01340294`.

[47] C. C. Roothaan, *New developments in molecular orbital theory*, Rev. Mod. Phys. **23** (1951) 69, DOI: `10.1103/RevModPhys.23.69`.

[48] H. Kümmel, *Origins of the Coupled Cluster Method*, Theor. Chim. Acta **80** (1991) 81, DOI: `10.1007/BF01119615`.

[49] P. Hohenberg and W. Kohn, *Inhomogeneous electron gas*, Phys. Rev. **136** (1964) B864, DOI: `10.1103/PhysRev.136.B864`.

[50] W. Kohn and L. J. Sham, *Self-consistent equations including exchange and correlation effects*, Phys. Rev. **140** (1965) A1133, DOI: `10.1103/PhysRev.140.A1133`.

[51] M. J. Dewar and W. Thiel, *Ground States of Molecules. 38. The MNDO Method. Approximations and Parameters*, J. Am. Chem. Soc. **99** (1977) 4899, DOI: `10.1021/ja00457a004`.

[52] P. O. Dral, B. Hourahine, and S. Grimme, *Modern semiempirical electronic structure methods*, J. Chem. Phys. **160** (2024), DOI: `10.1063/5.0196138`.

[53] L. Monticelli and D. P. Tieleman, “Force Fields for Classical Molecular Dynamics,” *Biomolecular Simulations: Methods and Protocols*, Totowa, NJ: Humana Press, 2013 197, DOI: `10.1007/978-1-62703-017-5_8`.

[54] S. Spicher and S. Grimme, *Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems*, Angew. Chem., Int. Ed. **59** (2020) 15665, DOI: `10.1002/anie.202004239`.

[55] P. A. M. Dirac, *Quantum mechanics of many-electron systems*, Proc. R. Soc. Lond. A **123** (1929) 714, DOI: `10.1098/rspa.1929.0094`.

[56] W. Pauli, *Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren*, Z. Phys. **31** (1925) 765, DOI: `10.1007/BF02980631`.

[57] D. P. Tew, W. Klopper, and T. Helgaker, *Electron correlation: The many-body problem at the heart of chemistry*, J. Comput. Chem. **28** (2007) 1307, DOI: `10.1002/jcc.20581`.

[58] C. Møller and M. S. Plesset, *Note on an Approximation Treatment for Many-Electron Systems*, Phys. Rev. **46** (1934) 618, DOI: `10.1103/PhysRev.46.618`.

[59] C. Riplinger and F. Neese, *An efficient and near linear scaling pair natural orbital based local coupled cluster method*, J. Chem. Phys. **138** (2013), DOI: `10.1063/1.4773581`.

[60] C. Riplinger, B. Sandhoefer, A. Hansen, and F. Neese, *Natural triple excitations in local coupled cluster calculations with pair natural orbitals*, J. Chem. Phys. **139** (2013), DOI: `10.1063/1.4821834`.

[61] C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, and F. Neese, *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. **144** (2016), DOI: `10.1063/1.4939030`.

[62] M. Saitow, U. Becker, C. Riplinger, E. F. Valeev, and F. Neese, *A new near-linear scaling, efficient and accurate, open-shell domain-based local pair natural orbital coupled cluster singles and doubles theory*, J. Chem. Phys. **146** (2017), DOI: `10.1063/1.4981521`.

[63] H.-J. Werner, G. Knizia, C. Krause, M. Schwilk, and M. Dornbach, *Scalable electron correlation methods I.: PNO-LMP2 with linear scaling in the molecular size and near-inverse-linear scaling in the number of processors*, J. Chem. Theory Comput. **11** (2015) 484, DOI: `10.1021/ct500725e`.

[64] Q. Ma and H.-J. Werner, *Scalable Electron Correlation Methods. 2. Parallel PNO-LMP2-F12 with Near Linear Scaling in the Molecular Size*, J. Chem. Theory Comput. **11** (2015) 5291, DOI: `10.1021/acs.jctc.5b00843`.

[65] M. Schwilk, Q. Ma, C. Köppl, and H.-J. Werner, *Scalable electron correlation methods. 3. Efficient and accurate parallel local coupled cluster with pair natural orbitals (PNO-LCCSD)*, J. Chem. Theory Comput. **13** (2017) 3650, DOI: `10.1021/acs.jctc.7b00554`.

[66] Q. Ma, M. Schwilk, C. Köppl, and H.-J. Werner, *Scalable Electron Correlation Methods. 4. Parallel Explicitly Correlated Local Coupled Cluster with Pair Natural Orbitals (PNO-LCCSD-F12)*, J. Chem. Theory Comput. **13** (2017) 4871, DOI: `10.1021/acs.jctc.7b00799`.

[67] Q. Ma and H.-J. Werner, *Scalable Electron Correlation Methods. 7. Local Open-Shell Coupled-Cluster Methods Using Pair Natural Orbitals: PNO-RCCSD and PNO-UCCSD*, J. Chem. Theory Comput. **16** (2020) 3135, DOI: `10.1021/acs.jctc.0c00192`.

[68] Q. Ma and H. J. Werner, *Scalable Electron Correlation Methods. 8. Explicitly Correlated Open-Shell Coupled-Cluster with Pair Natural Orbitals PNO-RCCSD(T)-F12 and PNO-UCCSD(T)-F12*, J. Chem. Theory Comput. **17** (2021) 902, DOI: `10.1021/acs.jctc.0c01129`.

[69] B. O. Roos, P. R. Taylor, and P. E. Sigbahn, *A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach*, Chem. Phys. **48** (1980) 157, DOI: `10.1016/0301-0104(80)80045-0`.

[70] R. J. Buenker and S. D. Peyerimhoff, *Individualized configuration selection in CI calculations with subsequent energy extrapolation*, Theor. Chim. Acta **35** (1974) 33, DOI: `10.1007/BF02394557`.

[71] R. J. Buenker, S. D. Peyerimhoff, and W. Butscher, *Applicability of the multi-reference double-excitation CI (MRD-CI) method to the calculation of electronic wavefunctions and comparison with related techniques*, Mol. Phys. **35** (1978) 771, DOI: `10.1080/00268977800100581`.

[72] H. J. Werner and P. J. Knowles, *An efficient internally contracted multiconfiguration-reference configuration interaction method*, J. Chem. Phys. **89** (1988) 5803, DOI: `10.1063/1.455556`.

[73] J. W. Park, R. Al-Saadon, M. K. MacLeod, T. Shiozaki, and B. Vlaisavljevich, *Multireference Electron Correlation Methods: Journeys along Potential Energy Surfaces*, Chem. Rev. **120** (2020) 5878, DOI: `10.1021/acs.chemrev.9b00496`.

[74] A. Khedkar and M. Roemelt, *Modern multireference methods and their application in transition metal chemistry*, Phys. Chem. Chem. Phys. **23** (2021) 17097, DOI: `10.1039/d1cp02640b`.

[75] S. H. Vosko, L. Wilk, and M. Nusair, *Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis*, Can. J. Phys. **58** (1980) 1200, DOI: `10.1139/p80-159`.

[76] J. P. Perdew and Y. Wang, *Accurate and simple analytic representation of the electron-gas correlation energy*, Phys. Rev. B **45** (1992) 13244, DOI: `10.1103/PhysRevB.45.13244`.

[77] A. D. Becke, *Density-functional exchange-energy approximation with correct asymptotic behavior*, Phys. Rev. A **38** (1988) 3098, DOI: `10.1103/PhysRevA.38.3098`.

[78] C. Lee, W. Yang, and R. G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density*, Phys. Rev. B **37** (1988) 785, DOI: `10.1103/PhysRevB.37.785`.

[79] J. P. Perdew, K. Burke, and M. Ernzerhof, *Generalized gradient approximation made simple*, Phys. Rev. Lett. **77** (1996) 3865, DOI: `10.1103/PhysRevLett.77.3865`.

[80] J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, and J. Sun, *Accurate and Numerically Efficient $r^2$SCAN Meta-Generalized Gradient Approximation*, J. Phys. Chem. Lett. **11** (2020) 8208, DOI: `10.1021/acs.jpclett.0c02405`.

[81] A. D. Becke, *Density-functional thermochemistry. III. The role of exact exchange*, J. Chem. Phys. **98** (1993) 5648, DOI: `10.1063/1.464913`.

[82] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, *Ab Initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields*, J. Phys. Chem. **98** (1994) 11623, DOI: `10.1021/j100096a001`.

[83] C. Adamo and V. Barone, *Toward reliable density functional methods without adjustable parameters: The PBE0 model*, J. Chem. Phys. **110** (1999) 6158, DOI: `10.1063/1.478522`.

[84] N. Mardirossian and M. Head-Gordon, *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. **16** (2014) 9904, DOI: `10.1039/c3cp54374a`.

[85] N. Mardirossian and M. Head-Gordon, *ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation*, J. Chem. Phys. **144** (2016), DOI: `10.1063/1.4952647`.

[86] T. M. Maier, A. V. Arbuznikov, and M. Kaupp, *Local hybrid functionals: Theory, implementation, and performance of an emerging new tool in quantum chemistry and beyond*, WIREs Comput. Mol. Sci. **9** (2019), doi: `10.1002/wcms.1378`.

[87] S. Grimme, *Semiempirical hybrid density functional with perturbative second-order correlation*, J. Chem. Phys. **124** (2006), doi: `10.1063/1.2148954`.

[88] G. Santra, M. Cho, and J. M. Martin, *Exploring Avenues beyond Revised DSD Functionals: I. Range Separation, with xDSD as a Special Case*, J. Phys. Chem. A **125** (2021) 4614, doi: `10.1021/acs.jpca.1c01294`.

[89] L. Wittmann, H. Neugebauer, S. Grimme, and M. Bursch, *Dispersion-corrected r2SCAN based double-hybrid functionals*, J. Chem. Phys. **159** (2023) 224103, doi: `10.1063/5.0174988`.

[90] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, *A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu*, J. Chem. Phys. **132** (2010) 154104, doi: `10.1063/1.3382344`.

[91] S. Grimme, S. Ehrlich, and L. Goerigk, *Effect of the damping function in dispersion corrected density functional theory*, J. Comput. Chem. **32** (2011) 1456, doi: `10.1002/jcc.21759`.

[92] S. Grimme, A. Hansen, J. G. Brandenburg, and C. Bannwarth, *Dispersion-Corrected Mean-Field Electronic Structure Methods*, Chem. Rev. **116** (2016) 5105, doi: `10.1021/acs.chemrev.5b00533`.

[93] E. Caldeweyher, C. Bannwarth, and S. Grimme, *Extension of the D3 dispersion coefficient model*, J. Chem. Phys. **147** (2017) 034112, doi: `10.1063/1.4993215`.

[94] L. Wittmann, I. Gordiy, M. Friede, B. Helmich-Paris, S. Grimme, A. Hansen, and M. Bursch, *Extension of the D3 and D4 London dispersion corrections to the full actinides series*, Phys. Chem. Chem. Phys. **26** (2024) 21379, doi: `10.1039/D4CP01514B`.

[95] D. E. Taylor et al., *Blind test of density-functional-based methods on intermolecular interaction energies*, J. Chem. Phys. **145** (2016) 124105, doi: `10.1063/1.4961095`.

[96] L. Gráfová, M. Pitoňák, J. Řezáč, and P. Hobza, *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. **6** (2010) 2365, doi: `10.1021/ct1002253`.

[97] J. Řezáč, K. E. Riley, and P. Hobza, *S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures*, J. Chem. Theory Comput. **7** (2011) 2427, doi: `10.1021/ct2002946`.

[98] O. A. Vydrov and T. Van Voorhis, *Nonlocal van der Waals density functional: The simpler the better*, J. Chem. Phys. **133** (2010) 244103, doi: `10.1063/1.3521275`.

[99] T. Kato, *On the eigenfunctions of many-particle systems in quantum mechanics*, Commun. Pure Appl. Math. **10** (1957) 151, doi: `10.1002/cpa.3160100201`.

[100] F. Weigend and R. Ahlrichs, *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. **7** (2005) 3297, doi: `10.1039/b508541a`.

[101] A. Hellweg and D. Rappoport, *Development of new auxiliary basis functions of the Karlsruhe segmented contracted basis sets including diffuse basis functions (def2-SVPD, def2-TZVPPD, and def2-QVPPD) for RI-MP2 and RI-CC calculations*, Phys. Chem. Chem. Phys. **17** (2015) 1010, DOI: `10.1039/C4CP04286G`.

[102] T. H. Dunning, *Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen*, J. Chem. Phys. **90** (1989) 1007, DOI: `10.1063/1.456153`.

[103] D. E. Woon and T. H. Dunning, *Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon*, J. Chem. Phys. **98** (1993) 1358, DOI: `10.1063/1.464303`.

[104] R. A. Kendall, T. H. Dunning, and R. J. Harrison, *Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions*, J. Chem. Phys. **96** (1992) 6796, DOI: `10.1063/1.462569`.

[105] D. P. Chong, E. Van Lenthe, S. Van Gisbergen, and E. J. Baerends, *Even-tempered slater-type orbitals revisited: From hydrogen to krypton*, J. Comput. Chem. **25** (2004) 1030, DOI: `10.1002/jcc.20030`.

[106] D. P. Chong, *Augmenting basis set for time-dependent density functional theory calculation of excitation energies: Slater-type orbitals for hydrogen to krypton*, Mol. Phys. **103** (2005) 749, DOI: `10.1080/00268970412331333618`.

[107] E. Van Lenthe and E. J. Baerends, *Optimized Slater-type basis sets for the elements 1-118*, J. Comput. Chem. **24** (2003) 1142, DOI: `10.1002/jcc.10255`.

[108] *AMS 2025.1, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands*, `http://www.scm.com`, 2025.

[109] E. J. Baerends et al., *The Amsterdam Modeling Suite*, J. Chem. Phys. **162** (2025) 162501, DOI: `10.1063/5.0228009`.

[110] S. Boys and F. Bernardi, *The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors*, Mol. Phys. **19** (1970) 553, DOI: `10.1080/00268977000101561`.

[111] H. Kruse and S. Grimme, *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. **136** (2012), DOI: `10.1063/1.3700154`.

[112] B. Metz, H. Stoll, and M. Dolg, *Small-core multiconfiguration-Dirac–Hartree–Fock-adjusted pseudopotentials for post-d main group elements: Application to PbH and PbO*, J. Chem. Phys. **113** (2000) 2563, DOI: `10.1063/1.1305880`.

[113] L. J. Norrby, *Why is mercury liquid? Or, why do relativistic effects not get into chemistry textbooks?* J. Chem. Educ. **68** (1991) 110, DOI: `10.1021/ed068p110`.

[114] M. Douglas and N. M. Kroll, *Quantum electrodynamical corrections to the fine structure of helium*, Ann. Phys. (N. Y). **82** (1974) 89, DOI: `10.1016/0003-4916(74)90333-9`.

[115] B. A. Hess, *Relativistic electronic-structure calculations employing a two-component no-pair formalism with external-field projection operators*, Phys. Rev. A **33** (1986) 3742, DOI: `10.1103/PhysRevA.33.3742`.

[116] A. Wolf, M. Reiher, and B. A. Hess, *The generalized Douglas-Kroll transformation*, J. Chem. Phys. **117** (2002) 9215, DOI: `10.1063/1.1515314`.

[117] M. Reiher, *Douglas-Kroll-Hess theory: A relativistic electrons-only theory for chemistry*, Theor. Chem. Acc. **116** (2006) 241, DOI: `10.1007/s00214-005-0003-2`.

[118] C. Chang, M. Pelissier, and P. Durand, *Regular two-component pauli-like effective hamiltonians in dirac theory*, Phys. Scr. **34** (1986) 394, DOI: `10.1088/0031-8949/34/5/007`.

[119] E. Van Lenthe, E. J. Baerends, and J. G. Snijders, *Relativistic regular two-component Hamiltonians*, J. Chem. Phys. **99** (1993) 4597, DOI: `10.1063/1.466059`.

[120] E. Van Lenthe, E. J. Baerends, and J. G. Snijders, *Relativistic total energy using regular approximations*, J. Chem. Phys. **101** (1994) 9783, DOI: `10.1063/1.467943`.

[121] W. Kutzelnigg and W. Liu, *Quasirelativistic theory equivalent to fully relativistic theory*, J. Chem. Phys. **123** (2005), DOI: `10.1063/1.2137315`.

[122] W. Kutzelnigg and W. Liu, *Quasirelativistic theory I. Theory in terms of a quasi-relativistic operator*, Mol. Phys. **104** (2006) 2225, DOI: `10.1080/00268970600662481`.

[123] W. Liu and W. Kutzelnigg, *Quasirelativistic theory. II. Theory at matrix level*, J. Chem. Phys. **126** (2007), DOI: `10.1063/1.2710258`.

[124] W. Liu and D. Peng, *Infinite-order quasirelativistic density functional method based on the exact matrix quasirelativistic theory*, J. Chem. Phys. **125** (2006), DOI: `10.1063/1.2222365`.

[125] D. Peng, W. Liu, Y. Xiao, and L. Cheng, *Making four- and two-component relativistic density functional methods fully equivalent based on the idea of "from atoms to molecule"*, J. Chem. Phys. **127** (2007), DOI: `10.1063/1.2772856`.

[126] M. Iliaš and T. Saue, *An infinite-order two-component relativistic Hamiltonian by a simple one-step transformation*, J. Chem. Phys. **126** (2007), DOI: `10.1063/1.2436882`.

[127] R. Sure and S. Grimme, *Corrected Small Basis Set Hartree-Fock Method for Large Systems*, J. Comput. Chem. **34** (2013) 1672, DOI: `10.1002/jcc.23317`.

[128] S. Grimme, J. G. Brandenburg, C. Bannwarth, and A. Hansen, *Consistent structures and interactions by density functional theory with small atomic orbital basis sets*, J. Chem. Phys. **143** (2015) 054107, DOI: `10.1063/1.4927476`.

[129] J. G. Brandenburg, C. Bannwarth, A. Hansen, and S. Grimme, *B97-3c: A revised low-cost variant of the B97-D density functional method*, J. Chem. Phys. **148** (2018) 064104, DOI: `10.1063/1.5012601`.

[130] J. G. Brandenburg, E. Caldeweyher, and S. Grimme, *Screened exchange hybrid density functional for accurate and efficient structures and interaction energies*, Phys. Chem. Chem. Phys. **18** (2016) 15519, DOI: `10.1039/c6cp01697a`.

[131] P. Pracht, D. F. Grant, and S. Grimme, *Comprehensive Assessment of GFN Tight-Binding and Composite Density Functional Theory Methods for Calculating Gas-Phase Infrared Spectra*, J. Chem. Theory Comput. **16** (2020) 7044, DOI: `10.1021/acs.jctc.0c00877`.

[132] M. Müller, A. Hansen, and S. Grimme, *$\omega$B97X-3c: A composite range-separated hybrid DFT method with a molecule-optimized polarized valence double-$\zeta$ basis set*, J. Chem. Phys. **158** (2023) 14103, DOI: `10.1063/5.0133026/2867476`.

[133] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, and S. Grimme, *Extended tight-binding quantum chemistry methods*, WIREs Comput. Mol. Sci. **11** (2021) e1493, DOI: `10.1002/wcms.1493`.

[134] S. Grimme, C. Bannwarth, and P. Shushkov, *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. **13** (2017) 1989, DOI: `10.1021/acs.jctc.7b00118`.

[135] C. Bannwarth, S. Ehlert, and S. Grimme, *GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions*, J. Chem. Theory Comput. **15** (2019) 1652, DOI: `10.1021/acs.jctc.8b01176`.

[136] D. M. Anstine, R. Zubatyuk, and O. Isayev, *AIMNet2: a neural network potential to meet your neutral, charged, organic, and elemental-organic needs*, Chem. Sci. **16** (2025) 10228, DOI: `10.1039/d4sc08572h`.

[137] B. M. Wood et al., *UMA: A Family of Universal Models for Atoms*, arXiv preprint (2025), DOI: `10.48550/arXiv.2506.23971`.

[138] P. Zimmerman, *Reliable transition state searches integrated with the growing string method*, J. Chem. Theory Comput. **9** (2013) 3043, DOI: `10.1021/ct400319w`.

[139] H. Jónsson, G. Mills, and K. W. Jacobsen, "Nudged elastic band method for finding minimum energy paths of transitions," *Classical and Quantum Dynamics in Condensed Phase Simulations*, World Scientific, 1998 385, DOI: `10.1142/9789812839664_0016`.

[140] V. Ásgeirsson, B. O. Birgisson, R. Bjornsson, U. Becker, F. Neese, C. Riplinger, and H. Jónsson, *Nudged Elastic Band Method for Molecular Reactions Using Energy-Weighted Springs Combined with Eigenvector following*, J. Chem. Theory Comput. **17** (2021) 4929, DOI: `10.1021/acs.jctc.1c00462`.

[141] B. de Souza, *GOAT: A Global Optimization Algorithm for Molecules and Atomic Clusters*, Angew. Chem., Int. Ed. **64** (2025), DOI: `10.1002/anie.202500393`.

[142] P. Pracht, F. Bohle, and S. Grimme, *Automated exploration of the low-energy chemical space with fast quantum chemical methods*, Phys. Chem. Chem. Phys. **22** (2020) 7169, DOI: `10.1039/c9cp06869d`.

[143] P. Pracht et al., *CREST—A program for the exploration of low-energy molecular chemical space*, J. Chem. Phys. **160** (2024), DOI: `10.1063/5.0197592`.

[144] S. Grimme, F. Bohle, A. Hansen, P. Pracht, S. Spicher, and M. Stahn, *Efficient Quantum Chemical Calculation of Structure Ensembles and Free Energies for Nonrigid Molecules*, J. Phys. Chem. A **125** (2021) 4039, DOI: `10.1021/acs.jpca.1c00971`.

[145] S. Grimme, *Supramolecular binding thermodynamics by dispersion-corrected density functional theory*, Chem. Eur. J. **18** (2012) 9955, DOI: `10.1002/chem.201200497`.

[146] P. Pracht and S. Grimme, *Calculation of absolute molecular entropies and heat capacities made simple*, Chem. Sci. **12** (2021) 6551, DOI: `10.1039/d1sc00621e`.

[147] W. Clark Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, *Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics*, J. Am. Chem. Soc. **112** (1990) 6127, DOI: `10.1021/ja00172a038`.

[148] A. W. Lange and J. M. Herbert, *Improving Generalized Born Models by Exploiting Connections to Polarizable Continuum Models. I. An Improved Effective Coulomb Operator*, J. Chem. Theory Comput. **8** (2012) 1999, DOI: `10.1021/ct300111m`.

[149] A. W. Lange and J. M. Herbert, *Improving Generalized Born Models by Exploiting Connections to Polarizable Continuum Models. II. Corrections for Salt Effects*, J. Chem. Theory Comput. **8** (2012) 4381, DOI: `10.1021/ct300493y`.

[150] S. Ehlert, M. Stahn, S. Spicher, and S. Grimme, *Robust and Efficient Implicit Solvation Model for Fast Semiempirical Methods*, J. Chem. Theory Comput. **17** (2021) 4250, DOI: `10.1021/acs.jctc.1c00471`.

[151] V. Barone and M. Cossi, *Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model*, J. Phys. Chem. A **102** (1998) 1995, DOI: `10.1021/jp9716997`.

[152] A. Klamt and G. Schüürmann, *COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient*, J. Chem. Soc., Perkin Trans. 2 (1993) 799, DOI: `10.1039/P29930000799`.

[153] A. Klamt, *The COSMO and COSMO-RS solvation models*, Wiley Interdiscip. Rev. Comput. Mol. Sci. **8** (2018), DOI: `10.1002/wcms.1338`.

[154] A. Klamt, *Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena*, J. Phys. Chem. **99** (1995) 2224, DOI: `10.1021/j100007a062`.

[155] A. Klamt, F. Eckert, and W. Arlt, *COSMO-RS: An alternative to simulation for calculating thermodynamic properties of liquid mixtures*, Annu. Rev. Chem. Biomol. Eng. **1** (2010) 101, DOI: `10.1146/annurev-chembioeng-073009-100903`.

[156] A. V. Marenich, C. J. Cramer, and D. G. Truhlar, *Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions*, J. Phys. Chem. B **113** (2009) 6378, DOI: `10.1021/jp810292n`.

[157] P. Su and H. Li, *Energy decomposition analysis of covalent bonds and intermolecular interactions*, J. Chem. Phys. **131** (2009), DOI: `10.1063/1.3159673`.

[158] Y. J. Franzke et al., *TURBOMOLE: Today and Tomorrow*, J. Chem. Theory Comput. **19** (2023) 6859, DOI: `10.1021/acs.jctc.3c00347`.

[159] G. Bistoni, *Finding chemical concepts in the Hilbert space: Coupled cluster analyses of noncovalent interactions*, Wiley Interdiscip. Rev. Comput. Mol. Sci. **10** (2020), DOI: `10.1002/wcms.1442`.

[160] A. Altun, R. Izsák, and G. Bistoni, *Local energy decomposition of coupled-cluster interaction energies: Interpretation, benchmarks, and comparison with symmetry-adapted perturbation theory*, Int. J. Quantum Chem. **121** (2021), DOI: `10.1002/qua.26339`.

[161] M. Hesse, H. Meier, and B. Zeeh, *Spektroskopische Methoden in der organischen Chemie*, Georg Thieme Verlag, 2005.

[162] M. Kaupp, M. Bühl, and V. G. Malkin, *Calculation of NMR and EPR Parameters: Theory and Applications*, ed. by M. Kaupp, M. Bühl, and V. G. Malkin, Wiley, 2004 1, doi: `10.1002/3527601678`.

[163] N. F. Ramsey, *Magnetic shielding of nuclei in molecules*, Phys. Rev. **78** (1950) 699, doi: `10.1103/PhysRev.78.699`.

[164] N. F. Ramsey, *Dependence of magnetic shielding of nuclei upon molecular orientation*, Phys. Rev. **83** (1951) 540, doi: `10.1103/PhysRev.83.540`.

[165] S. Grimme, C. Bannwarth, S. Dohm, A. Hansen, J. Pisarek, P. Pracht, J. Seibert, and F. Neese, *Fully Automated Quantum-Chemistry-Based Computation of Spin–Spin-Coupled Nuclear Magnetic Resonance Spectra*, Angew. Chem., Int. Ed. **56** (2017) 14763, doi: `10.1002/anie.201708266`.

[166] M. Jaszuński, K. V. Mikkelsen, A. Rizzo, and M. Witanowski, *A Study of the Nitrogen NMR Spectra of Azoles and their Solvent Dependence*, J. Phys. Chem. A **104** (2000) 1466, doi: `10.1021/jp994204h`.

[167] S. Ehlert, S. Grimme, and A. Hansen, *Conformational Energy Benchmark for Longer n -Alkane Chains*, J. Phys. Chem. A **126** (2022) 3521, doi: `10.1021/acs.jpca.2c02439`.

[168] L. Wittmann, C. E. Selzer, and S. Grimme, *A Diverse and Chemically Relevant Solvation Model Benchmark Set with Flexible Molecules and Conformer Ensembles*, Chem. Sci. (2025), doi: `10.1039/D5SC06406F`.

[169] H. J. Werner, T. B. Adler, and F. R. Manby, *General orbital invariant MP2-F12 theory*, J. Chem. Phys. **126** (2007), doi: `10.1063/1.2712434`.

[170] T. B. Adler, G. Knizia, and H. J. Werner, *A simple and efficient CCSD(T)-F12 approximation*, J. Chem. Phys. **127** (2007), doi: `10.1063/1.2817618`.

[171] H. J. Werner, G. Knizia, and F. R. Manby, *Explicitly correlated coupled cluster methods with pair-specific geminals*, Mol. Phys. **109** (2011) 407, doi: `10.1080/00268976.2010.526641`.

[172] C. Plett, S. Grimme, and A. Hansen, *Toward Reliable Conformational Energies of Amino Acids and Dipeptides - The DipCONFS Benchmark and DipCONL Datasets*, J. Chem. Theory Comput. **20** (2024) 8329, doi: `10.1021/acs.jctc.4c00801`.

[173] C. Plett, S. Grimme, and A. Hansen, *Conformational energies of biomolecules in solution: Extending the MPCONF196 benchmark with explicit water molecules*, J. Comput. Chem. **45** (2024) 419, doi: `10.1002/jcc.27248`.

[174] M. Friede, S. Ehlert, S. Grimme, and J.-M. Mewes, *Do Optimally Tuned Range-Separated Hybrid Functionals Require a Reparametrization of the Dispersion Correction? It Depends*, J. Chem. Theory Comput. **19** (2023) 8097, doi: `10.1021/acs.jctc.3c00717`.

[175] J. P. Perdew, *Density-functional approximation for the correlation energy of the inhomogeneous electron gas*, Phys. Rev. B **33** (1986) 8822, doi: `10.1103/PhysRevB.33.8822`.

[176] Y. Zhao and D. G. Truhlar, *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 function*, Theor. Chem. Acc. **120** (2008) 215, doi: `10.1007/s00214-007-0310-x`.

[177] C. Adamo and V. Barone, *Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models*, J. Chem. Phys. **108** (1998) 664, DOI: `10.1063/1.475428`.

[178] F. Neese, *Software Update: The ORCA Program System—Version 6.0*, WIREs Comput. Mol. Sci. **15** (2025) e70019, DOI: `10.1002/wcms.70019`.

[179] T. Gasevic, M. Bursch, Q. Ma, S. Grimme, H.-J. Werner, and A. Hansen, *Correction: The p-block challenge: assessing quantum chemistry methods for inorganic heterocycle dimerizations*, Phys. Chem. Chem. Phys. **27** (2025) 8572, DOI: `10.1039/d5cp90062j`.

[180] M. Veith, *Cage Compounds with Main-Group Metals*, Chem. Rev. **90** (1990) 3, DOI: `10.1021/CR00099A001/ASSET/CR00099A001.FP.PNG_V03`.

[181] S. Nakarnura, *Chapter 8 Group 111-V Nitride-Based Ultraviolet Blue-Green-Yellow Light-Emitting Diodes and Laser Diodes*, Semicond. Semimetals **48** (1997) 391, DOI: `10.1016/S0080-8784(08)62409-6`.

[182] H. Asahi, "Tl-Based III-V Alloy Semiconductors," *Infrared Detect. Emit. Mater. Devices*, Springer, Boston, MA, 2001 233, DOI: `10.1007/978-1-4615-1607-1_9`.

[183] K. Ota and R. Kinjo, *Inorganic Benzene Valence Isomers*, Chem. Asian J. **15** (2020) 2558, DOI: `10.1002/asia.202000535`.

[184] J. G. Ekerdt, Y. M. Sun, A. Szabo, G. J. Szulczewski, and J. M. White, *Role of surface chemistry in semiconductor thin film processing*, Chem. Rev. **96** (1996) 1499, DOI: `10.1021/CR950236Z/ASSET/CR950236Z.FP.PNG_V03`.

[185] F. C. Sauls and L. V. Interrante, *Coordination compounds of aluminum as precursors to aluminum nitride*, Coord. Chem. Rev. **128** (1993) 193, DOI: `10.1016/0010-8545(93)80030-9`.

[186] E. E. Foos, R. J. Jouet, R. L. Wells, A. L. Rheingold, and L. M. Liable-Sands, *Preparation of single-source precursors to nanocrystalline gallium arsenide and gallium antimonide. X-ray crystal structures of* $[Et_2GaAs(SiMe_3)_2]_2$, $[Et_2GaSb(SiMe_3)_2]_2$ *and* $Et_2GaAs(SiMe_3)_2Ga(Et)_2Sb(SiMe_3)_2$, J. Organomet. Chem. **582** (1999) 45, DOI: `10.1016/S0022-328X(98)01184-X`.

[187] A. M. Priegert, B. W. Rawe, S. C. Serin, and D. P. Gates, *Polymers and the p-block elements*, Chem. Soc. Rev. **45** (2016) 922, DOI: `10.1039/C5CS00725A`.

[188] S. Grimme and A. Hansen, *A Practicable Real-Space Measure and Visualization of Static Electron-Correlation Effects*, Angew. Chem., Int. Ed. **54** (2015) 12308, DOI: `10.1002/anie.201501887`.

[189] C. A. Bauer, A. Hansen, and S. Grimme, *The Fractional Occupation Number Weighted Density as a Versatile Analysis Tool for Molecules with a Complicated Electronic Structure*, Chem. Eur. J. **23** (2017) 6150, DOI: `10.1002/chem.201604682`.

[190] E. Ramos-Cordoba, P. Salvador, and E. Matito, *Separation of dynamic and nondynamic correlation*, Phys. Chem. Chem. Phys. **18** (2016) 24015, DOI: `10.1039/c6cp03072f`.

[191] K. A. Peterson, T. B. Adler, and H.-J. Werner, *Systematically convergent basis sets for explicitly correlated wavefunctions: The atoms H, He, B-Ne, and Al-Ar*, J. Chem. Phys. **128** (2008) 084102, DOI: `10.1063/1.2831537`.

[192] M. Bursch, H. Neugebauer, S. Ehlert, and S. Grimme, *Dispersion Corrected $r^2$SCAN Based Global Hybrid Functionals : $r^2$SCANh , $r^2$SCAN0 , and $r^2$SCAN50*, J. Chem. Phys. **156** (2022) 10, DOI: `10.1063/5.0086040`.

[193] S. Samanipour, L. P. Barron, D. van Herwerden, A. Praetorius, K. V. Thomas, and J. W. O'Brien, *Exploring the Chemical Space of the Exposome: How Far Have We Gone?* JACS Au **4** (2024) 2412, DOI: `10.1021/jacsau.4c00220`.

[194] Y. Khalak, G. Tresadern, D. F. Hahn, B. L. de Groot, and V. Gapsys, *Chemical Space Exploration with Active Learning and Alchemical Free Energies*, J. Chem. Theory Comput. **18** (2022) 6259, DOI: `10.1021/acs.jctc.2c00752`.

[195] A. M. Mroz, V. Posligua, A. Tarzia, E. H. Wolpert, and K. E. Jelfs, *Into the Unknown: How Computation Can Help Explore Uncharted Material Space*, J. Am. Chem. Soc. **144** (2022) 18730, DOI: `10.1021/jacs.2c06833`.

[196] H. Park, A. Onwuli, K. T. Butler, and A. Walsh, *Mapping inorganic crystal chemical space*, Faraday Discuss. **256** (0 2025) 601, DOI: `10.1039/D4FD00063C`.

[197] P. Pyykkö and M. Atsumi, *Molecular Single-Bond Covalent Radii for Elements 1–118*, Chem. Eur. J. **15** (2009) 186, DOI: `10.1002/chem.200800987`.

[198] R. Bauernschmitt and R. Ahlrichs, *Stability analysis for solutions of the closed shell Kohn–Sham equation*, J. Chem. Phys. **104** (1996) 9047, DOI: `10.1063/1.471637`.

[199] T. Froitzheim, M. Müller, A. Hansen, and S. Grimme, *g-xTB: A General-Purpose Extended Tight-Binding Electronic Structure Method For the Elements H to Lr (Z=1–103)*, ChemRxiv preprint (2025), DOI: `10.26434/chemrxiv-2025-bjxvt`.

[200] D. S. De, B. Schaefer, B. Von Issendorff, and S. Goedecker, *Nonexistence of the decahedral $Si_{20}H_{20}$ cage: Levinthal's paradox revisited*, Phys. Rev. B **101** (2020) 214303, DOI: `10.1103/PhysRevB.101.214303`.

[201] U. Mayer, V. Gutmann, and W. Gerger, *The acceptor number - A quantitative empirical parameter for the electrophilic properties of solvents*, Monatshefte für Chemie **106** (1975) 1235, DOI: `10.1007/BF00913599`.

[202] M. A. Beckett, G. C. Strickland, J. R. Holland, and K. S. Varma, *A convenient n.m.r. method for the measurement of Lewis acidity at boron centres: Correlation of reaction rates of Lewis acid initiated epoxide polymerizations with Lewis acidity*, Polymer (Guildf). **37** (1996) 4629, DOI: `10.1016/0032-3861(96)00323-0`.

[203] M. Bamberg, M. Bursch, A. Hansen, M. Brandl, G. Sentis, L. Kunze, M. Bolte, H.-W. Lerner, S. Grimme, and M. Wagner, *$[Cl@Si_{20}H_{20}]^-$: Parent Siladodecahedrane with Endohedral Chloride Ion*, J. Am. Chem. Soc. **143** (2021) 10865, DOI: `10.1021/jacs.1c05598`.

[204] J. Zheng, X. Xu, and D. G. Truhlar, *Minimally augmented Karlsruhe basis sets*, Theor. Chem. Acc. **128** (2011) 295, DOI: `10.1007/s00214-010-0846-z`.

# List of Figures

# List of Abbreviations

| | |
|---|---|
| **ADF** | Amsterdam Density Functional |
| **ALPB** | analytical linearized Poisson-Boltzmann |
| **AMAX** | absolute maximum error |
| **AMS** | Amsterdam Modeling Suite |
| **BJ** | Becke-Johnson |
| **BSIE** | basis set incompleteness error |
| **BSSE** | basis set superposition error |
| **CASSCF** | complete active space self-consistent field |
| **CBS** | complete basis set |
| **CC** | coupled cluster |
| **CENSO** | commandline energetic sorting |
| **CGTO** | contracted Gaussian-type orbital |
| **CLP** | confined Lewis pair |
| **COSMO** | conductor-like screening model |
| **CP** | counter-poise |
| **CPCM** | conductor-like polarizable continuum model |
| **CREST** | conformer-rotamer ensemble sampling tool |
| **DFA** | density functional approximation |
| **DFT** | density functional theory |
| **DH** | double hybrid |
| **DKH** | Douglas–Kroll–Hess |
| **EDA** | energy decomposition analysis |
| **ECP** | effective core potential |
| **EFG** | electric field gradient |
| **FF** | force field |
| **FOD** | fractional occupation density |

**GBSA** generalized Born and surface area

**gCP** geometrical counter-poise

**GGA** generalized gradient approximation

**GFN** geometry, frequency, noncovalent interactions

**GOAT** global optimizer algorithm

**GTO** Gaussian-type orbital

**HAHA** heavy-atom on heavy-atom

**HALA** heavy-atom on light-atom

**HF** Hartree-Fock

**LDA** local density approximation

**LCAO** linear combination of atomic orbitals

**LED** local energy decomposition

**LSDA** local spin density approximation

**MAD** mean absolute deviation

**MAE** mean absolute error

**MLIP** machine learning interatomic potential

**MLM** "mindless" molecule

**MO** molecular orbital

**MP2** second-order Møller-Plesset perturbation theory

**MR** multireference

**MR-CI** multi-reference configurational interaction

**MSE** mean signed error

**NCI** non-covalent interaction

**NMR** Nuclear Magnetic Resonance

**oDFB** ortho-difluorobenzene

**PGTO** primitive Gaussian-type orbital

**PP** pseudopotential

**PT2** second-order perturbation theory

**QC** quantum chemistry

**RMSE** root mean square error

**RRHO** rigid-rotor harmonic oscillator

**SCF** self-consistent field

**SD** standard deviation

**SIE** self-interaction error

| | |
|---|---|
| **SOC** | spin-orbit coupling |
| **SO** | spin-orbit |
| **SMD** | solvation model based on density |
| **SQM** | semiempirical quantum mechanical |
| **SR** | scalar relativistic |
| **SRB** | short-range bond |
| **STO** | Slater-type orbital |
| **UMA-sm** | small Universal Model for Atoms |
| **WFT** | wave function theory |
| **WTMAD-2** | weighted total mean absolute deviation in version 2 |
| **X2C** | exact two-component |
| **xTB** | extended tight-binding |
| **ZORA** | Zeroth Order Regular Approximation |