Chemical Space Exploration with Artificial "Mindless" Molecules

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

†*Mulliken Center for Theoretical Chemistry, University of Bonn, Germany*
‡*Microsoft Research AI for Science, Microsoft Research, Germany*

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

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 decomposition reactions mediated by H_2. This set provides a challenging benchmark for testing, validation, and training of density functional approximations (DFAs), semiempirical methods, force fields, and machine learning potentials using molecular structures beyond the conventional chemical space. For DFAs, we initially hypothesized that highly parameterized functionals might perform poorly on this set. However, no consistent relationship between fitting strategy and accuracy was observed. A clear Jacob's ladder trend emerges, with ωB97X-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 learned 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 with 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 energies of bonded systems, often at the expense of accurately modeling electron densities. Benchmark studies against high-level wavefunction-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.[25–30] Research into unconventional molecular structures has already been explored through datasets such as the "mindless" benchmarking set MB08-165,[31] its successor MB16-43,[30] the VQM24 dataset by von Lilienfeld et al.,[32] as well as the QCML dataset by Google DeepMind.[33] 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 dataset 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)[31] 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.[31] MLMs are particularly valuable for methods, that rely heavily on the 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 datasets. This enhanced diversity should improve the generalizability of ML-based approaches while reducing the required number of training samples.[34] 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 datasets within multi-fidelity 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.[35]

To assess the performance of computational methods and to reach uncharted chemical space, specialized generative strategies are required. Commonly, ML approaches are employed to generate new molecules by extrapolating from known structures.[36–38] 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 DFA parameterization schemes impact the accuracy and robustness for 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 semi-automatic 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 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 molecule generator: `MindlessGen`. Its key features are summarized in Figure 1 and will be described in the following.

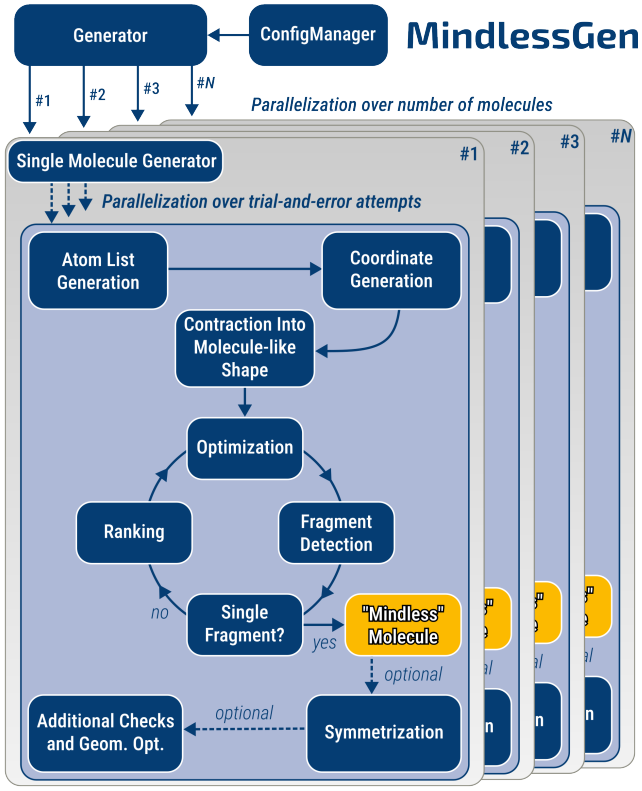

Figure 1: Outline of the working principle of `MindlessGen`.

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 GFNn-xtb,[39–42] 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 example `mindlessgen.toml` files available on the 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 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 GFNn-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 one for continued optimization. Graph nodes, reflecting covalent bonds, are set up by comparing interatomic distances with covalent radii by Pyykkö and Atsumi.[43] 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 post-processing steps may then be applied:

- Generation of symmetric arrangements for supramolecular complexes based on non-covalent 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 post-processing 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 MLM generation.
- Interfaces to various (S)QM programs, including `xtb-6.7.1`,[40] `TURBOMOLE-7.9.0`,[44] and `ORCA-6.0.1`[45] with support for easy extension through a standardized backend interface.

Exemplary "Mindless" Molecules

Figure 2 showcases exemplary "mindless" molecules (MLMs) generated 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[46] or amorphous indium oxide semiconductors.[47] 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 post-processing 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 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 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 dataset. Also, noble gases were not included since they are chemically too inert. Suiting system charges between -2 and +2 were randomly selected to ensure closed-shell systems and to add diversity. Additionally, various MR diagnostics were applied,

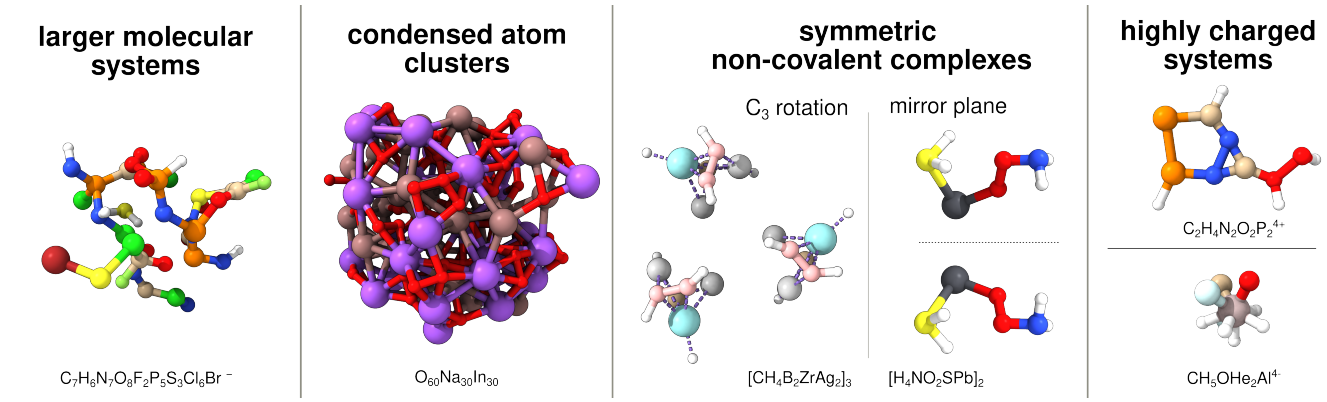

Figure 2: Exemplary "mindless" molecules generated using `MindlessGen`, clustered into different categories. The corresponding `mindlessgen.toml` input files used to generate these structures are available in the code repository at github.com/grimme-lab/MindlessGen.

including fractional occupation density (FOD)-based static correlation diagnostics[48–50] or the SCF stability analysis[51] (see Supporting Information for details).[52] 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 dataset if this is not restricted. Moreover, many generated MLMs also contain linear "snake"-like 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:

$$\begin{aligned} &\square\ \mathrm{C}_{\square}\mathrm{H}_{\square}\mathrm{B}_{\square}\mathrm{N}_{\square}\ldots + \square\ \mathrm{H}_2 \\ &\rightarrow \square\ \mathrm{CH}_4 + \square\ \mathrm{BH}_3 + \square\ \mathrm{N}_2 + \ldots \end{aligned} \quad (1)$$

If multiple molecular fragments are present within a single structure, 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. Charges were balanced with charged hydride species that were generated by adding or removing one hydrogen atom of the neutral products (e.g., OH^-, H_2O, H_3O^+). 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 towards hydride species and while conventional molecules are used as products, "mindless" molecules are only applied as reactants in the investigated reactions.

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).

Final geometry optimizations on the DFT level (PBE0-D4/def2-TZVP)[53–55] failed to converge for 9 systems, and 17 structures dissociated during this geometry optimization. Two more structures were excluded due to the FOD

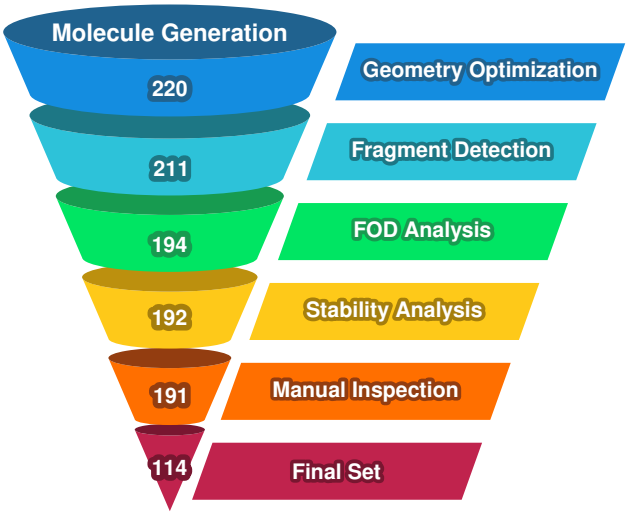

Figure 3: Exemplary screening process of MLMs after molecule generation. The respective volume elements of the funnel do not represent the actual set size. See text for more details.

analysis, and one more 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 diversity across the dataset. However, we note that parts of these filtering steps could also be included in the fully automated process.

The *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. Po-

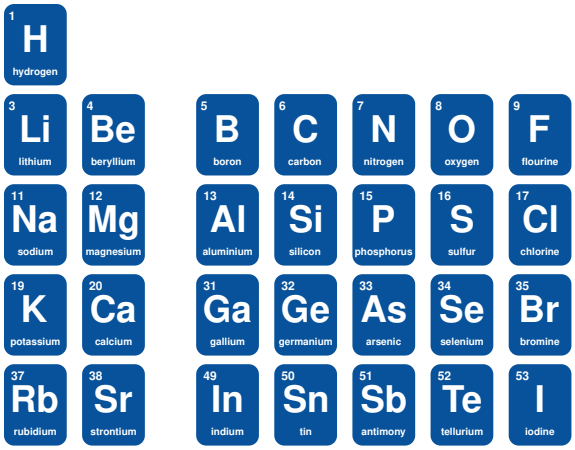

Figure 4: All elements included in the *MB2061* benchmark set.

tential 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[56–62] (as implemented in Molpro 2025.1)[60,63,64] applied with `TIGHT` domain settings, subvalence 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[65] and several further studies.[66–69] 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 *p*-block reactions, respectively, and may be a little bit larger for the heavier *s*-block reactions due to slightly less accurate basis set setup (see 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 our dataset. The new *MB2061* set is designed to be exceptionally diverse by incorporating various system

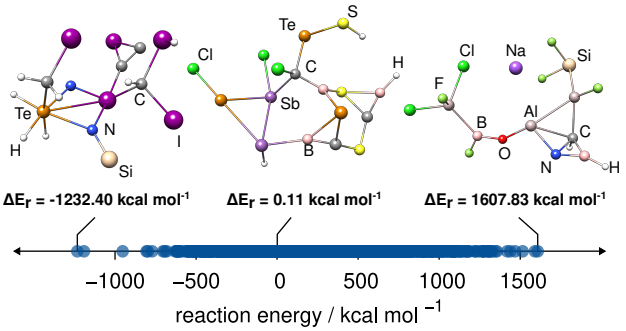

Figure 5: Range of reference reaction energies in the *MB2061* set with three example molecules.

charges, sizes, elements, binding partners, and binding motifs (cf. Figure 6).

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.

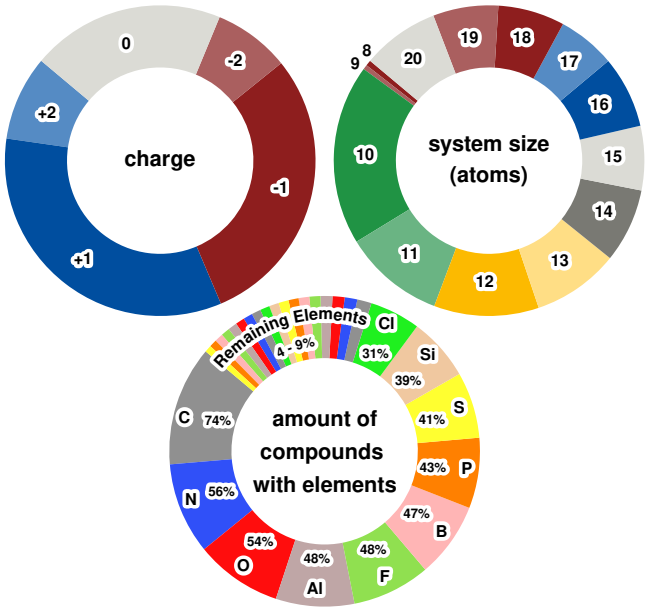

Figure 6: Composition of the *MB2061* set. The "amount of compounds with elements" indicates, for example, that 74 % of all compounds contain at least one carbon atom.

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 quantum mechanical (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 datasets 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. All molecular geometries, reference values, and reaction energy coefficients are provided in the Supporting Information.

Computational Details

Using the *MB2061* benchmark set, we assess the performance and robustness of one force field (FF),[70] six SQM methods,[39,40,71–75] seven (meta-) generalized gradient approximations (GGAs),[76–85] eleven (range-separated) hybrids,[53,86–95] nine double-hybrid,[96,96–101] five composite methods[102–106], and four MLIPs.[107–109] All DFAs (apart from composite methods) are used in combination with Ahlrichs' quadruple-ζ def2-QZVPP[55] basis set (with def2-ECPs) or with the aug-cc-pVQZ-PP-KS[65] (AVQZ-PP-KS) basis set, which is specifically re-contracted for the use of ECP10MDF[110–112] pseudopotentials (PPs) for the elements Ga-Br.[65] As density functionals do not sufficiently cover London dispersion effects, we tested the D3 London dispersion correction with zero damping[113] for Minnesota functionals and Becke-Johnson (BJ) damping[114]

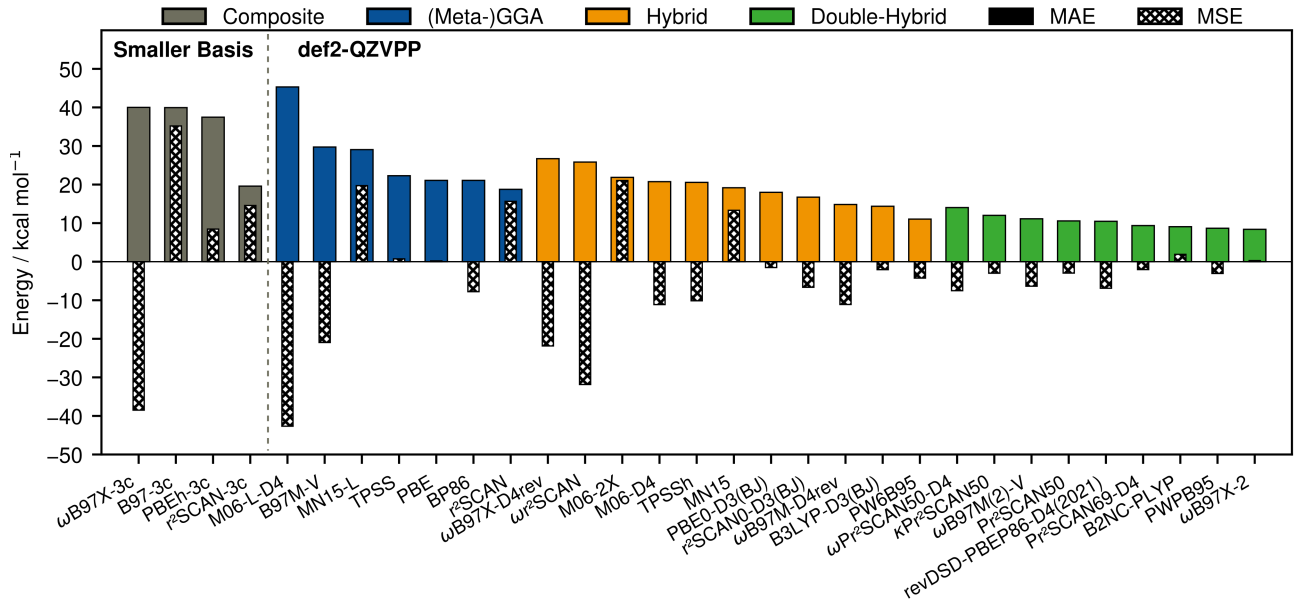

Figure 7: Mean absolute error (MAE) and mean signed error (MSE) of the best-performing combination of DFA and London dispersion correction for the whole *MB2061* set.

for the remaining functionals, the D4 dispersion correction[54,115–117] as well as the non-local DFT-NL dispersion correction[118,119] as introduced by Vydrov and van Voorhis (VV10)[120] in a post-SCF way. All respective calculations were conducted with the ORCA 6.0.0,[45] xTB 6.7.1,[41] Molpro_2025.1,[63,64], MOPAC 2016,[121] and fairchem 2.1.0[122] program packages. All tested methods are listed in Table S1 and Table S3 (see Supporting Information). D3 and D4 London dispersion corrections were computed with the respective standalone programs `s-dftd3`[123,124] and `dftd4`.[125] 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.*,[126] 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, especially considering their computational cost.

The heavily parameterized 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 MSEs 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.[31]

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 dataset 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 dataset can be divided into three subsets: *MB727-Light*, *MB782-s-Block*, and *MB552-Heavy*. The performance of selected DFA 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 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 outliers. This is particularly evident for functionals such as PW6B95 and ωB97M-D4rev, which show more pronounced deviations compared to 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 6.0 kcal·mol^{-1} to 13.7 kcal·mol^{-1} for ωB97X-2.

Relativity

Since relativistic effects become more significant for heavier elements, these trends might suggest that they are responsible for the observed deviations. However, it is important to note that such 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. One important difference remains: the reference energies were obtained using the ECP10MDF PP for elements Ga–Br, whereas the DFT calculations employed def2-ECPs, which are only applied 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[65] basis set, as well as with the exact two-component (X2C) relativistic method and the x2c-QZVPPall(-2c) 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 re-contracted basis for Ga–Br to support the ECP10MDF PP, ensuring a more consistent comparison in this context (see ref. 65 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 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 range and is therefore neglected in this

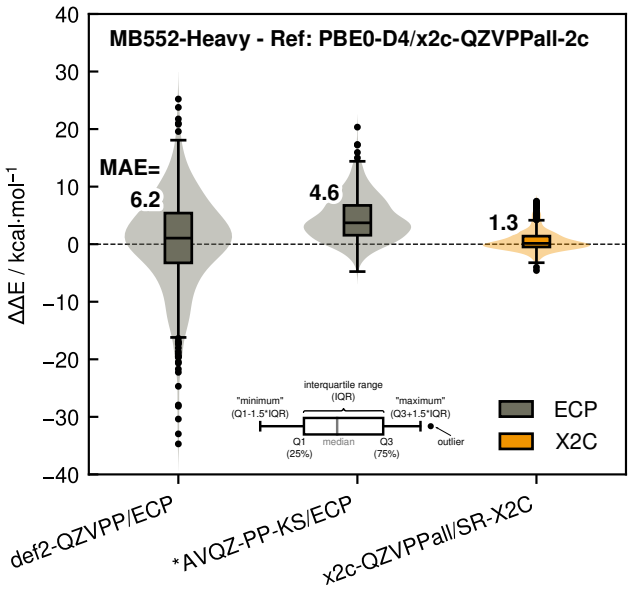

Figure 8: Violin plot showing the performance of PBE0-D4 in combination with ECPs and the explicit treatment of relativity using X2C. All errors are given relative to PBE0-D4/x2c-QZVPPall-2c (SO-X2C) reaction energies. *AVQZ-PP-KS is identical to def2-QZVPP, except for Ga–Br, where the basis functions were re-contracted for use with ECPs (see Supporting Information or Ref. 65).

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 also tested its performance with other density functionals. However, we observed SCF convergence issues with certain functionals such as r^{2}SCAN or ωB97M-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.

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)[70], all tight-binding methods from the GFNn-xTB family (n = 0, 1, 2),[39,40,74] the aTB[75] atomic density-based tight-binding model, the PM6-D3H4X[71,72] and PM7[73] 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 a 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) 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 dataset serves as a challenging benchmark for evaluating the robustness of computational methods and their underlying implementations. This is reflected by the number of non-converging systems or cases that converge to incorrect SCF minima. For example, aTB failed for 61 systems, GFN1-xTB for 67, and PM6-D3H4X for 90 systems.

ML-based models

By employing MLIP models, such as Aimnet2[107,127], 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* dataset that can be described using AIMNet2. We then compared its performance against DFT results (see 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 many of the tested DFAs. For example, PBE0 yields an MAE of 16.2 kcal·mol^{-1}, while PW6B95 achieves 10.5 kcal·mol^{-1} on the same subset. This indicates that the training set of AIMNet2 is already sufficiently broad and the model 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 non-metals. 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*)[108] data set was published, alongside some MLIPs that were trained on it.[108,109] 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 the *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 up to 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 datasets.

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 dataset contains 2061 high-level reference data points computed at the PNO-LCCSD(T)-F12 level, covering decomposition reactions of MLMs mediated by H_2, 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.

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 ωB97X-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 parameterized 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 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.
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 indicates 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.

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 chemical space. 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 *et al.*[128] Looking ahead, we anticipate that `MindlessGen` and the *MB2061* set will serve as a valuable resource for the parameterization and testing of next-generation models, ultimately contributing to more robust and transferable methods across the chemical space.

Acknowledgement 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 the access to the Marvin Computing Cluster of the University of Bonn. The authors also acknowledge Microsoft Accelerating Foundation Models Research Program for access to Microsoft Azure cloud compute. T.G. thanks Microsoft Research for funding. M.M. thanks the Fonds der Chemischen Industrie (FCI) for funding via a Kekulé scholarship.

Conflict of Interest

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.

Data and Software 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.
All calculated reaction energies are listed in the `esi2.xlsx` file (Supporting Information).
Additional technical details and supporting figures are also available in the Supporting Information (`esi.pdf`).

Supporting Information Available

Geometries (XYZ files), reference energies, and reaction energy coefficients of the *MB2061* benchmark set (ZIP)
Calculated reaction energies with SQM, DFT, and MLIPs (XLSX)
Additional technical details and supporting figures (PDF)

References

(1) Reymond, J.-L. The Chemical Space Project. *Accounts 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 Discov. 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. *Nature* **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. *Nature* **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.-M.; Hansen, A.; Grimme, S. Best-Practice DFT Protocols for Basic Molecular Computational

Chemistry. *Angew. Chem. - Int. Ed.* **2022**, *61*, 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. *Science* **2017**, *355*, 49–52.

(22) Kepp, K. P. Comment on "Density functional theory is straying from the path toward the exact functional". *Science* **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". *Science* **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*, e1737, e1737 CMS-1118.R1.

(25) 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.

(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.; others ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* **2019**, *47*, D930–D940.

(27) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; others PubChem 2023 update. *Nucleic Acids Res.* **2023**, *51*, D1373–D1380.

(28) 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.

(29) Axelrod, S.; Gómez-Bombarelli, R. GEOM, energy-annotated molecular conformations for property prediction and molecular generation. *Sci. Data* **2022**, *9*, 185.

(30) 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.

(31) Korth, M.; Grimme, S. Mindless DFT benchmarking. *J. Chem. Theo. Comp.* **2009**, *5*, 993–1003, Publisher: American Chemical Society.

(32) 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; `https://arxiv.org/abs/2405.05961`.

(33) 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*.

(34) 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.

(35) Gould, T.; Vuckovic, S. "Slim" benchmark sets for faster method development. 2025; http://dx.doi.org/10.26434/chemrxiv-2025-949xn.

(36) 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.

(37) 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.

(38) Anstine, D. M.; Isayev, O. Generative Models as an Emerging Paradigm in the Chemical Sciences. *J. Am. Chem. Soc.* **2023**, *145*, 8736–8750.

(39) 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.

(40) 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.

(41) 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*, e1493.

(42) 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.

(43) Pyykkö, P.; Atsumi, M. Molecular Single-Bond Covalent Radii for Elements 1–118. *Chem. Eur. J.* **2009**, *15*, 186–197.

(44) TURBOMOLE GmbH TURBOMOLE V7.9, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007. 2025; https://www.turbomole.org/.

(45) Neese, F. Software Update: The ORCA Program System—Version 6.0. *WIREs Comput. Mol. Sci.* **2025**, *15*, e70019.

(46) Shelby, J.; Noonan, J. Formation and properties of sodium indium silicate glasses. *Phys. Chem. Glas.* **1998**, *39*, 36–40.

(47) 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.

(48) Grimme, S.; Hansen, A. A practicable real-space measure and visualization of static electron-correlation effects. *Angew. Chem. - Int. Ed.* **2015**, *54*, 12308–12313.

(49) 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.

(50) Ramos-Cordoba, E.; Salvador, P.; Matito, E. Separation of dynamic and nondynamic correlation. *Phys. Chem. Chem. Phys.* **2016**, *18*, 24015–24023.

(51) Bauernschmitt, R.; Ahlrichs, R. Stability analysis for solutions of the closed shell Kohn–Sham equation. *J. Chem. Phys.* **1996**, *104*, 9047–9052.

(52) 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.

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

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

(55) 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.

(56) 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.

(57) 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.

(58) 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.

(59) 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). *Journal of chemical theory and computation* **2017**, *13*, 4871–4896.

(60) 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.

(61) 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.

(62) 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.

(63) Werner, H.-J. et al. MOLPRO, version 2025.1, a package of ab initio programs. 2025; see https://www.molpro.net.

(64) Werner, H.-J. et al. The Molpro quantum chemistry package. *J. Chem. Phys.* **2020**, *152*, 144107.

(65) 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.

(66) 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.

(67) 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. Theor. Comput. **2023**, *19*, 7007–7030.

(68) 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.

(69) 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.

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

(71) 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.

(72) S. Brahmkshatriya, P.; Dobeš, P.; Fanfrlík, J.; Řezáč, J.; Paruch, K.; Bronowska, A.; Lepšík, M.; Hobza, P. Quantum Mechanical Scoring: Structural and Energetic Insights into Cyclin-Dependent Kinase 2 Inhibition by Pyrazolo[1,5-a]pyrimidines. *Curr. Comput.-Aid. Drug.* **2013**, *9*, 118–129.

(73) 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.

(74) Pracht, P.; Caldeweyher, E.; Ehlert, S.; Grimme, S. A robust non-self-consistent tight-binding quantum chemistry method for large molecules. 2019; `http://dx.doi.org/10.26434/chemrxiv.8326202.v1`.

(75) 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.

(76) 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.

(77) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(78) Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33*, 8822–8824.

(79) 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.

(80) 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.

(81) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(82) 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.

(83) 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**, 9248.

(84) 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*, 061101.

(85) 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.

(86) Axel D. Becke Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5656.

(87) 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.

(88) 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.

(89) 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 noncovalent interactions. *Chem. Sci.* **2016**, *7*, 5032–5051.

(90) 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.

(91) 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.

(92) 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.

(93) 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.

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

(95) 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.

(96) Wittmann, L.; Neugebauer, H.; Grimme, S.; Bursch, M. Dispersion-corrected r2SCAN based double-hybrid

functionals. *J. Chem. Phys.* **2023**, *159*, 224103.

(97) Yu, F. Double-Hybrid Density Functionals Free of Dispersion and Counterpoise Corrections for Non-Covalent Interactions. *J. Phys. Chem. A* **2014**, *118*, 3175–3182.

(98) 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.

(99) 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.

(100) 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.

(101) Chai, J.-D.; Head-Gordon, M. Long-range corrected double-hybrid density functionals. *J. Chem. Phys* **2009**, *131*, 174105.

(102) Sure, R.; Grimme, S. Corrected Small Basis Set Hartree-Fock Method for Large Systems. *J. Comput. Chem.* **2013**, *34*, 1672–1685.

(103) 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.

(104) Grimme, S.; Hansen, A.; Ehlert, S.; Mewes, J. M. R2SCAN-3c: A "swiss army knife" composite electronic-structure method. *J. Chem. Phys.* **2021**, *154*, 064103.

(105) 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.

(106) 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.

(107) Isayev, O.; Anstine, D.; Zubaiuk, R. AIMNet2: A Neural Network Potential to Meet your Neutral, Charged, Organic, and Elemental-Organic Needs. *Chem. Sci.* **2025**, –.

(108) Levine, D. S. et al. The Open Molecules 2025 (OMol25) Dataset, Evaluations, and Models. 2025; `https://arxiv.org/abs/2505.08762v1`.

(109) UMA: A Family of Universal Models for Atoms | Research - AI at Meta (accessed: 22 Jun 2025). `https://ai.meta.com/research/publications/uma-a-family-of-universal-models-for-atoms`

(110) 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.

(111) 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.

(112) 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.

(113) 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.

(114) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry* **2011**, *32*, 1456–1465.

(115) 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.

(116) 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.

(117) 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.

(118) Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals density functional made simple. *Phys. Rev. Lett.* **2009**, *103*, 063004.

(119) 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.

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

(121) MOPAC2016, James J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, http://OpenMOPAC.net (2016).

(122) GitHub - facebookresearch/fairchem: FAIR Chemistry's library of machine learning methods for chemistry — github.com. `https://github.com/facebookresearch/fairchem`, [Accessed 04-06-2025].

(123) s-DFT-D3 at GitHub, https://github.com/dftd3/simple-dftd3 (accessed 6 Feb 2025).

(124) Ehlert, S. Simple DFT-D3: Library first implementation of the D3 dispersion correction. *Journal of Open Source Software* **2024**, *9*, 7169.

(125) DFT-D4 at GitHub, https://github.com/dftd4/dftd4 (accessed 6 Feb 2025).

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

(127) GitHub - isayevlab/AIMNet2 — github.com. `https://github.com/isayevlab/AIMNet2`, [Accessed 04-06-2025].

(128) Surzhikova, E.; Proppe, J. regAL: Python Package for Active Learning of Regression Problems. 2024; `https://arxiv.org/abs/2410.17917`.

TOC Graphic