

BH9, a New Comprehensive Benchmark Dataset for Barrier Heights and Reaction Energies: Assessment of Density Functional Approximations and Basis Set Incompleteness Potentials

Viki Kumar Prasad,^{1,*} Zhipeng Pei,¹ Simon Edelmann,^{1,†} Alberto Otero-de-la-Roza,^{2,‡} and Gino A. DiLabio^{1,§}

¹*Department of Chemistry, University of British Columbia,
3247 University Way, Kelowna, British Columbia, Canada V1V 1V7*

²*Departamento de Química Física y Analítica and MALTA Consolider Team,
Facultad de Química, Universidad de Oviedo, 33006 Oviedo, Spain*

(Dated: July 10, 2021)

The calculation of accurate reaction energies and barrier heights is essential in computational studies of reaction mechanisms and thermochemistry. In order to assess methods regarding their ability to predict these two properties, high-quality benchmark sets are required that comprise a reasonably large and diverse set of organic reactions. Due to the time-consuming nature of both locating transition states and computing accurate reference energies for reactions involving large molecules, previous benchmark sets have been limited in scope, the number of reactions considered, and the size of the reactant and product molecules. Recent advances in coupled-cluster theory, in particular local correlation methods like DLPNO-CCSD(T), now allow the calculation of reaction energies and barrier heights for relatively large systems. In this work, we present a comprehensive, and diverse benchmark set of barrier heights and reaction energies based on DLPNO-CCSD(T)/CBS, called BH9. BH9 comprises 449 chemical reactions belonging to nine types common in organic chemistry and biochemistry. We examine the accuracy of DLPNO-CCSD(T) vis-a-vis canonical CCSD(T) for a subset of BH9 and conclude that, although there is a penalty in using the DLPNO approximation, the reference data are accurate enough to serve as benchmark for density-functional theory (DFT) methods. We then present two applications of the BH9 set. First, we examine the performance of several density functional approximations commonly used in thermochemical and mechanistic studies. Second, we assess our basis set incompleteness potentials regarding their ability to mitigate basis set incompleteness error. The number of data points, the diversity of the reactions considered, and the relatively large size of the reactant molecules make BH9 the most comprehensive thermochemical benchmark set to date, and a useful tool for the development and assessment of computational methods.

INTRODUCTION

The prediction of barrier heights (BHs) and reaction energies (REs) using computational methods, combined with the application of transition-state theory,¹⁻³ is a powerful tool for the elucidation of reaction mechanisms in chemistry.^{4,5} The prediction of kinetic and thermochemical properties is also important in biochemistry, and has contributed greatly to the understanding of the catalytic activity of enzymes,⁶⁻⁹ as well as to the discovery of new drugs.^{10,11}

The main bottleneck for the successful prediction of rate constants and equilibrium constants is the accuracy in the determination of BHs and REs.^{3,10} Because of the exponential dependence of these constants on the corresponding energies, an accuracy of about RT (0.6 kcal/mol at room temperature) or better is required.¹² Quantum mechanical methods based on wavefunction theory,¹²⁻¹⁵ particularly recent composite methods, are able to calculate BHs and REs to this level of accuracy,^{16,17} but they are not applicable to molecules with sizes typically encountered in organic chemistry, let alone biochemistry.³ As a consequence of the tradeoff between accuracy and computational cost, the most popular method for thermochemical and kinetic calculations

in organic reactions is density-functional theory (DFT).¹⁰ In reactions of biochemical interest, where the reactant molecules are much larger, DFT is typically combined with force fields in hybrid quantum mechanics/molecular mechanics (QM/MM) approaches.^{6,7,18} In either case, the accuracy of the methods typically employed is often sufficient for gauging the relative energies of various mechanistic pathways but not enough to reliably predict rate constants of chemical reactions.^{10,19} Consequently, the search for a standard method for kinetic and mechanistic studies is still ongoing.^{4,5}

To develop new computational methods for the study of chemical reactions, and to assess the existing ones, high-quality benchmark sets are necessary.²⁰⁻²⁶ These benchmark sets comprise REs and BHs of model reactions calculated at a very accurate level of theory, typically coupled-cluster theory (CC) with large basis sets and a complete-basis-set (CBS) extrapolation¹⁶ (CCSD(T)/CBS is a very popular method¹²). Besides the obvious requirement that the reference data be accurate, there are a number of additional desirable traits for BH and RE benchmark sets. First, the set of reactions must be sufficiently large for the analysis to be statistically significant, and diverse enough to catch any particularities or biases of the method under study. For

instance, most DFT methods tend to underestimate BH of pericyclic reactions because the transition state (TS) is overstabilized due to delocalization error.^{27,28} Second, non-covalent interactions between reactants play an important role in stabilizing the TS.²⁹ The importance of this stabilization increases with the size of the reactant molecules and it is particularly important in biochemical studies where, for instance, the shape of the active site determines the activity and specificity of enzymes.⁴ Therefore, it is essential that the reactant molecules in the benchmark set are large enough to correctly assess the method under study regarding its ability to describe non-covalent interactions.^{18,29-32}

There are difficulties with the creation of benchmark sets for BHs and REs with the aforementioned characteristics. The generation of TSs is not easily automatized.^{10,19} More importantly, the computational cost involved in the calculation of accurate reference data limits the number of reactions in the set and the size of the reactant molecules. As a consequence, previously proposed benchmark sets use model reactions with small reactant molecules that are not representative of the typical reactions commonly found in mechanistic studies.³³⁻³⁶ Other benchmark sets either focus on specific types of reactions, or they contain only a handful of data points, or they are not evaluated using a reference level of enough quality to allow benchmarking commonly used quantum mechanical methods.^{28,31,32,37-43} The current necessity of a benchmark set for enzymatically catalyzed reactions has been emphasized several times recently.^{31,32,44}

Local correlation methods, particularly DLPNO-CCSD(T), have become very popular recently due to a favorable combination of relatively high accuracy and modest computational cost.⁴⁵⁻⁵⁰ Thanks to its near-linear-scaling nature, DLPNO-CCSD(T) can be applied to reasonably large systems.¹⁸ Since conventional CCSD(T)/CBS is at least two orders of magnitude more accurate than the methods typically assessed with BH and RE benchmark sets, a tradeoff is used in this work. By using DLPNO-CCSD(T)/CBS for the reference energies, we designed a benchmark set (called BH9) that has the desirable features listed above, namely, the reactions in BH9 are numerous and diverse and the reactants are relatively large. The accuracy penalty in using the DLPNO approximation^{30,51} is evaluated, providing an accuracy limit for the assessment of approximate methods. Variants of the DLPNO-CCSD(T)/CBS approach have been used in other recently proposed benchmark sets.^{31,32,52,53}

To our knowledge, BH9 is the most comprehensive benchmark set for BHs and REs of organic and bio-organic reactions to date. Our particular objective with this set is to aid in the development of atom-centered potentials⁵⁴⁻⁵⁶ (ACPs), whose training requires a large and diverse set of molecular properties. However, re-

cent machine-learning-based methods can equally benefit from using the BH9 data. Furthermore, the reference BHs and REs in BH9 can be recalculated should further developments in computational methods or computer hardware occur, without the need to find TSs for new reactions, a task that is often non-trivial.

We also present two simple applications of the new benchmark set. First, we use BH9 to assess several popular density functional approximations used in mechanistic studies. The effect of including corrections for dispersion interactions is considered, and we analyze the performance of these functionals individually for the different types of reactions included in the BH9 set. Second, the application of DFT to reaction mechanisms in practice often requires using a finite basis set due to computational constraints. Therefore, we also study the performance of our basis set incompleteness potentials^{54,55} (BSIP) regarding their ability to mitigate basis set incompleteness error in the calculation of REs and BHs.

DESIGN OF BH9 AND COMPUTATIONAL DETAILS

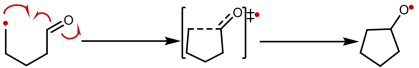
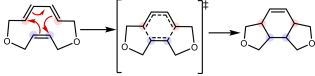
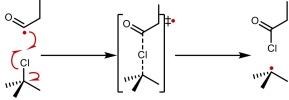
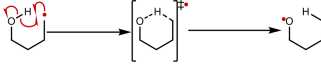
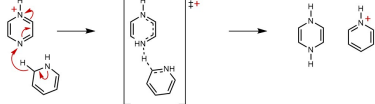
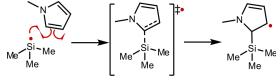
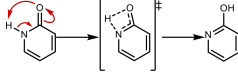
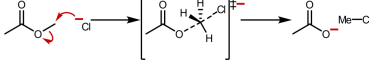
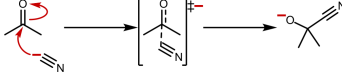
Design of the BH9 benchmark set

The BH9 set contains 449 elementary chemical reactions, categorized in the reaction types shown in Table I. The reference data comprises the corresponding 449 REs and 898 BHs (forward and reverse), as well as the structures of reactants, products, and transition states. Table I also shows a prototype reaction for each type. The full list of diagrams for each reaction is given in the Supporting Information (SI), as well as the reference BHs, REs, and the geometries of all the molecular species. The data for each reaction is given in the form of "db" files. This plain-text file format has been described elsewhere.^{57,58}

The reaction types in Table I represent a diverse set of reactions that are common in organic and bio-organic chemistry, although the list is by no means exhaustive. Most reaction types and many of the particular reactions included in the BH9 set were adopted from the Mechanism and Catalytic Site Atlas (M-CSA) database,⁵⁹ and, are known to occur in biological systems. However, some reaction types that are important in organic chemistry, such as pericyclic, hydride-transfer, and halogen-atom transfer reactions, are relatively rare in biological contexts or are not sufficiently represented in the M-CSA. For these reaction types, we explored the literature and compiled a number of reactions from various published mechanistic studies in order to complete our database.⁶⁰⁻¹⁰⁴

For the sake of simplicity, and due to our desire for this set to serve as a basis for ACP development, all molecules in BH9 contain exclusively elements common in organic

TABLE I. Reaction types in the BH9 set.

	Reaction type	Number ^a	Example reaction ^b
I	Radical rearrangement and addition	48	
II	Pericyclic	140	
III	Halogen atom transfer	43	
IV	Hydrogen atom transfer	90	
V	Hydride transfer	42	
VI	B- and Si- containing reactions	35	
VII	Proton transfer	10	
VIII	Nucleophilic substitution	15	
IX	Nucleophilic addition	26	

^a Number of reactions in each type.^b Example reaction for each type.

chemistry (H, C, N, O, F, P, S, Cl). We also included a specific set of reactions containing Si and B (reaction type VI in Table I). The fact that there are no transition metals in the BH9 reactions simplifies the application and interpretation of the tests based on this set, particularly regarding the application of BH9 to the assessment of DFT methods. RE and BH benchmark sets for reactions containing transition metals have been proposed recently, also at DLPNO-CCSD(T) level.^{52,53} The sizes of the reactant and TS molecules in BH9 range from 11 to 71 atoms—significantly larger than most previous sets, and typical of mechanistic studies.

Some of the reactions in BH9, particularly nucleophilic substitutions, nucleophilic additions, and proton transfer reactions, involve charged species. In this case, we expect the species involved in the reaction, and particularly reactants and products, to be greatly stabilized by interactions with the solvent or the environment. We ex-

perienced difficulties finding some of these TSs, which is why the number of reactions in these three categories is smaller than the others (see Table I). In addition, some of the BHs are negative, possibly because the solvent stabilizes reactants and products more than it stabilizes the TS. Although we eliminated very negative BHs from the set, some were left for diversity sake. A similar decision was taken by Iron et al. for their BH set for reactions involving transition metals.⁵²

Location of the transition states

Guess TSs were built for the 449 reactions in the BH9 set. This was a laborious process because of the difficulty in locating TS with the currently available algorithms in standard software packages. In addition to not being automatic, the TS search often failed entirely,

which explains the uneven number of reactions in each category of Table I. Because of their relatively large size and the abundance of reactions, reliably locating the minimum-energy conformer for each species in the BH9 is a formidable problem. However, in order for the BHs to still be representative of the corresponding reactions, we devised a protocol that explores the conformational landscape of reactants, products, and TS. This protocol, described below, ensures that the proposed structures are reasonably close in energy, if not identical, to the global energy minima of all species.

In all cases, we used the Gaussian 09/16^{105,106} software package. Our calculations employed a default SCF convergence criterion of 10^{-8} Hartree, “ultrafine” integration grid (pruned 99×590 grid), and tight optimization convergence criteria (maximum force = 1.5×10^{-5} Hartree/Bohr, RMS force = 1×10^{-5} Hartree/Bohr, maximum displacement = 6×10^{-5} Bohr, RMS displacement = 4×10^{-5} Bohr). The CalcFC and NoEigenTest options were used to specify the computation of force constants in the first step of the optimization and to suppress the curvature test during optimization, respectively. All calculations were carried out in the gas phase.

In the first step, preliminary TS were located by geometry optimization followed by a frequency calculation. Finding the TS is often difficult because a good initial guess for the TS geometry is required for the optimization to succeed. In difficult cases, we ran series of constrained geometry minimizations where we fixed a few geometric parameters, then used the resulting structure as the initial guess for the TS search. The preliminary TS optimizations used the B3LYP hybrid density functional,^{107,108} except in a few cases where the range-separated density functional CAM-B3LYP¹⁰⁹ was used. (The change in functional was prompted by the instability of B3LYP in the calculation of zwitterionic systems.) The D3 dispersion correction¹¹⁰ with Becke-Johnson damping^{111,112} was used in all cases. Due to the different sizes of the reactant molecules, depending on the reaction type, various Pople basis sets^{113–115} (6-31G*, 6-31+G*, 6-31+G**) were used together with their associated basis set incompleteness potentials⁵⁵ (BSIPs) to mitigate the effect of basis set incompleteness error (BSIE). 6-31G*-BSIP was used to model radical addition and pericyclic reactions. 6-31+G*-BSIP was used to model halogen atom transfer, nucleophilic substitution, nucleophilic addition, and the B- and Si-containing reactions. 6-31+G**-BSIP was used to model hydrogen atom transfer, hydride transfer, and proton transfer reactions. Each preliminary TS was checked for the presence of a single imaginary frequency and visually inspected to confirm the imaginary-frequency eigenvector was oriented along the reaction coordinate.

The preliminary TS were then subjected to a constrained conformer search using the commercial Schrödinger’s MacroModel Suite^{116,117} implemented in

the Maestro¹¹⁸ software package. This search is similar to the one used in previous works.^{119–121} Bonds undergoing breaking and formation in the TS had their bond distances constrained to their values in the preliminary TS. The conformational search was performed using the mixed torsional/large-scale low-mode sampling option in Maestro, followed by a constrained post-optimization with the OPLS all-atom force field.¹²² From this sampling, a maximum of 100 structures (fewer if the molecule was not sufficiently flexible) were then subjected to a single-point calculation using the same calculation level as in the preliminary TS optimization. A maximum of 9 lowest-energy conformers were chosen to undergo further refinement.

For each reaction, the nine TS conformers obtained in this manner plus the TS from the preliminary optimization were subjected to unconstrained optimization using the same method as above. We discarded all the structures whose optimization failed to locate a new TS or whose eigenvectors did not point in the direction of reactants and products. The lowest-energy conformer was then subjected to a final TS optimization and frequency calculation at a higher level of theory (CAM-B3LYP-D3(BJ)/6-311++G**^{123,124}). After verification of the imaginary frequencies and the direction of the imaginary-frequency eigenvector, this last structure was adopted as the TS for the reaction.

Reactant and product structures

The initial reactant and product structures were constructed from the optimized TS and subjected to geometry optimizations using CAM-B3LYP-D3(BJ) with the same combination of Pople basis sets and BSIPs as above (6-31G*, 6-31+G*, or 6-31+G** depending on reaction type). All geometry optimizations employed a default SCF convergence criterion of 10^{-8} Hartree, “ultrafine” integration grid, and the default optimization convergence criteria (maximum force = 4.5×10^{-4} Hartrees/Bohr, RMS force = 3×10^{-4} Hartrees/Bohr, maximum displacement = 1.8×10^{-3} Bohr, RMS displacement = 1.2×10^{-3} Bohr).

After this initial relaxation, a 100-step Monte-Carlo multiple minimum¹²⁵ (MCM) conformational search was carried out using the FullMonte^{126,127} software package. The conformers generated in this way were optimized with the semi-empirical PM6-DH2¹²⁸ method using the MOPAC2016¹²⁹ software package. All conformers were then subjected to a single point calculation at the same level of theory used for their initial optimization. The ten lowest-energy conformers were selected for further optimization at the same level. The resulting lowest energy conformer was subjected to a final optimization and frequency calculation at a higher level of theory (CAM-B3LYP-D3(BJ)/6-311++G**).

Reference energy calculations

The reference BHs and REs were obtained using single-point DLPNO-CCSD(T)^{45–50} at the equilibrium geometries of reactants, products, and TSs calculated as above. The favorable scaling of DLPNO-CCSD(T) makes it possible to apply CC to the fairly large molecules included in BH9, which is why this method has been often used to generate reference data in recent benchmark sets.^{31,52,53} Naturally, the use of the DLPNO approximation introduces an error compared to canonical CCSD(T). The reference energies are calculated using a focal-point approach to minimize the computational cost associated with using large basis sets.¹² The error introduced by DLPNO as well as the convergence of the reference data with respect to basis set size are examined in the Results and Discussion section.

For the calculation of the reference energies, we used the ORCA program, version 4.2.1.^{130,131} The aug-cc-pVNZ basis sets (in the following, aNZ for short) of Dunning and co-workers^{132–134} were used for the complete-basis-set extrapolation, as well as the resolution of the identity MP2 method^{135–137} (RI-MP2) with the aug-cc-pVNZ/C auxiliary basis sets.¹³⁸ The TightPNO and TightSCF threshold settings were used in the DLPNO-CCSD(T) calculation. The use of TightPNO was shown to be very important in the calculation of REs and BHs, particularly those of Diels-Alder reactions.³⁰ The frozen core approximation was used in all calculations. It has been shown to have a relatively minor impact on the accuracy of calculated thermochemical properties.¹²

DFT calculation details

DFT calculations were used to assess the performance of various density functional approximations commonly used in mechanistic studies¹⁰ on the BH9 set. We used Gaussian 16¹⁰⁶ to calculate the BH9 reactions using B3LYP,^{107,108} LC- ω PBE,^{139,140} M05-2X,¹⁴¹ M06-2X,¹⁴² revTPSS,¹⁴³ and ω B97XD.^{144,145} Ultrafine grids were used for all calculations. The BLYP,^{108,146} PBE,¹⁴⁷ TPSS,¹⁴⁸ BH&HLYP,^{108,149} PBE0,¹⁵⁰ and CAM-B3LYP¹⁰⁹ functionals were evaluated using ORCA, version 4.2.1.^{130,131} The tight SCF convergence criteria and the “grid4” integration grid were used. Second-order SCF was deactivated. The resolution of the identity (RI) method was used in all cases. For the hybrid functionals (BH&HLYP and PBE0), RI was applied to both the Coulomb and exchange integrals (RI-JK keyword). For the range-separated hybrid functional (CAM-B3LYP), the chain-of-spheres approximation¹⁵¹ was used to calculate the exchange energy. In all cases, the Def2-QZVPP basis set was used,¹⁵² with the corresponding auxiliary basis sets (Def2/JK and Def2/J) used where appropriate.¹⁵³ Contrary to previous reports,¹⁵⁴ we did

not observe any SCF convergence problems using the Minnesota functionals.

To evaluate the importance of dispersion, the exchange-hole dipole moment (XDM) model was used in combination with some of the functionals above.^{155,156} The canonical complete-basis-set XDM damping function parameters and the postg program were used.^{157,158} We expect the conclusions from this analysis to be transferable to other dispersion corrections, such as Grimme’s Dn family.¹¹⁰

Due to the typical size of the molecular species involved, the availability of computationally inexpensive methods for thermochemistry and kinetics is very important in the study of biochemical reactions.⁵⁶ One of the major factors impacting the accuracy of DFT methods in this context is BSIE, which arises from the finite nature of the basis sets employed. For this reason, we also examine the performance of our recently proposed basis set incompleteness potentials⁵⁵ (BSIPs) combined with several double- ζ basis sets in the description of REs and BHs. In particular, we evaluated the PBE0-XDM functional^{150,156} combined with the BSIP-corrected 6-31G*, 6-31+G*, 6-31+G**,^{113–115} Def2-SV(P), Def2-SVP,¹⁵² and pc-1 basis sets.^{159–162}

RESULTS AND DISCUSSION

Evaluation of the reference data

The most popular calculation level for benchmark sets is CCSD(T) with CBS extrapolation, which is known to yield sub-kcal/mol accuracy.^{12,163} As mentioned above, this level of theory is too computationally demanding and cannot be used to generate reference data for BH9, so we opted for DLPNO-CCSD(T) instead. We expect the two primary sources of error are our choice of basis set extrapolation strategy¹² and the application of the DLPNO approximation.³⁰ In this section, we evaluate the importance of both sources of error and we provide a reasonable estimate for the error bars associated with the BH9 reference data. Ultimately, this error estimate constitutes the accuracy limit of the BH9 set; methods more accurate than those applied here cannot be reliably assessed with this set.

The reference data was calculated using a focal-point approach,^{164,165} which has been shown to be an effective way of approaching the CBS limit in similar calculations.^{12,166,167} The BH9 reference energies are calculated as:

$$E = E_{\text{HF}}^{\text{a}\{\text{T},\text{Q}\}\text{Z}} + \Delta E_{\text{MP2}}^{\text{a}\{\text{T},\text{Q}\}\text{Z}} + \Delta E_{\text{DLPNO-CCSD(T)}}^{\text{aTZ}} \quad (1)$$

where $E_{\text{HF/a}\{\text{T},\text{Q}\}\text{Z}}$ is the HF energy calculated from the aTZ and aQZ energies using the CBS extrapolation for-

TABLE II. Subset of the BH9 reactions used for assessing the quality of the reference data.

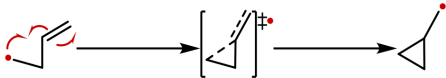
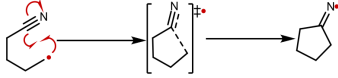
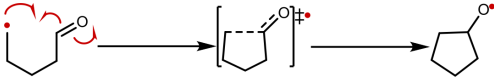
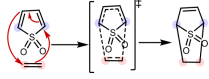
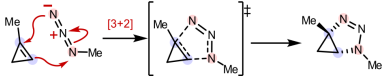
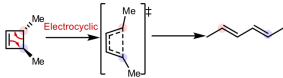
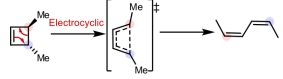
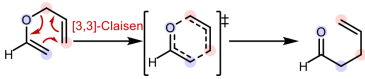
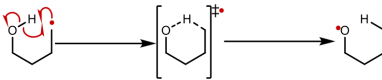
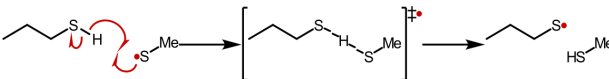
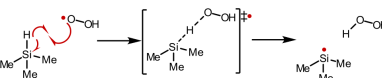
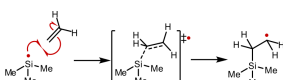
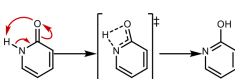
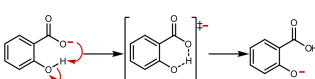
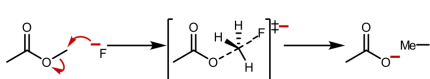
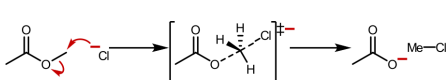
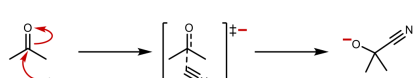
	Reaction	Type
1		Radical rearrangement (I)
2		Radical rearrangement (I)
3		Radical rearrangement (I)
4		Pericyclic (II)
5		Pericyclic (II)
6		Pericyclic (II)
7		Pericyclic (II)
8		Pericyclic (II)
9		Hydrogen atom transfer (IV)
10		Hydrogen atom transfer (IV)
11		Si-containing (VI)
12		Si-containing (VI)
13		Proton transfer (VII)
14		Proton transfer (VII)
15		Nucleophilic substitution (VIII)
16		Nucleophilic substitution (VIII)
17		Nucleophilic addition (IX)

TABLE III. Barrier heights for the reactions in Table II calculated using various levels of theory.^a

Reaction ^g	HF				ΔMP2 ^f				ΔCCSD(T) ^f		CCSD(T)			
	aTZ ^b	aQZ ^b	a5Z ^b	a{T,Q}Z ^c	a{Q,5}Z ^c	aTZ ^b	aQZ ^b	a5Z ^b	a{T,Q}Z ^c	a{Q,5}Z ^c	DLPNO aTZ ^b	Can. ^d aTZ ^b	DLPNO CBS ^e	Can. ^d CBS ^e
1	F	16.84	16.85	16.85	16.85	16.85	-1.25	-1.08	-1.27	-0.96	-4.15	-3.76	11.74	12.13
	R	10.09	10.15	10.15	10.17	10.14	5.30	5.46	5.58	5.57	-6.74	-6.35	9.01	9.39
2	F	16.38	16.42	16.44	16.43	16.44	4.80	5.13	5.24	5.36	-11.00	-10.80	10.80	11.00
	R	26.39	26.42	26.41	26.43	26.41	-4.18	-4.03	-3.96	-3.92	-1.98	-1.69	20.54	20.82
3	F	15.67	15.84	15.86	15.89	15.86	-0.96	-0.68	-0.55	-0.48	-6.85	-6.78	8.56	8.63
	R	19.79	19.70	19.69	19.67	19.69	-6.12	-6.01	-5.98	-5.93	-2.90	-2.62	10.84	11.13
4	F	42.42	43.16	43.46	43.36	43.51	-39.38	-38.45	-38.06	-37.78	11.41	10.46	16.98	16.03
	R	70.71	70.48	70.29	70.41	70.26	-15.48	-15.46	-15.39	-15.44	4.49	3.74	59.46	58.71
5	F	41.06	41.34	41.45	41.41	41.47	-35.95	-35.32	-35.19	-34.87	8.89	8.02	15.43	14.56
	R	87.78	87.91	87.88	87.95	87.88	-32.15	-31.60	-31.34	-31.21	11.55	10.69	68.29	67.44
6	F	41.85	41.97	41.97	42.00	41.97	-11.06	-10.93	-10.82	-10.83	1.91	1.62	33.08	32.78
	R	59.96	60.12	60.15	60.16	60.15	-18.88	-18.66	-18.61	-18.50	3.76	3.58	45.42	45.24
7	F	56.32	56.42	56.43	56.45	56.43	-14.28	-13.95	-13.81	-13.72	1.55	1.12	44.28	43.84
	R	70.99	71.10	71.15	71.13	71.15	-20.69	-20.34	-20.25	-20.08	3.30	3.01	54.35	54.06
8	F	50.55	50.77	50.83	50.83	50.84	-25.91	-25.53	-25.41	-25.27	6.51	5.68	32.08	31.25
	R	69.02	69.36	69.39	69.44	69.40	-25.53	-25.43	-25.30	-25.36	5.53	4.62	49.61	48.70
9	F	28.85	29.05	29.08	29.11	29.09	-9.04	-8.73	-8.60	-8.51	-4.18	-4.47	16.42	16.13
	R	33.81	33.93	33.96	33.96	33.96	-22.61	-22.81	-22.80	-22.95	1.36	1.06	12.36	12.06
10	F	21.25	21.42	21.55	21.47	21.57	-16.87	-16.83	-16.82	-16.81	0.13	-0.36	4.80	4.30
	R	21.34	21.51	21.63	21.55	21.65	-17.17	-17.10	-17.08	-17.05	0.08	-0.25	4.58	4.25
11	F	34.65	35.21	35.38	35.36	35.41	-22.00	-22.02	-21.91	-22.04	2.40	1.44	15.72	14.76
	R	19.87	19.85	19.78	19.84	19.77	-4.60	-4.32	-4.16	-4.12	-5.04	-5.70	10.68	10.02
12	F	8.54	8.64	8.66	8.66	8.66	-2.00	-1.60	-1.48	-1.32	-5.11	-5.10	2.24	2.25
	R	30.06	30.47	30.72	30.57	30.76	3.11	3.41	3.57	3.62	-5.84	-5.80	28.35	28.39
13	F	49.98	50.22	50.24	50.28	50.25	-15.90	-16.08	-16.11	-16.20	3.21	2.93	37.29	37.01
	R	51.08	51.28	51.31	51.33	51.32	-14.50	-14.64	-14.69	-14.75	1.62	1.37	38.20	37.96
14	F	2.70	2.76	2.77	2.78	2.77	-2.24	-2.19	-2.16	-2.16	0.22	0.26	0.84	0.87
	R	2.96	3.03	3.05	3.05	3.05	-2.95	-2.99	-2.96	-3.02	0.72	0.75	0.75	0.78
15	F	3.80	3.93	3.99	3.97	4.00	-6.03	-5.94	-5.77	-5.88	-1.29	-1.88	-3.21	-3.80
	R	25.57	25.87	26.01	25.95	26.03	-11.34	-11.27	-11.23	-11.23	0.34	-0.34	15.07	14.39
16	F	25.57	25.71	25.73	25.75	25.73	-11.53	-11.24	-11.20	-11.04	-0.42	-1.09	14.28	13.62
	R	2.72	3.28	3.63	3.43	3.68	-1.87	-1.90	-1.92	-1.91	-1.56	-2.12	-0.05	-0.61
17	F	13.08	13.40	13.43	13.48	13.44	-16.69	-16.60	-16.42	-16.54	2.11	1.68	-0.94	-1.38
	R	3.48	3.39	3.39	3.36	3.39	2.93	3.02	3.02	3.09	-1.64	-1.68	4.81	4.77
MAE		0.27	0.08	0.01	0.05	0.05	0.42	0.21	0.11	0.11	0.43		0.43	

^a Units are kcal/mol. ^b aNZ = aug-cc-pV/NZ basis set. ^c a{X,Y}Z = two-point HF or correlation energy extrapolation as described in the text. ^d Can. = canonical.^e CBS = complete-basis-set extrapolation (Eq. 1). ^f $\Delta\text{MP2} = E_{\text{MP2}} - E_{\text{HF}}$. $\Delta\text{CCSD(T)} = E_{\text{CCSD(T)}} - E_{\text{MP2}}$. ^g F = forward reaction, R = reverse reaction.

mula:

$$E_{\text{HF}}^L = E_{\text{HF}}^{\text{CBS}} + A \exp(-\alpha\sqrt{L}) \quad (2)$$

where L is the cardinal number of the basis set (3 for aTZ, 4 for aQZ, etc.). From this formula, a two-point extrapolation approach can be easily derived:

$$E_{\text{HF}}^{\text{CBS}} = \frac{E_{\text{HF}}^X \times e^{-\alpha\sqrt{Y}} - E_{\text{HF}}^Y \times e^{-\alpha\sqrt{X}}}{e^{-\alpha\sqrt{Y}} - e^{-\alpha\sqrt{X}}} \quad (3)$$

where X and Y are the cardinal numbers of the basis set pair. Following the recommendations of Neese and Valeev, we used the optimized $\alpha = 5.79$ value for the aTZ/aQZ pair.¹⁶⁸

The MP2 correlation energy (ΔE_{MP2}) is calculated using the known inverse cube dependence of the correlation energy with the basis set cardinal number:^{169,170}

$$\Delta E_{\text{MP2}}^L = \Delta E_{\text{MP2}}^{\text{CBS}} + AL^{-\beta} \quad (4)$$

with $\beta = 3$ in the large- L limit. This yields the two-point extrapolation formula:

$$\Delta E_{\text{MP2}}^{\text{CBS}} = \frac{E_{\text{MP2}}^X \times X^\beta - E_{\text{MP2}}^Y \times Y^\beta}{X^\beta - Y^\beta} \quad (5)$$

In practice, optimized β parameters have been proposed for some basis set pairs, and it has been shown that $\beta < 3$ improves the CBS estimate for low cardinal numbers.¹⁷¹ The $\beta = 3.05$ value proposed by Neese and Valeev for the aTZ/aQZ pair is used here.¹⁶⁸ Finally, the last component in our reference energy is:

$$\Delta E_{\text{DLPNO-CCSD(T)}}^{\text{aTZ}} = E_{\text{DLPNO-CCSD(T)}}^{\text{aTZ}} - E_{\text{MP2}}^{\text{aTZ}} \quad (6)$$

which is calculated using the aTZ basis set. The CCSD(T)/MP2 energy difference is routinely calculated at the aTZ level in the “gold standard” focal-point approach for non-covalent interactions,^{172–174} and it is justified by the observation that high-order contributions to the correlation energy converge relatively quickly with the basis set size.¹⁶³ In fact, our method for the calculation of reference data is very similar to the “gold standard” method for intermolecular interactions, except for the use of DLPNO for the CCSD(T) calculation.

To estimate the overall error in the BH9 reference data and to assess each of the approximations made, we selected a small subset of BH9 containing 17 reactions with relatively small molecules. This subset is shown in Table II. Since errors in the calculation of REs are typically lower than BHs,³⁰ we focus on the latter. The small size of the molecules in this subset allows the calculation of canonical CCSD(T)/aTZ energies, as well as HF/a5Z and MP2/a5Z. These last two quantities permit the calculation of the extrapolated aQZ/a5Z HF and MP2 barrier heights. For the MP2 correlation energy, we used the same two-point extrapolation formula (Eq. 5) with the

asymptotic value $\beta = 3$.^{169,170} For the HF energy, we used the extrapolation formula proposed by Karton and Martin for this particular basis set pair.^{175,176}

The BHs obtained with these methods are shown in Table III. We first consider the impact of basis set incompleteness on the individual components of our reference BHs. In the case of HF, our best CBS estimate (a{Q,5}Z extrapolation) agrees with HF/a5Z to within 0.01 kcal/mol on average, indicating that both are converged to within this value. Our chosen reference method for the HF component (a{T,Q}Z extrapolation) has a mean absolute error (MAE) of only 0.05 kcal/mol with respect to the aQZ/a5Z result. The highest deviations happen for reactions involving second-row atoms: numbers 6 (0.10 kcal/mol, both directions), 4 (0.15 kcal/mol, both directions), 8 (0.19 kcal/mol, reverse), and 12 (0.25 kcal/mol, reverse).

As noted above, the MP2 correlation energy converges more slowly to the CBS than the HF energy so, as expected, the basis set incompleteness errors are higher. The MP2 correlation contribution used in our reference method (a{T,Q}Z extrapolation) has the same MAE as MP2/a5Z (0.11 kcal/mol) compared to our best MP2/CBS estimate (a{Q,5}Z extrapolation). In this case, the large errors are not associated with second-row atoms, and they can be as high as half a kcal/mol (0.51 kcal/mol for forward reaction 1 and 0.32 kcal/mol for forward reaction 13). Combining the HF and MP2 results, we expect the average error from the HF+MP2 contribution to be in the vicinity of 0.2 kcal/mol, with worst cases being between 0.5 and 1 kcal/mol. Due to computational constraints, we cannot estimate the error introduced by calculating $\Delta E_{\text{CCSD(T)}}$ at aTZ level, although past experience with non-covalent interactions suggests that it is in the range of tenths of a kcal/mol or lower.¹⁷²

The last four columns in Table III show the error introduced by the DLPNO approximation by comparing the $\Delta E_{\text{CCSD(T)}}$ contribution and the total BH with and without DLPNO. The MAE from the DLPNO approximation is 0.43 kcal/mol, which is very similar to the 0.51 kcal/mol reported by Paiva et al. for enzymatic reactions.⁵¹ However, there are a few reactions where the deviations between DLPNO and canonical CCSD(T) are significantly higher, although lower than 1 kcal/mol in all cases: reactions 17 (0.83 kcal/mol and 0.91 kcal/mol), 14 (0.86 and 0.87 kcal/mol), and 4 (0.95 and 0.75 kcal/mol). The reactions for which the maximum deviation is observed are all pericyclic reactions, which agrees with the recent report by Sandler et al. who showed that DLPNO error is higher for dispersion-dominated and Diels-Alder BHs, with errors that can be as high as 1.2 kcal/mol.³⁰ The behavior of the errors in Table III confirm the relative difficulty of the DLPNO approximation in modeling large dispersion-dominated systems: All bimolecular BHs are overestimated, and the error for the forward and re-

verse reactions is approximately the same, indicating that the TSs are predicted to be too unstable by DLPNO. Based on these observations and the fact that the reaction subset in Table II contains the smallest molecules in BH9, we expect the 0.43 kcal/mol to be an overly optimistic error bar. An average error from the DLPNO approximation of around 0.5–1 kcal/mol for the reference data in BH9 is probably a more realistic estimate.

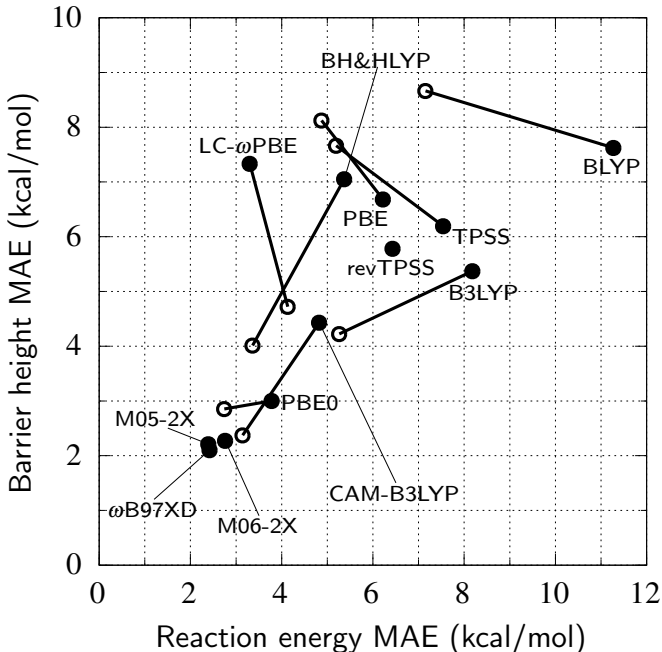
On the grounds of the preceding analysis, it is clear that the DLPNO approximation is the main contributing factor to the error in the BH9 reference data. Since basis set incompleteness is not the leading contribution to the error, our basis set extrapolation approach is justified.¹⁶⁶ Our analysis also shows that the estimated error is low enough to benchmark density functional approximations, which have typical errors in the range of a few kcal/mol²¹ (see below). The reference data can be revised in the future as more powerful computers and better algorithms become available.

Assessment of density functional theory methods

We now proceed to assess a few density functionals that are popular in mechanistic studies with the BH9 set. Our objectives are: i) evaluate whether the increased number of reactions and the larger molecules in the BH9 offer a picture of the performance of these functionals for thermochemistry and kinetics that is different from previous studies,^{21,154,156} ii) analyze the errors in the BH and RE calculations as a function of reaction type, and iii) benchmark the available XDM-corrected density functionals regarding their ability to calculate REs and BHs, something that has been done previously only with a very limited set of reactions.¹⁵⁶ We expect the inclusion of XDM dispersion to have a similar effect to D3, which has been extensively studied.^{21,154}

Tables IV and V show the mean absolute error (MAE) and mean error (ME) of the selected functionals for REs and BHs, grouped by type. The overall RE and BH MAEs are shown graphically in Figure 1. In agreement with previous studies,^{21,154,156} the performance of hybrid and range-separated hybrid functionals is, in general, much better than that of GGA functionals. Also, Figure 1 shows that there is a degree of positive correlation between RE and BH average errors, indicating that functionals that perform well for REs tend to work for BHs as well. The best-performing functionals are ω B97XD, M05-2X, and M06-2X with MAEs for both BHs and REs between 2 and 3 kcal/mol. The good performance of these functionals (or variants of ω B97X in combination with other dispersion corrections) has been noted in previous works.^{21,52} Close in performance, but with slightly higher MAEs are PBE0-XDM (MAE(RE) = 2.74 kcal/mol; MAE(BH) = 2.85 kcal/mol) and CAM-B3LYP-XDM (MAE(RE) = 3.14 kcal/mol; MAE(BH) =

FIG. 1. Barrier height vs. reaction energy mean absolute errors (MAE) for the chosen functionals. Open symbols represent the XDM-corrected version of each functional.



2.37 kcal/mol). B3LYP performs relatively poorly both in RE and BH, and so does its XDM-corrected version with average errors slightly over 4 kcal/mol.

The good performance of M05-2X and M06-2X is interesting. It is known that these functionals underestimate non-covalent interaction energies at long range.^{177,178} Since the importance of long-range dispersion increases with molecular size,¹⁷⁹ one would have expected a degradation in the performance of these functionals for the BH9 relative to previous studies of REs and BHs involving smaller molecules.^{21,177} However, this does not seem to be the case, and our average errors are similar to those reported by Mardirossian et al.¹⁷⁷ and Goerigk et al.²¹ Because TSs of addition reactions are larger than either the reactant or product molecules, an underestimation of non-covalent binding would lead to an erroneously unstable TS and an overestimation of the BHs. This seems to be the case for M05-2X and M06-2X, as indicated by the MEs in Table V. However, the average bias is only 0.72 (M05-2X) and 1.30 kcal/mol (M06-2X) for REs and 0.86 (M05-2X) and 1.05 kcal/mol (M06-2X) for BHs, suggesting that capturing the correct asymptotic dependence of the dispersion contribution seems not to be as important for the calculation of REs and BHs as previously argued.²¹ This point is reinforced by the fact that the performance of M05-2X and M06-2X in the GMTKN database is only marginally improved by their combination with the D3 dispersion correction.^{21,154}

Compared to our previous analysis of the performance of XDM-corrected functionals for REs and BHs,¹⁵⁶

TABLE IV. Average errors in the BH9 reaction energies using various density functionals.^a

Functional		I	II	III	IV	V	VI	VII	VIII	IX	Total
BLYP	MAE	8.42	22.27	8.75	3.17	2.35	11.70	2.82	4.01	10.84	11.27
	ME	8.42	20.96	-1.76	-1.24	0.01	10.59	0.87	-0.78	10.84	8.47
BLYP-XDM	MAE	5.27	13.87	7.77	2.77	2.24	3.92	2.70	3.50	4.65	7.15
	ME	5.24	12.51	-1.09	-1.17	-0.02	3.11	0.64	-0.87	4.56	4.61
PBE	MAE	3.34	10.52	8.71	3.24	2.49	5.79	3.00	4.04	3.67	6.22
	ME	1.56	9.36	-1.95	-1.02	1.52	4.78	0.50	-1.29	2.44	3.32
PBE-XDM	MAE	2.68	7.59	8.36	3.12	2.48	2.26	3.00	4.00	3.14	4.87
	ME	0.68	6.36	-1.72	-1.01	1.49	1.36	0.44	-1.43	0.02	1.90
TPSS	MAE	4.10	13.81	8.34	3.14	2.45	8.22	2.48	3.72	5.57	7.54
	ME	2.96	12.84	-1.56	-0.88	1.24	7.12	0.66	-1.24	5.49	4.95
TPSS-XDM	MAE	2.82	8.87	7.73	2.94	2.51	3.01	2.44	3.13	2.87	5.19
	ME	1.26	7.79	-1.16	-0.85	1.24	2.10	0.54	-1.37	1.64	2.62
revTPSS	MAE	3.32	10.91	7.82	3.37	2.82	7.46	2.12	3.76	3.95	6.43
	ME	1.22	9.92	-1.20	-0.76	1.58	6.48	0.64	-1.36	3.76	3.80
B3LYP	MAE	5.63	15.65	7.12	2.40	2.56	8.91	1.81	2.77	8.02	8.18
	ME	5.62	14.75	-0.47	-0.91	-1.67	7.94	0.61	-0.28	8.02	5.90
B3LYP-XDM	MAE	3.69	9.85	6.37	2.06	2.45	2.78	1.76	2.16	3.68	5.26
	ME	3.62	8.89	0.00	-0.87	-1.69	2.20	0.47	-0.44	3.49	3.20
PBE0	MAE	2.69	5.52	6.25	2.19	1.73	4.14	1.77	2.57	2.12	3.78
	ME	-0.81	3.91	-1.05	-0.46	-0.50	3.18	0.23	-0.59	0.92	1.18
PBE0-XDM	MAE	2.40	3.18	5.89	2.02	1.63	1.39	1.77	2.22	2.49	2.74
	ME	-1.67	0.95	-0.83	-0.46	-0.52	-0.22	0.17	-0.72	-1.48	-0.22
BH&HLYP	MAE	3.46	9.42	4.80	1.60	4.11	6.31	0.87	2.07	5.44	5.37
	ME	3.32	8.78	0.14	-0.31	-3.02	5.42	0.45	0.66	5.42	3.53
BH&HLYP-XDM	MAE	2.22	5.36	4.33	1.33	3.97	2.39	0.86	1.43	2.39	3.36
	ME	2.01	4.58	0.46	-0.32	-3.01	0.95	0.36	0.53	2.10	1.57
M05-2X	MAE	1.61	3.12	4.55	1.26	2.54	1.73	0.92	1.83	1.69	2.39
	ME	0.28	2.30	1.10	-0.43	-1.65	0.89	-0.08	0.03	0.75	0.72
M06-2X	MAE	2.31	4.00	4.72	1.56	1.47	2.36	0.91	1.84	1.71	2.76
	ME	0.51	3.71	0.80	-0.63	-0.58	1.76	-0.14	-0.56	1.36	1.30
CAM-B3LYP	MAE	2.98	8.37	5.29	1.62	2.62	5.89	1.08	1.82	4.67	4.82
	ME	2.72	7.83	-0.29	-0.52	-1.60	5.10	0.57	0.07	4.65	3.13
CAM-B3LYP-XDM	MAE	2.10	5.03	4.87	1.43	2.57	2.11	1.08	1.30	2.14	3.14
	ME	1.71	4.42	-0.02	-0.51	-1.64	1.19	0.50	-0.12	1.80	1.51
LC- ω PBE	MAE	4.28	4.38	5.90	1.66	1.73	2.98	1.19	1.82	1.72	3.30
	ME	-3.68	-2.57	0.11	-0.18	1.00	1.68	0.76	0.20	0.01	-0.97
LC- ω PBE-XDM	MAE	4.83	6.84	5.54	1.63	1.69	2.99	1.18	1.50	2.62	4.13
	ME	-4.83	-6.18	0.39	-0.16	1.13	-2.04	0.69	0.21	-2.57	-2.62
ω B97XD	MAE	1.76	3.03	5.27	1.71	1.32	1.94	1.34	1.19	1.69	2.42
	ME	0.08	2.04	-0.08	-0.48	-0.11	-0.42	0.49	-0.22	1.24	0.57

^a Units are kcal/mol. MAE = mean absolute error. ME = mean error. The roman numerals represent the reaction types in Table II.

the advantages of a more complete benchmark set are very evident. In our previous work, LC- ω PBE-XDM (MAE = 1.43 kcal/mol) and BH&HLYP-XDM (MAE = 2.38 kcal/mol) were the best-performing functionals for BHs.¹⁵⁶ This is in stark contrast with the results in Table V and Figure 1, where the MAEs of these functionals rise to 4.72 and 4.01 kcal/mol, respectively. The cause of this disagreement is very likely the limited size of the benchmark set used in our previous work,¹⁵⁶ the small size of the molecular species, and the fact that it contained only hydrogen atom transfer reactions. Still, the results in Table IV, Table V, and Figure 1 are encouraging and suggest that expanding the list of functionals

with which XDM has been combined could increase the applicability of the method to chemical problems other than modeling non-covalent interactions.

As expected, the effect of including the XDM dispersion energy agrees, in general terms, with previous reports in the literature using D3.^{21,29,154,180,181} The inclusion of XDM has a noticeable impact on REs and BHs. Uncorrected GGAs severely overestimate REs, with an MAE that can be as high as 11.27 kcal/mol (BLYP). The overestimation is less pronounced for uncorrected hybrid and range-separated hybrid functionals. The inclusion of XDM dispersion partially corrects the overestimation of the REs and reduces the MAEs by several kcal/mol

TABLE V. Average errors in the BH9 barrier heights using various density functionals.^a

Functional		I	II	III	IV	V	VI	VII	VIII	IX	Total
BLYP	MAE	5.01	12.19	7.54	4.87	5.05	7.35	3.40	4.00	5.73	7.62
	ME	-3.77	-5.94	-6.46	-1.97	0.29	2.74	-1.56	-2.09	-3.45	-3.33
BLYP-XDM	MAE	5.05	10.26	11.86	8.51	10.67	4.45	3.62	8.72	6.16	8.66
	ME	-5.05	-10.07	-11.86	-8.49	-10.42	-3.06	-3.53	-8.72	-5.89	-8.44
PBE	MAE	3.85	7.98	8.59	7.01	7.59	3.99	5.51	4.04	4.77	6.68
	ME	-3.69	-6.55	-8.48	-6.62	-6.57	-1.46	-4.86	-3.27	-4.51	-5.78
PBE-XDM	MAE	4.18	8.20	10.99	9.63	12.04	4.39	5.65	6.32	5.57	8.12
	ME	-4.18	-8.16	-10.99	-9.60	-11.96	-4.29	-5.64	-6.29	-5.44	-8.07
TPSS	MAE	3.89	9.04	6.88	4.84	4.78	5.23	3.84	4.04	4.44	6.19
	ME	-3.52	-6.41	-6.38	-3.08	-1.81	-0.10	-2.63	-3.13	-3.85	-4.17
TPSS-XDM	MAE	4.33	9.07	10.04	7.56	9.66	4.30	3.97	7.57	5.47	7.66
	ME	-4.33	-9.05	-10.04	-7.52	-9.47	-4.13	-3.85	-7.57	-5.36	-7.61
revTPSS	MAE	3.86	8.12	6.99	4.53	4.43	5.03	3.14	4.08	4.19	5.78
	ME	-3.65	-6.54	-6.66	-2.92	-2.06	-1.13	-1.83	-3.44	-3.85	-4.31
B3LYP	MAE	2.98	8.08	4.27	3.66	5.69	5.88	1.75	3.58	4.10	5.37
	ME	-1.04	-0.89	-1.08	1.08	4.81	4.23	0.11	2.02	-0.95	0.52
B3LYP-XDM	MAE	2.34	5.51	5.81	4.03	4.05	2.43	1.58	3.09	3.13	4.22
	ME	-1.97	-3.91	-5.17	-3.95	-3.73	-0.33	-1.28	-2.99	-2.68	-3.38
PBE0	MAE	1.46	3.34	3.47	3.08	3.82	2.72	2.76	3.85	1.62	3.00
	ME	-0.28	-0.05	-0.76	-0.95	1.53	1.56	-1.91	2.93	-1.01	-0.05
PBE0-XDM	MAE	1.43	2.54	3.87	3.98	4.09	1.55	2.76	1.52	2.13	2.85
	ME	-0.77	-1.66	-3.25	-3.91	-3.89	-1.28	-2.69	-0.06	-1.93	-2.33
BH&HLYP	MAE	2.55	7.27	7.54	6.82	14.49	7.21	3.36	7.03	3.39	7.05
	ME	2.25	6.68	6.69	6.81	14.49	7.18	3.36	6.98	3.07	6.73
BH&HLYP-XDM	MAE	1.81	4.92	4.43	2.99	7.50	3.71	2.30	3.22	1.96	4.01
	ME	1.55	4.41	3.45	2.85	7.48	3.52	2.30	3.06	1.77	3.67
M05-2X	MAE	0.96	1.94	2.90	1.56	6.30	1.72	1.26	2.97	1.06	2.21
	ME	0.03	0.15	1.38	0.36	5.70	0.53	-0.32	2.25	-0.57	0.86
M06-2X	MAE	1.61	2.39	2.96	1.36	4.99	1.66	1.11	3.14	1.13	2.27
	ME	1.06	0.92	2.03	-0.30	4.44	0.37	-0.22	2.76	-0.28	1.05
CAM-B3LYP	MAE	1.83	5.06	4.18	3.20	9.34	5.21	1.21	4.89	2.41	4.43
	ME	1.19	3.86	2.28	2.47	9.05	5.05	0.16	4.65	1.29	3.52
CAM-B3LYP-XDM	MAE	1.27	3.16	2.55	1.68	3.50	2.26	1.07	2.13	1.19	2.37
	ME	0.63	2.03	-0.46	-0.86	3.13	1.85	-0.70	1.31	0.25	0.96
LC- ω PBE	MAE	3.70	9.15	8.47	4.82	13.79	5.38	1.40	10.38	3.70	7.33
	ME	3.49	9.12	8.31	4.66	13.37	5.18	0.13	10.38	3.63	7.16
LC- ω PBE-XDM	MAE	3.16	7.15	5.59	1.52	7.35	2.60	1.22	6.87	2.84	4.72
	ME	2.88	7.13	5.32	1.12	6.86	2.09	-0.89	6.87	2.46	4.42
ω B97XD	MAE	1.13	2.66	2.97	1.73	2.17	1.58	0.90	2.99	1.21	2.10
	ME	0.65	2.26	1.91	-1.20	1.46	0.80	0.04	2.79	0.58	1.04

^a Units are kcal/mol. MAE = mean absolute error. ME = mean error. The roman numerals represent the reaction types in Table II.

in general, except in the case of LC- ω PBE. These observations can be explained by the fact that the overall RE error is dominated by addition reactions, where the product molecule is the combination of both reactants. Functionals without dispersion underestimate the stability of the addition products, resulting in erroneously high REs.

In the case of BHs, Figure 1 and Table V show that the MAEs for the uncorrected functionals are in the range 4–8 kcal/mol. In particular, all GGA and meta-GGA functionals severely underestimate the BHs. This is explained by delocalization error,²⁷ the tendency of approximate density functionals to overstabilize delocalized molecules.

The TS are, in general, more delocalized than reactants and products. Consequently, they are spuriously stabilized, resulting in an erroneously low BH.²⁷ Delocalization error severely affects GGAs, while admixture of exact exchange in global and range-separated hybrids mitigates, but does not eliminate, this problem. The effect of including the XDM dispersion energy on BHs can be understood as well. The dispersion stabilization increases with the size of the molecule, so including the dispersion energy always leads to lower BHs. For GGAs functionals, which spuriously underestimate BHs, inclusion of XDM results in an increased MAE. For hybrid and range-separated hybrid functionals, which do not suffer as much

from delocalization error, the use of XDM decreases the MAEs. This is consistent with previous analysis in the literature.^{27,154,156}

We now analyze the performance of the chosen functionals on the various reaction types of the BH9 set using the data in Tables IV and V. For a few representative functionals, the RE and BH MAEs as a function of reaction type are shown in Figure 2. In the case of the REs, there are large differences between reaction types regarding the performance of various functionals and the effect of dispersion. Reaction types IV (hydrogen atom transfer), V (hydride transfer), VII (proton transfer), and VIII (nucleophilic substitution) seem to be modelable with approximately the same error by all uncorrected and dispersion-corrected functionals, in the range 2–4 kcal/mol. However, hybrid and range-separated hybrid functionals are, again, slightly better than GGAs. Reactions I (mostly radical rearrangements) and III (halogen atom transfer) show higher errors (up to around 8 kcal/mol for BLYP), are better modeled by hybrid or range-separated functionals, and the inclusion of dispersion corrections has a relatively minor impact. Reaction types VI (B- and Si-containing reactions) and IX (nucleophilic addition) show similar or larger errors than I and III and hybrid and range-separated hybrid functionals outperform GGAs. However, in this case, the inclusion of dispersion interactions improves the functional performance by several kcal/mol. These observations are easily explained by the fact that types VI and IX comprise addition reactions, while the other reaction types mentioned are rearrangements or atom transfer reactions. Since dispersion interactions stabilize larger molecules, their inclusion alleviates the overestimation of the REs in these reactions by the uncorrected functionals, as mentioned above. (Note that most REs for addition reactions are negative. Uncorrected functionals yield overestimated REs in general; their REs are above this negative reference value, but smaller in magnitude.) Lastly, the pericyclic reactions (II) show the highest errors, possibly due to the effect of varying delocalization between reactants and products, and benefit from dispersion for the same reason as VI and IX, since most of the members of this category are addition reactions.

Figure 2 and Table V show that the MAEs for BHs are higher than for REs, and that the inclusion of dispersion has comparatively more impact. BHs are more accurately represented by hybrid functionals, particularly if they are dispersion corrected, than by GGA functionals for all reaction types. For types I (radical rearrangement) and VII (proton transfer), the effect of including dispersion is minimal, and the accuracy is entirely controlled by the base functional, with hybrid and range-separated hybrid functionals showing much better performance. In reactions II (pericyclic) and VI (B- and Si-containing reactions), including dispersion interactions either has no effect or is beneficial, regardless of the functional type.

Figure 2 also shows the MAE for the forward and reverse BHs separately. Reactions II and VI are particular in that the effect of dispersion is very noticeable in the forward reaction BHs, but it is not for the reverse reaction BHs. This is reasonable because both categories comprise addition reactions. The dispersion stabilization of the product is essentially the same as the TS, but higher than for the reactant molecules. For the rest of the reactions, the inclusion of dispersion increases the MAE of the GGA functionals and decreases (in general) the MAEs of hybrid and range-separated hybrid functionals, for the reasons stated above.

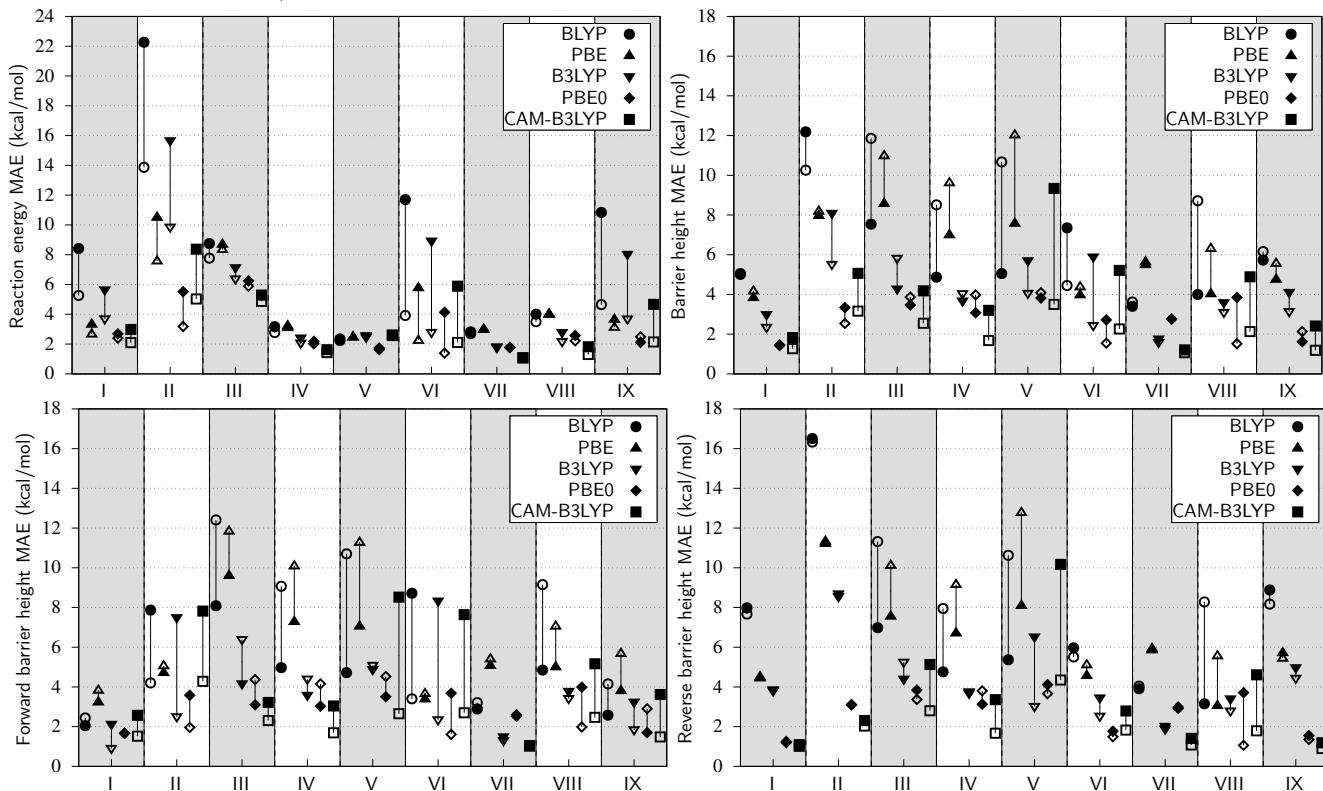
The fact that the inclusion of dispersion interactions decreases the MAE for (forward) BHs in pericyclic reactions (II) is slightly surprising in light of our previous discussion regarding delocalization error. Given the delocalized nature of the TS in pericyclic reactions, we expected a severe underestimation of the forward BH by GGAs and a subsequent increase in the MAE upon application of XDM. We can interpret this by noting that the reacting molecules are larger in the pericyclic reactions than in other reactions of the BH9 set, which suggests that non-covalent interactions have a comparatively more important role in the stabilization of the TS than electronic delocalization. Omitting dispersion interactions from the functional destabilizes the TS more than the spurious stabilizing effect from delocalization error, and therefore the inclusion of XDM is beneficial. In any case, these results emphasize that relying on error cancellation for the calculation of molecular properties is ill-advised.

Assessment of basis set incompleteness potentials

One of our objectives in the construction of the BH9 set is to provide training data for the development of atom-centered potentials (ACPs).^{54–56,182} ACPs are one-electron potentials that are designed to correct for the shortcomings of the DFT method to which they are applied. One particular flavor of ACPs are the basis set incompleteness potentials^{54,55} (BSIPs) whose purpose is to minimize the basis set incompleteness error (BSIE) that originates from using small or minimal basis sets in DFT calculations. The application of BSIPs allows computing molecular properties with a quality similar to a complete basis set but at a much reduced computational cost.

The development of ACPs requires a relatively large training set of molecular properties. BSIPs, in particular, are constructed by minimizing the deviation between the BSIP-corrected small-basis-set values and the complete-basis-set values for a number of molecular properties. Since both the approximate and the reference molecular properties are calculated using the same functional and BSIE is mostly functional-independent, this ensures that BSIPs are mostly transferable between functionals,

FIG. 2. Reaction energy (top left), overall barrier height (top right), forward reaction barrier height (bottom left), and reverse reaction barrier height (bottom right) mean absolute errors (MAEs) as a function of reaction type and density functional (using the Def2-QZVPP basis set). Open symbols represent the XDM-corrected version of each functional.



and are tied only to the basis set for which they were developed.⁵⁴

The training set for the recently developed BSIPs contained only 316 REs and 102 BHs, out of a total of 9,372 molecular properties.⁵⁵ Therefore, it is interesting to examine whether BSIPs decrease BSIE in the calculation of REs and BHs by applying them to the BH9 set. For this test, we used the BSIPs from our previous work.⁵⁵ Of the 15 basis sets for which BSIPs were developed, only six had thermochemical data in their training set (6-31G*, 6-31+G*, 6-31+G**, Def2-SV(P), Def2-SVP, and pc-1), so we restrict our analysis to these basis sets. We chose PBE0-XDM as the base functional for this analysis in order to check the transferability of BSIPs across functionals. (These BSIPs were developed using B3LYP.⁵⁵)

The MAEs for the REs and BHs in the BH9 set using BSIP-corrected and uncorrected PBE0-XDM in combination with the aforementioned basis sets are shown in Table VI. The table shows two sets of MAEs with respect to different reference data: the DLPNO-CCSD(T)/CBS values (“Ref.”) and our complete-basis-set PBE0-XDM estimate using the Def2-QZVPP basis set (“CBS”). Since the objective of BSIP development is to minimize BSIE, comparison with the CBS results is the purest measure of performance. The uncorrected MAEs (“Bare”) in Table VI show that the magnitude of the BSIE on average

is between 1.4 kcal/mol and 4 kcal/mol, depending on the basis set, and that there are no significant differences between BHs and REs regarding BSIE. When BSIPs are applied, the MAEs with respect to the CBS values decrease in all cases, by up to 2 kcal/mol, bringing the results to a reasonably close agreement with the Def2-QZVPP results. The performance of BSIPs is better for the larger basis sets, 6-31+G* and 6-31+G**, where the MAEs with respect to the CBS reference are lower than 1 kcal/mol for both BHs and REs. There seems to be no salient differences between the effect of BSIPs on REs and BHs. These results are encouraging because of the aforementioned sparsity of thermochemical and kinetic data in the training set, which suggests that BSIPs have robust performance for systems significantly different from those in the training set. Also, because a functional different from PBE0-XDM was used in their development, our results suggest a strong transferability of BSIPs across functionals.

The MAEs between the small-basis-set BSIP-corrected and uncorrected PBE0-XDM results and the DLPNO-CCSD(T)/CBS data are also shown in Table VI (“Ref.” column). In this case, the particular MAE values result from a combination of two errors: the uncorrected BSIE and the errors from the PBE0-XDM functional itself. Application of BSIPs reduces the MAEs in general to values

TABLE VI. Mean absolute errors (MAE) for the BH9 barrier heights and reaction energies of PBE0-XDM with several BSIP-corrected and uncorrected basis sets.^a

		Reaction energies		Barrier heights	
		CBS ^b	Ref. ^c	CBS ^b	Ref. ^c
6-31G*	Bare	3.59	3.97	2.46	4.27
	BSIP	1.91	3.46	1.62	3.26
6-31+G*	Bare	2.49	3.01	1.48	3.30
	BSIP	0.89	2.65	0.75	2.83
6-31+G**	Bare	1.81	2.66	1.42	3.48
	BSIP	0.64	2.65	0.63	2.81
Def2-SV(P)	Bare	4.12	4.65	3.11	4.96
	BSIP	1.95	3.64	1.80	2.91
Def2-SVP	Bare	3.12	3.97	2.90	4.87
	BSIP	2.19	3.97	1.74	3.06
pc-1	Bare	2.42	2.86	2.60	4.80
	BSIP	1.41	3.17	1.32	3.18
Def2-QZVPP		2.74		2.85	

^a Units are kcal/mol. The statistics correspond to the whole BH9 set. ^b CBS = MAEs calculated with respect to our complete-basis-set estimate (Def2-QZVPP).

^c Ref. = MAEs calculated with respect to the DLPNO-CCSD(T)/CBS reference data for the BH9.

that are close to the MAE of PBE0-XDM/Def2-QZVPP, particularly for the BHs, for which the BSIP-corrected MAEs are at most 0.41 kcal/mol above the Def2-QZVPP MAE. However, in some cases the MAE is unaffected or increases slightly due to favorable error cancellation in the uncorrected results.

CONCLUSIONS

In this article we introduce the BH9 set, an extensive and diverse benchmark dataset for reaction energies (REs) and barrier heights (BHs) in organic and bio-organic reactions. The BH9 set comprises 449 diverse reactions (449 REs and 898 BHs) involving relatively large molecular species (up to 71 atoms), similar to those found in thermochemical and mechanistic studies. The molecular species in BH9 comprise main-group elements, particularly those typically found in organic and bio-organic chemistry (H, C, N, O, F, P, S, and Cl) plus B and Si.

The computational level for the BH9 reference data is DLPNO-CCSD(T) combined with a focal-point approach in order to minimize errors from basis set incompleteness. We used a small subset of the BH9 composed of small molecular species to evaluate the errors introduced by our approximations and to estimate an error bar for the BH9 reference data. The DLPNO approximation is the main source of error, in comparison with basis set incompleteness. We estimate that the overall accuracy of the benchmark is in the vicinity of 1 kcal/mol or better.

The newly created BH9 was applied in two ways. First, we benchmarked a few popular density function-

als used in the literature for calculating REs and BHs, as well as some XDM-corrected functionals to evaluate the effect of dispersion interactions on REs and BHs. In general, hybrid and range-separated hybrid functionals perform much better than GGA functionals. The two Minnesota functionals M05-2X and M06-2X and the ω B97XD functional had the lowest mean absolute errors (MAEs), between 2 and 3 kcal/mol for both BHs and REs. The XDM-corrected functionals PBE0-XDM and CAM-B3LYP-XDM closely followed these functionals in terms of performance, with MAEs not above 3 kcal/mol. We also verified that delocalization error is a major contribution to the BH and RE errors. However, for the reactions involving large molecular species (e.g. some pericyclic reactions), the incorrect treatment of dispersion seems to outweigh delocalization error for non-dispersion-corrected functionals.

Lastly, we applied the BH9 set to analyze the performance of our basis set incompleteness potentials (BSIPs) for REs and BHs in combination with a few double- ζ basis sets and the PBE0-XDM functional. We found that, despite the fact that thermochemical and kinetic data were only a small part of their training set and that they were developed using a different functional (B3LYP), BSIPs performed excellently, reducing the discrepancy between the double- ζ and the complete-basis-set results by a factor of around 1.5 to 2. For BHs, the BSIP-corrected double- ζ MAEs were at most 0.41 kcal/mol higher than the Def2-QZVPP MAE, and the calculations were immensely less expensive. This confirms BSIPs are a robust way of minimizing basis set incompleteness from finite basis sets.

To our knowledge, BH9 is the most comprehensive BH and RE benchmark set to date. We hope that it will be useful to assess and develop new methods for thermochemical and kinetic work.

SUPPORTING INFORMATION

Table containing the list of reactions for the BH9 set organized by type, and the reference reaction energy and barrier heights. Plain-text db files for all reactions. Plain-text xyz files for all molecular structures involved in the reactions.

ACKNOWLEDGEMENTS

GAD would like to thank the Natural Sciences and Engineering Research Council of Canada, the Canadian Foundation for Innovation, and the British Columbia Knowledge Development Fund for financial support. We are grateful to Compute Canada, Westgrid, and the University of British Columbia’s Advanced Research Computing platforms for the generous allocation of comput-

ing resources. AOR. thanks the Spanish Ministerio de Ciencia e Innovación and the Agencia Estatal de Investigación (AEI) for financial support (projects PGC2018-097520-A-100 and RED2018-102612-T), and the Spanish MINECO for a Ramón y Cajal fellowship (RyC-2016-20301).

* ORCID:0000-0003-0982-3129

† Current address: Department of Chemistry, McGill University, 801 Sherbrooke St. West, Montréal, Québec, Canada H3A 0B8.

‡ aoterodelaroza@gmail.com

§ gino.dilabio@ubc.ca

- [1] Laidler, K. *Chemical Kinetics*; Harper & Row, 1987.
- [2] Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. Current status of transition-state theory. *J. Phys. Chem.* **1996**, *100*, 12771–12800.
- [3] Klippenstein, S. J.; Pande, V. S.; Truhlar, D. G. Chemical kinetics and mechanisms of complex systems: a perspective on recent theoretical advances. *J. Am. Chem. Soc.* **2014**, *136*, 528–546.
- [4] Bachrach, S. M. Challenges in computational organic chemistry. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 482–487.
- [5] Houk, K.; Liu, F. Holy grails for computational organic chemistry and biochemistry. *Acc. Chem. Res.* **2017**, *50*, 539–543.
- [6] van der Kamp, M. W.; Mulholland, A. J. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry* **2013**, *52*, 2708–2728.
- [7] Ahmadi, S.; Barrios Herrera, L.; Chehelamirani, M.; Hostaš, J.; Jalife, S.; Salahub, D. R. Multiscale modeling of enzymes: QM-cluster, QM/MM, and QM/MM/MD: A tutorial review. *Int. J. Quantum Chem.* **2018**, *118*, e25558.
- [8] Himo, F. Recent trends in quantum chemical modeling of enzymatic reactions. *J. Am. Chem. Soc.* **2017**, *139*, 6780–6786.
- [9] Quesne, M. G.; Borowski, T.; de Visser, S. P. Quantum mechanics/molecular mechanics modeling of enzymatic processes: Caveats and breakthroughs. *Chem. Eur. J.* **2016**, *22*, 2562–2581.
- [10] Engkvist, O.; Norrby, P.-O.; Selmi, N.; Lam, Y.-h.; Peng, Z.; Sherer, E. C.; Amberg, W.; Erhard, T.; Smyth, L. A. Computational prediction of chemical reactions: current status and outlook. *Drug Discov. Today* **2018**, *23*, 1203–1218.
- [11] Lam, Y.-h.; Abramov, Y.; Ananthula, R. S.; Elward, J. M.; Hilden, L. R.; Nilsson Lill, S. O.; Norrby, P.-O.; Ramirez, A.; Sherer, E. C.; Mustakis, J., et al. Applications of Quantum Chemistry in Pharmaceutical Process Development: Current State and Opportunities. *Org. Process Res. Dev.* **2020**, *24*, 1496–1507.
- [12] Peterson, K. A.; Feller, D.; Dixon, D. A. Chemical accuracy in ab initio thermochemistry and spectroscopy: current strategies and future challenges. *Theor. Chem. Acc.* **2012**, *131*, 1–20.
- [13] Karton, A.; Rabinovich, E.; Martin, J. M.; Ruscic, B. W4 theory for computational thermochemistry: In pursuit of confident sub-kJ/mol predictions. *J. Chem. Phys.* **2006**, *125*, 144108.
- [14] Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 theory. *J. Chem. Phys.* **2007**, *126*, 084108.
- [15] Harding, M. E.; Vázquez, J.; Ruscic, B.; Wilson, A. K.; Gauss, J.; Stanton, J. F. High-accuracy extrapolated ab initio thermochemistry. III. Additional improvements and overview. *J. Chem. Phys.* **2008**, *128*, 114111.
- [16] Karton, A. A computational chemist's guide to accurate thermochemistry for organic molecules. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2016**, *6*, 292–310.
- [17] Karton, A.; Sylvetsky, N.; Martin, J. M. W4-17: A diverse and high-confidence dataset of atomization energies for benchmarking high-level electronic structure methods. *J. Comput. Chem.* **2017**, *38*, 2063–2075.
- [18] Bistoni, G.; Polyak, I.; Sparta, M.; Thiel, W.; Neese, F. Toward accurate QM/MM reaction barriers with large QM regions using domain based pair natural orbital coupled cluster theory. *J. Chem. Theory Comput.* **2018**, *14*, 3524–3531.
- [19] Deglmann, P.; Schäfer, A.; Lennartz, C. Application of quantum calculations in the chemical industry—An overview. *Int. J. Quantum Chem.* **2015**, *115*, 107–136.
- [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] 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 non-covalent interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184–32215.
- [22] Goerigk, L.; Mehta, N. A trip to the density functional theory zoo: warnings and recommendations for the user. *Aust. J. Chem.* **2019**, *72*, 563–573.
- [23] Morgante, P.; Peverati, R. ACCDB: A collection of chemistry databases for broad computational purposes. *J. Comput. Chem.* **2019**, *40*, 839–848.
- [24] Yu, H. S.; Zhang, W.; Verma, P.; He, X.; Truhlar, D. G. Nonseparable exchange–correlation functional for molecules, including homogeneous catalysis involving transition metals. *Phys. Chem. Chem. Phys.* **2015**, *17*.
- [25] 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.
- [26] Haoyu, S. Y.; 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.
- [27] Johnson, E. R.; Mori-Sánchez, P.; Cohen, A. J.; Yang, W. Delocalization errors in density functionals and implications for main-group thermochemistry. *J. Chem. Phys.* **2008**, *129*, 204112.
- [28] Yu, L.-J.; Sarrami, F.; O'Reilly, R. J.; Karton, A. Reaction barrier heights for cycloreversion of heterocyclic rings: An Achilles' heel for DFT and standard ab initio procedures. *Chem. Phys.* **2015**, *458*, 1–8.

- [29] Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. Inclusion of dispersion effects significantly improves accuracy of calculated reaction barriers for cytochrome P450 catalyzed reactions. *J. Phys. Chem. Lett.* **2010**, *1*, 3232–3237.
- [30] Sandler, I.; Chen, J.; Taylor, M.; Sharma, S.; Ho, J. Accuracy of DLPNO-CCSD (T): Effect of Basis Set and System Size. *J. Phys. Chem. A* **2021**, *125*, 1553–1563.
- [31] Wappett, D. A.; Goerigk, L. Toward a quantum-chemical benchmark set for enzymatically catalyzed reactions: important steps and insights. *J. Phys. Chem. A* **2019**, *123*, 7057–7074.
- [32] Wappett, D. A.; Goerigk, L. A guide to benchmarking enzymatically catalysed reactions: the importance of accurate reference energies and the chemical environment. *Theor. Chem. Acc.* **2021**, *140*, 1–15.
- [33] Zheng, J.; Zhao, Y.; Truhlar, D. G. Representative benchmark suites for barrier heights of diverse reaction types and assessment of electronic structure methods for thermochemical kinetics. *J. Chem. Theory Comput.* **2007**, *3*, 569–582.
- [34] Karton, A.; Tarnopolsky, A.; Lam re, J.-F.; Schatz, G. C.; Martin, J. M. Highly accurate first-principles benchmark data sets for the parametrization and validation of density functional and other approximate methods. Derivation of a robust, generally applicable, double-hybrid functional for thermochemistry and thermochemical kinetics. *J. Phys. Chem. A* **2008**, *112*, 12868–12886.
- [35] Zhao, Y.; Gonz lez-Garc a, 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* **2005**, *109*, 2012–2018.
- [36] 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.
- [37] Chan, B.; Simmie, J. M. Barriometry—an enhanced database of accurate barrier heights for gas-phase reactions. *Phys. Chem. Chem. Phys.* **2018**, *20*, 10732–10740.
- [38] Goerigk, L.; Sharma, R. The INV24 test set: how well do quantum-chemical methods describe inversion and racemization barriers? *Can. J. Chem.* **2016**, *94*, 1133–1143.
- [39] Karton, A.; Goerigk, L. Accurate reaction barrier heights of pericyclic reactions: Surprisingly large deviations for the CBS-QB3 composite method and their consequences in DFT benchmark studies. *J. Comput. Chem.* **2015**, *36*, 622–632.
- [40] Karton, A.; O’Reilly, R. J.; Radom, L. Assessment of theoretical procedures for calculating barrier heights for a diverse set of water-catalyzed proton-transfer reactions. *J. Phys. Chem. A* **2012**, *116*, 4211–4221.
- [41] Karton, A.; O’Reilly, R. J.; Chan, B.; Radom, L. Determination of Barrier Heights for Proton Exchange in Small Water, Ammonia, and Hydrogen Fluoride Clusters with G4 (MP2)-Type, MP *n*, and SCS-MP*n* Procedures—A Caveat. *J. Chem. Theory Comput.* **2012**, *8*, 3128–3136.
- [42] Grambow, C. A.; Pattanaik, L.; Green, W. H. Reactants, products, and transition states of elementary chemical reactions based on quantum chemistry. *Sci. Data* **2020**, *7*, 1–8.
- [43] von Rudorff, G. F.; Heinen, S. N.; Bragato, M.; von Lilienfeld, O. A. Thousands of reactants and transition states for competing E2 and S2 reactions. *Mach. Learn. Sci. Technol.* **2020**, *1*, 045026.
- [44] Kromann, J. C.; Christensen, A. S.; Cui, Q.; Jensen, J. H. Towards a barrier height benchmark set for biologically relevant systems. *PeerJ* **2016**, *4*, e1994.
- [45] Riplinger, C.; Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *J. Chem. Phys.* **2013**, *138*, 034106.
- [46] Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *J. Chem. Phys.* **2013**, *139*, 134101.
- [47] Liakos, D. G.; Sparta, M.; Kesharwani, M. K.; Martin, J. M.; Neese, F. Exploring the accuracy limits of local pair natural orbital coupled-cluster theory. *J. Chem. Theory Comput.* **2015**, *11*, 1525–1539.
- [48] Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory. *J. Chem. Phys.* **2016**, *144*, 024109.
- [49] Saitow, M.; Becker, U.; Riplinger, C.; Valeev, E. F.; Neese, F. 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.* **2017**, *146*, 164105.
- [50] Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D. G.; Minenkov, Y.; Cavallo, L.; Neese, F. Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [DLPNO-CCSD (T)]. *J. Chem. Phys.* **2018**, *148*, 011101.
- [51] Paiva, P.; Ramos, M. J.; Fernandes, P. A. Assessing the validity of DLPNO-CCSD(T) in the calculation of activation and reaction energies of ubiquitous enzymatic reactions. *J. Comput. Chem.* **2020**, *41*, 2459–2468.
- [52] 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.
- [53] 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.
- [54] Otero-de-la-Roza, A.; DiLabio, G. A. Transferable atom-centered potentials for the correction of basis set incompleteness errors in density-functional theory. *J. Chem. Theory Comput.* **2017**, *13*, 3505–3524.
- [55] Otero-de-la-Roza, A.; DiLabio, G. A. Improved basis-set incompleteness potentials for accurate dft calculations in large systems. *J. Chem. Theory Comput.* **2020**, *16*, 4176–4191.
- [56] Prasad, V. K.; Otero-de-la-Roza, A.; DiLabio, G. A. Atom-centered potentials with dispersion-corrected minimal basis set Hartree-Fock: an efficient and accurate computational approach for large molecular systems. *J. Chem. Theory Comput.* **2018**, *14*, 726–738.
- [57] Prasad, V. K.; Otero-de La-Roza, A.; DiLabio, G. A. PEPCONF, a diverse data set of peptide conformational energies. *Sci. Data* **2019**, *6*, 1–9.

- [58] Prasad, V. K.; Khalilian, M. H.; Otero-de-la-Roza, A.; DiLabio, G. A. BSE49, a diverse, high-quality benchmark dataset of separation energies of chemical bonds. *Sci. Data* **2021**, (submitted).
- [59] Ribeiro, A. J. M.; Holliday, G. L.; Furnham, N.; Tyzack, J. D.; Ferris, K.; Thornton, J. M. Mechanism and Catalytic Site Atlas (M-CSA): a database of enzyme reaction mechanisms and active sites. *Nucleic Acids Res.* **2018**, *46*, D618–D623.
- [60] Newcomb, M. In *Encyclopedia of Radicals in Chemistry, Biology and Materials*; Chatgililoglu, C., Studer, A., Eds.; Wiley Online Library, 2012.
- [61] Mackie, I. D.; DiLabio, G. A. Ring-opening radical clock reactions: many density functionals have difficulty keeping time. *Org. Biomol. Chem.* **2011**, *9*, 3158–3164.
- [62] Lucas, M. d. F.; Ramos, M. J. Theoretical study of the suicide inhibition mechanism of the enzyme pyruvate formate lyase by methacrylate. *J. Am. Chem. Soc.* **2005**, *127*, 6902–6909.
- [63] Zou, Y.; Xue, X.-S.; Deng, Y.; Smith III, A. B.; Houk, K. Factors Controlling Reactivity in the Hydrogen Atom Transfer and Radical Addition Steps of a Radical Relay Cascade. *Org. Lett.* **2019**, *21*, 5894–5897.
- [64] Romero-Silva, A.; Mora-Diez, N.; Alvarez-Idaboy, J. R. Theoretical study of the reactivity and selectivity of various free radicals with cysteine residues. *ACS Omega* **2018**, *3*, 16519–16528.
- [65] Hayden, A. E.; Paton, R. S.; Becker, J.; Lim, Y. H.; Nicolaou, K.; Houk, K. Origins of regioselectivity of Diels–Alder reactions for the synthesis of bisanthraquinone antibiotic BE-43472B. *J. Org. Chem.* **2010**, *75*, 922–928.
- [66] Lan, Y.; Danheiser, R. L.; Houk, K. Why Nature Eschews the Concerted [2+ 2+ 2] Cycloaddition of a Nonconjugated Cyanodiyne. Computational Study of a Pyridine Synthesis Involving an Ene–Diels–Alder–Bimolecular Hydrogen-Transfer Mechanism. *J. Org. Chem.* **2012**, *77*, 1533–1538.
- [67] Zheng, Y.; Thiel, W. Computational insights into an enzyme-catalyzed [4+ 2] cycloaddition. *J. Org. Chem.* **2017**, *82*, 13563–13571.
- [68] Sato, M.; Yagishita, F.; Mino, T.; Uchiyama, N.; Patel, A.; Chooi, Y.-H.; Goda, Y.; Xu, W.; Noguchi, H.; Yamamoto, T., et al. Involvement of lipocalin-like CghA in decalin-forming stereoselective intramolecular [4+ 2] cycloaddition. *ChemBioChem* **2015**, *16*, 2294.
- [69] Zhang, B.; Wang, K. B.; Wang, W.; Wang, X.; Liu, F.; Zhu, J.; Shi, J.; Li, L. Y.; Han, H.; Xu, K., et al. Enzyme-catalysed [6+ 4] cycloadditions in the biosynthesis of natural products. *Nature* **2019**, *568*, 122–126.
- [70] Maiga-Wandiam, B.; Corbu, A.; Massiot, G.; Sautel, F.; Yu, P.; Lin, B. W.-Y.; Houk, K. N.; Cossy, J. Intramolecular Diels–Alder Approaches to the Decalin Core of Verongidolide: The Origin of the exo-Selectivity, a DFT Analysis. *J. Org. Chem.* **2018**, *83*, 5975–5985.
- [71] Byrne, M. J.; Lees, N. R.; Han, L.-C.; van der Kamp, M. W.; Mulholland, A. J.; Stach, J. E.; Willis, C. L.; Race, P. R. The catalytic mechanism of a natural Diels–Alderase revealed in molecular detail. *J. Am. Chem. Soc.* **2016**, *138*, 6095–6098.
- [72] He, C. Q.; Chen, T. Q.; Patel, A.; Karabiyikoglu, S.; Merlic, C. A.; Houk, K. Distortion, tether, and entropy effects on transannular Diels–Alder cycloaddition reactions of 10–18-membered rings. *J. Org. Chem.* **2015**, *80*, 11039–11047.
- [73] Yu, P.; Patel, A.; Houk, K. Transannular [6+ 4] and ambimodal cycloaddition in the biosynthesis of heronamide A. *J. Am. Chem. Soc.* **2015**, *137*, 13518–13523.
- [74] Pham, H. V.; Paton, R. S.; Ross, A. G.; Danishefsky, S. J.; Houk, K. Intramolecular Diels–Alder reactions of cycloalkenones: stereoselectivity, Lewis acid acceleration, and halogen substituent effects. *J. Am. Chem. Soc.* **2014**, *136*, 2397–2403.
- [75] Duan, A.; Yu, P.; Liu, F.; Qiu, H.; Gu, F. L.; Doyle, M. P.; Houk, K. Diazo esters as dienophiles in intramolecular (4+ 2) cycloadditions: Computational explorations of mechanism. *J. Am. Chem. Soc.* **2017**, *139*, 2766–2770.
- [76] Fell, J. S.; Lopez, S. A.; Higginson, C. J.; Finn, M.; Houk, K. Theoretical Analysis of the Retro-Diels–Alder Reactivity of Oxanorbornadiene Thiol and Amine Adducts. *Org. Lett.* **2017**, *19*, 4504–4507.
- [77] Levandowski, B. J.; Herath, D.; Gallup, N. M.; Houk, K. Origin of π -Facial Stereoselectivity in Thiophene 1-Oxide Cycloadditions. *J. Org. Chem.* **2018**, *83*, 2611–2616.
- [78] Scholl, K.; Dillashaw, J.; Timpy, E.; Lam, Y.-h.; DeRatt, L.; Benton, T. R.; Powell, J. P.; Houk, K.; Morgan, J. B. Quinine-Promoted, Enantioselective Boron-Tethered Diels–Alder Reaction by Anomeric Control of Transition-State Conformation. *J. Org. Chem.* **2018**, *83*, 5756–5765.
- [79] Suh, S.-E.; Chen, S.; Houk, K.; Chenoweth, D. M. The mechanism of the triple aryne–tetrazine reaction cascade: theory and experiment. *Chem. Sci.* **2018**, *9*, 7688–7693.
- [80] Tan, D.; Jamieson, C. S.; Ohashi, M.; Tang, M.-C.; Houk, K.; Tang, Y. Genome-mined Diels–Alderase catalyzes formation of the cis-octahydrodecalins of varicidin A and B. *J. Am. Chem. Soc.* **2019**, *141*, 769–773.
- [81] Schmidt, Y.; Lam, J. K.; Pham, H. V.; Houk, K.; Vanderwal, C. D. Studies on the Himbert intramolecular Arene/Allene Diels–Alder cycloaddition. Mechanistic studies and expansion of scope to all-carbon tethers. *J. Am. Chem. Soc.* **2013**, *135*, 7339–7348.
- [82] Liang, Y.; Mackey, J. L.; Lopez, S. A.; Liu, F.; Houk, K. Control and design of mutual orthogonality in bioorthogonal cycloadditions. *J. Am. Chem. Soc.* **2012**, *134*, 17904–17907.
- [83] Liu, S.; Lei, Y.; Qi, X.; Lan, Y. Reactivity for the Diels–Alder reaction of cumulenes: a distortion–interaction analysis along the reaction pathway. *J. Phys. Chem. A* **2014**, *118*, 2638–2645.
- [84] Ma, Z.-X.; Patel, A.; Houk, K.; Hsung, R. P. Highly Torquoselective Electrocyclizations and Competing 1, 7-Hydrogen Shifts of 1-Azatrienenes with Silyl Substitution at the Allylic Carbon. *Org. Lett.* **2015**, *17*, 2138–2141.
- [85] Ohashi, M.; Liu, F.; Hai, Y.; Chen, M.; Tang, M.-c.; Yang, Z.; Sato, M.; Watanabe, K.; Houk, K.; Tang, Y. SAM-dependent enzyme-catalysed pericyclic reactions in natural product biosynthesis. *Nature* **2017**, *549*, 502–506.
- [86] Yang, Z.; Dong, X.; Yu, Y.; Yu, P.; Li, Y.; Jamieson, C.; Houk, K. Relationships between product ratios in ambimodal pericyclic reactions and bond lengths in transition structures. *J. Am. Chem. Soc.* **2018**, *140*, 3061–3067.

- [87] Xue, X.-S.; Jamieson, C. S.; Garcia-Borràs, M.; Dong, X.; Yang, Z.; Houk, K. Ambimodal trisubstituted transition state and dynamic control of periselectivity. *J. Am. Chem. Soc.* **2019**, *141*, 1217–1221.
- [88] Wiest, O.; Houk, K. Stabilization of the transition state of the chorismate-prephenate rearrangement: An ab initio study of enzyme and antibody catalysis. *J. Am. Chem. Soc.* **1995**, *117*, 11628–11639.
- [89] Scott, S. K.; Sanders, J. N.; White, K. E.; Yu, R. A.; Houk, K.; Grenning, A. J. Controlling, understanding, and redirecting the thermal rearrangement of 3, 3-dicyano-1, 5-enynes. *J. Am. Chem. Soc.* **2018**, *140*, 16134–16139.
- [90] Boon, B. A.; Green, A. G.; Liu, P.; Houk, K.; Merlic, C. A. Using ring strain to control 4 π -electrocyclization reactions: torquoselectivity in ring closing of medium-ring dienes and ring opening of bicyclic cyclobutenes. *J. Org. Chem.* **2017**, *82*, 4613–4624.
- [91] Patel, A.; Barcan, G. A.; Kwon, O.; Houk, K. Origins of 1, 6-Stereinduction in Torquoselective 6 π Electrocyclizations. *J. Am. Chem. Soc.* **2013**, *135*, 4878–4883.
- [92] Sader, C. A.; Houk, K. A Theoretical Study of Cyclohexyne Addition to Carbonyl-C α Bonds: Allowed and Forbidden Electrocyclic and Nonpericyclic Ring-Openings of Strained Cyclobutenes. *J. Org. Chem.* **2012**, *77*, 4939–4948.
- [93] Hong, X.; Liang, Y.; Griffith, A. K.; Lambert, T. H.; Houk, K. Distortion-accelerated cycloadditions and strain-release-promoted cycloreversions in the organocatalytic carbonyl-olefin metathesis. *Chem. Sci.* **2014**, *5*, 471–475.
- [94] Xie, S.; Lopez, S. A.; Ramström, O.; Yan, M.; Houk, K. 1, 3-Dipolar cycloaddition reactivities of perfluorinated aryl azides with enamines and strained dipolarophiles. *J. Am. Chem. Soc.* **2015**, *137*, 2958–2966.
- [95] Krenke, E. H.; Patel, A.; Houk, K. Does nature click? Theoretical prediction of an enzyme-catalyzed transannular 1, 3-dipolar cycloaddition in the biosynthesis of lycojaponicumins A and B. *J. Am. Chem. Soc.* **2013**, *135*, 17638–17642.
- [96] Çelebi-Ölçüm, N.; Lam, Y.-h.; Richmond, E.; Ling, K. B.; Smith, A. D.; Houk, K. N. Pericyclic Cascade with Chirality Transfer: Reaction Pathway and Origin of Enantioselectivity of the Hetero-Claisen Approach to Oxindoles. *Angew. Chem. Intl. Ed.* **2011**, *50*, 11478–11482.
- [97] Hamlin, T. A.; Kelly, C. B.; Oviatt, J. M.; Wiles, R. J.; Tilley, L. J.; Leadbeater, N. E. Toward a unified mechanism for oxoammonium salt-mediated oxidation reactions: a theoretical and experimental study using a hydride transfer model. *J. Org. Chem.* **2015**, *80*, 8150–8167.
- [98] Vitkovskaya, N. M.; Kobayev, V. B.; Bobkov, A. S.; Orel, V. B.; Schmidt, E. Y.; Trofimov, B. A. Nucleophilic Addition of Ketones To Acetylenes and Allenes: A Quantum-Chemical Insight. *J. Org. Chem.* **2017**, *82*, 12467–12476.
- [99] Kister, J.; Ess, D. H.; Roush, W. R. Enantio- and Diastereoselective Synthesis of syn- β -Hydroxy- α -vinyl Carboxylic Esters via Reductive Aldol Reactions of Ethyl Allenecarboxylate with 10-TMS-9-Borabicyclo [3.3. 2] decane and DFT Analysis of the Hydroboration Pathway. *Org. Lett.* **2013**, *15*, 5436–5439.
- [100] Johnson, E. R.; Clarkin, O. J.; Dale, S. G.; DiLabio, G. A. Kinetics of the addition of olefins to Si-centered radicals: the critical role of dispersion interactions revealed by theory and experiment. *J. Phys. Chem. A* **2015**, *119*, 5883–5888.
- [101] Zhang, Z.; Collum, D. B. Wittig Rearrangements of Boron-Based Oxazolidinone Enolates. *J. Org. Chem.* **2019**, *84*, 10892–10900.
- [102] Liu, W.-B.; Schuman, D. P.; Yang, Y.-F.; Toutov, A. A.; Liang, Y.; Klare, H. F.; Nesnas, N.; Oestreich, M.; Blackmond, D. G.; Virgil, S. C., et al. Potassium tert-butoxide-catalyzed dehydrogenative C–H silylation of heteroaromatics: a combined experimental and computational mechanistic study. *J. Am. Chem. Soc.* **2017**, *139*, 6867–6879.
- [103] D'Alfonso, C.; Bietti, M.; DiLabio, G. A.; Lanzalunga, O.; Salamone, M. Reactions of the Phthalimide N-Oxyl Radical (PINO) with Activated Phenols: The Contribution of π -Stacking Interactions to Hydrogen Atom Transfer Rates. *J. Org. Chem.* **2013**, *78*, 1026–1037.
- [104] DiLabio, G. A.; Franchi, P.; Lanzalunga, O.; Lapi, A.; Lucarini, F.; Lucarini, M.; Mazzonna, M.; Prasad, V. K.; Ticconi, B. Hydrogen atom transfer (HAT) processes promoted by the quinolinimide-N-oxyl radical. A kinetic and theoretical study. *J. Org. Chem.* **2017**, *82*, 6133–6141.
- [105] Frisch, M. J. et al. Gaussian 09 Revision A.1. Gaussian Inc. Wallingford CT 2009.
- [106] Frisch, M. J. et al. Gaussian 16 Revision A.03. 2016; Gaussian Inc. Wallingford CT.
- [107] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- [108] 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.
- [109] Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51–57.
- [110] 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.
- [111] Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.
- [112] Johnson, E. R.; Becke, A. D. A post-Hartree-Fock model of intermolecular interactions: Inclusion of higher-order corrections. *J. Chem. Phys.* **2006**, *124*, 174104.
- [113] Francel, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, *77*, 3654–3665.
- [114] Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* **1973**, *28*, 213–222.
- [115] Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-consistent molecular orbital methods. XII. Further extensions of Gaussian—type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.* **1972**, *56*,

- 2257–2261.
- [116] Mohamadi, F.; Richards, N. G.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. MacroModel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.* **1990**, *11*, 440–467.
 - [117] Schrödinger Release 2020-3: MacroModel, Schrödinger, LLC. 2020; New York, NY.
 - [118] Schrödinger Release 2020-3: Maestro, Schrödinger, LLC. 2020; New York, NY.
 - [119] Medvedev, M. G.; Zeifman, A. A.; Novikov, F. N.; Bushmarinov, I. S.; Stroganov, O. V.; Titov, I. Y.; Chilov, G. G.; Svitanko, I. V. Quantifying possible routes for SpnF-catalyzed formal Diels–Alder cycloaddition. *J. Am. Chem. Soc.* **2017**, *139*, 3942–3945.
 - [120] Medvedev, M. G.; Panova, M. V.; Chilov, G. G.; Bushmarinov, I. S.; Novikov, F. N.; Stroganov, O. V.; Zeifman, A. A.; Svitanko, I. V. Exhaustive conformational search for transition states: the case of catechol O-methyltransferase active site. *Mendeleev Commun.* **2017**, *27*, 224–227.
 - [121] Fukaya, K.; Saito, A.; Nakajima, N.; Urabe, D. A computational study on the stereo- and regioselective formation of the C4 α –C6' bond of tethered catechin moieties by an exhaustive search of the transition states. *J. Org. Chem.* **2019**, *84*, 2840–2849.
 - [122] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
 - [123] Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **1980**, *72*, 650–654.
 - [124] McLean, A.; Chandler, G. Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z=11–18. *J. Chem. Phys.* **1980**, *72*, 5639–5648.
 - [125] Chang, G.; Guida, W. C.; Still, W. C. An internal-coordinate Monte Carlo method for searching conformational space. *J. Am. Chem. Soc.* **1989**, *111*, 4379–4386.
 - [126] <https://github.com/bobbypaton/FullMonte>.
 - [127] Kim, S.; Chmely, S. C.; Nimlos, M. R.; Bomble, Y. J.; Foust, T. D.; Paton, R. S.; Beckham, G. T. Computational study of bond dissociation enthalpies for a large range of native and modified lignins. *J. Phys. Chem. Lett.* **2011**, *2*, 2846–2852.
 - [128] Korth, M.; Pitonak, M.; Rezac, J.; Hobza, P. A transferable H-bonding correction for semiempirical quantum-chemical methods. *J. Chem. Theory Comput.* **2010**, *6*, 344–352.
 - [129] Stewart, J. J. P. MOPAC2016 (<http://OpenMOPAC.Net>). 2016; Colorado Springs, CO, USA.
 - [130] Neese, F. The ORCA program system. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 73–78.
 - [131] Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, e1327.
 - [132] Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
 - [133] Kendall, R. A.; Dunning Jr, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
 - [134] Woon, D. E.; Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358–1371.
 - [135] Feyereisen, M.; Fitzgerald, G.; Komornicki, A. Use of approximate integrals in ab initio theory. An application in MP2 energy calculations. *Chem. Phys. Lett.* **1993**, *208*, 359–363.
 - [136] Bernholdt, D. E.; Harrison, R. J. Large-scale correlated electronic structure calculations: the RI-MP2 method on parallel computers. *Chem. Phys. Lett.* **1996**, *250*, 477–484.
 - [137] Weigend, F.; Häser, M. RI-MP2: first derivatives and global consistency. *Theor. Chem. Acc.* **1997**, *97*, 331–340.
 - [138] Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002**, *116*, 3175–3183.
 - [139] Vydrov, O. A.; Scuseria, G. E. Assessment of a long-range corrected hybrid functional. *J. Chem. Phys.* **2006**, *125*, 234109.
 - [140] Vydrov, O. A.; Heyd, J.; Krukau, A. V.; Scuseria, G. E. Importance of short-range versus long-range Hartree–Fock exchange for the performance of hybrid density functionals. *J. Chem. Phys.* **2006**, *125*, 074106.
 - [141] Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. *J. Chem. Theory Comput.* **2006**, *2*, 364–382.
 - [142] Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
 - [143] Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Constantin, L. A.; Sun, J. Workhorse semilocal density functional for condensed matter physics and quantum chemistry. *Phys. Rev. Lett.* **2009**, *103*, 026403.
 - [144] Chai, J.-D.; Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **2008**, *128*, 084106.
 - [145] Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.
 - [146] Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098.
 - [147] Perdew, J.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
 - [148] 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.
 - [149] Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372.

- [150] Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.
- [151] Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. A ‘chain-of-spheres’ algorithm for the Hartree–Fock exchange. *Chem. Phys.* **2009**, *356*, 98–109.
- [152] 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.
- [153] Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.
- [154] Goerigk, L.; Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670–6688.
- [155] Becke, A. D.; Johnson, E. R. Exchange-Hole Dipole Moment and the Dispersion Interaction Revisited. *J. Chem. Phys.* **2007**, *127*, 154108.
- [156] Otero-de-la-Roza, A.; Johnson, E. R. Non-Covalent Interactions and Thermochemistry using XDM-Corrected Hybrid and Range-Separated Hybrid Density Functionals. *J. Chem. Phys.* **2013**, *138*, 204109.
- [157] The postg program is freely available from <https://github.com/aoterodelaroza/postg>.
- [158] The canonical XDM coefficients can be found at <http://schooner.chem.dal.ca> and also in the xdm.param file of the postg distribution.
- [159] Jensen, F. Polarization consistent basis sets: Principles. *J. Chem. Phys.* **2001**, *115*, 9113–9125.
- [160] Jensen, F. Polarization consistent basis sets. II. Estimating the Kohn–Sham basis set limit. *J. Chem. Phys.* **2002**, *116*, 7372–7379.
- [161] Jensen, F. Polarization consistent basis sets. III. The importance of diffuse functions. *J. Chem. Phys.* **2002**, *117*, 9234–9240.
- [162] Jensen, F.; Helgaker, T. Polarization consistent basis sets. V. The elements Si–Cl. *J. Chem. Phys.* **2004**, *121*, 3463–3470.
- [163] Karton, A. Highly accurate CCSDT(Q)/CBS reaction barrier heights for a diverse set of transition structures: basis set convergence and cost-effective approaches for estimating post-CCSD(T) contributions. *J. Phys. Chem. A* **2019**, *123*, 6720–6732.
- [164] East, A. L.; Allen, W. D. The heat of formation of NCO. *J. Chem. Phys.* **1993**, *99*, 4638–4650.
- [165] East, A. L.; Johnson, C. S.; Allen, W. D. Characterization of the \tilde{X}^1A' state of isocyanic acid. *J. Chem. Phys.* **1993**, *98*, 1299–1328.
- [166] Papajak, E.; Truhlar, D. G. What are the most efficient basis set strategies for correlated wave function calculations of reaction energies and barrier heights? *J. Chem. Phys.* **2012**, *137*, 064110.
- [167] Warden, C. E.; Smith, D. G.; Burns, L. A.; Bozkaya, U.; Sherrill, C. D. Efficient and automated computation of accurate molecular geometries using focal-point approximations to large-basis coupled-cluster theory. *J. Chem. Phys.* **2020**, *152*, 124109.
- [168] Neese, F.; Valeev, E. F. Revisiting the atomic natural orbital approach for basis sets: Robust systematic basis sets for explicitly correlated and conventional correlated ab initio methods? *J. Chem. Theory Comput.* **2011**, *7*, 33–43.
- [169] Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set convergence of correlated calculations on water. *J. Chem. Phys.* **1997**, *106*, 9639–9646.
- [170] Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N₂, and H₂O. *Chem. Phys. Lett.* **1998**, *286*, 243–252.
- [171] Truhlar, D. G. Basis-set extrapolation. *Chem. Phys. Lett.* **1998**, *294*, 45–48.
- [172] Burns, L. A.; Marshall, M. S.; Sherrill, C. D. Appointing silver and bronze standards for noncovalent interactions: A comparison of spin-component-scaled (SCS), explicitly correlated (F12), and specialized wavefunction approaches. *J. Chem. Phys.* **2014**, *141*, 234111.
- [173] 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.
- [174] Rezac, J.; Hobza, P. Describing noncovalent interactions beyond the common approximations: How accurate is the “gold standard,” CCSD (T) at the complete basis set limit? *J. Chem. Theory Comput.* **2013**, *9*, 2151–2155.
- [175] Karton, A.; Martin, J. M. Comment on: Estimating the Hartree–Fock limit from finite basis set calculations [Jensen F (2005) *Theor Chem Acc* 113: 267]. *Theor. Chem. Acc.* **2006**, *115*, 330–333.
- [176] Jensen, F. Estimating the Hartree–Fock limit from finite basis set calculations. *Theor. Chem. Acc.* **2005**, *113*, 267–273.
- [177] Mardirossian, N.; Head-Gordon, M. How accurate are the Minnesota density functionals for noncovalent interactions, isomerization energies, thermochemistry, and barrier heights involving molecules composed of main-group elements? *J. Chem. Theory Comput.* **2016**, *12*, 4303–4325.
- [178] Goerigk, L.; Kruse, H.; Grimme, S. Benchmarking density functional methods against the S66 and S66x8 datasets for non-covalent interactions. *Chem. Phys. Chem* **2011**, *12*, 3421–3433.
- [179] Grimme, S.; Huenerbein, R.; Ehrlich, S. On the importance of the dispersion energy for the thermodynamic stability of molecules. *Chem. Phys. Chem* **2011**, *12*, 1258–1261.
- [180] Grimme, S. Seemingly simple stereoelectronic effects in alkane isomers and the implications for Kohn–Sham density functional theory. *Angew. Chem. Intl. Ed.* **2006**, *45*, 4460–4464.
- [181] Goerigk, L. How do DFT-DCP, DFT-NL, and DFT-D3 compare for the description of London-dispersion effects in conformers and general thermochemistry? *J. Chem. Theory Comput.* **2014**, *10*, 968–980.
- [182] Holmes, J. D.; Otero-de-la-Roza, A.; DiLabio, G. A. Accurate modeling of water clusters with density-functional theory using atom-centered potentials. *J. Chem. Theory Comput.* **2017**, *13*, 4205–4215.