

PCCCP

Physical Chemistry Chemical Physics

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: O. Koeksal, *Phys. Chem. Chem. Phys.*, 2026, DOI: 10.1039/D6CP00985A.



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

Δ -Machine Learning toward CCSD Accuracy for Homohalogenated Borane–Phosphine Adducts: Screening Low-Energy Structures from DFT and MP2 Libraries

Okan Köksal^{1,*}

¹Center for Molecular Modeling Ghent University,
Technologiepark-Zwijnaarde 46, BE-9052 Gent, Belgium

(Dated: April 14, 2026)

Accurate formation energies for weak boron–phosphorus Lewis adducts are challenging because low-order correlation and density-functional methods can mis-rank low-energy motifs on shallow potential-energy landscapes and are sensitive to basis-set superposition error (BSSE). Herein, we combine large-scale structure-library sampling with Δ -machine learning (Δ -ML) to approach coupled-cluster accuracy for the homohalogenated adducts F_3B-PF_3 , Cl_3B-PCl_3 , and Br_3B-PBr_3 . DFT (B3LYP-D3) and MP2 libraries comprising several thousand geometries per system are generated and used to train CCSD-referenced Δ -ML models that predict E_{CCSD} from low-level inputs. The resulting models reproduce CCSD energies with low errors and enable efficient screening of the full libraries, after which a compact low-energy subset is refined with targeted CCSD(T) calculations. Counterpoise-corrected results show that MP2 substantially overbinds, especially for the chlorinated and brominated adducts. At the highest level, Cl_3B-PCl_3 is found to lie at the threshold of binding, whereas Br_3B-PBr_3 remains clearly bound and F_3B-PF_3 is weakly bound. Distance-resolved scans and Morokuma-type energy decomposition analyses rationalize the distinct binding regimes across F/Cl/Br in terms of the balance between Pauli repulsion, polarization/exchange, and dispersion. The proposed workflow enables reliable coupled-cluster-level screening of weak donor–acceptor adducts at greatly reduced cost.

I. INTRODUCTION

Lewis acid–base adducts featuring B–P donor–acceptor bonds remain central to main-group chemistry, with relevance spanning classical synthesis and reactive intermediates to applications in molecular materials and hydrogen-storage concepts [1–4]. Although the textbook picture of a phosphine→borane donation into the vacant boron *p* orbital is conceptually straightforward, the nature of the B–P interaction can vary continuously from weak, predominantly electrostatic association to substantially covalent bonding. This variability depends sensitively on fragment electronic structure, substituent effects, and the deformation required to bring the Lewis acid into a binding-ready (pyramidalized) geometry [5–7]. In haloborane adducts, the stability of the B–P contact reflects a delicate interplay between donor→acceptor charge transfer, the pyramidalization penalty of trigonal-planar BX_3 , and halogen-dependent polarization, back-donation, and dispersion effects [6, 7].

A recurring theme in theoretical work on B–P donor–acceptor complexes is that computed geometries and formation energies can be strongly method dependent, especially for shallow potentials where multiple low-lying minima or near-degenerate motifs may exist [8–19]. This motivates a PES-focused perspective and benchmarking against correlated wave-function methods rather than reliance on a single optimized structure. A striking illustration of the underlying complexity was highlighted by

Almas and Pearson [20], who demonstrated that the B–P bond-stretching potential energy surface (PES) in phosphine haloboranes can deviate strongly from a conventional Morse-like form and may exhibit anomalous inflection points. In the extreme case of Cl_3B-PH_3 , the PES even shows two distinct minima [20]. Such findings emphasize that reliable ranking and mechanistic interpretation of weak-to-intermediate adducts require accurate characterization of the local PES landscape in regions where shallow basins, shoulders, or competing minima may occur. In this context, high-quality experimental benchmarks for closely related B–P adducts remain valuable reference points for assessing computed structures and trends [21]. Recently, we presented a systematic *ab initio* analysis of donor–acceptor bonding in a broad set of $X_3B \cdots PY_3$ and $X_3B \cdots AsY_3$ complexes ($X = H, F, Cl, Br$; $Y = CH_3, F, Cl, Br$), combining MP2 energies with coupled-cluster benchmarking and complementary bonding analyses (EDA, RDG, NBO) [22]. That work established clear trends from predominantly electrostatic interactions in weakly bound adducts to strongly bound, more covalent systems, and underscored the stabilizing role of halogen back-donation and charge polarization in BCl_3 and BBr_3 adducts relative to BF_3 complexes [22]. Importantly, the three symmetric halogenated phosphane–haloborane pairs F_3B-PF_3 , Cl_3B-PCl_3 , and Br_3B-PBr_3 were already included in that dataset and served as representative examples of weak-to-intermediate B–P triel bonding motifs, and all three were verified as genuine minima at the MP2/aug-cc-pVTZ level [23], exhibiting no imaginary frequencies in harmonic vibrational analyses [22].

In the present contribution, we revisit precisely these three prototypical systems: F_3B-PF_3 , Cl_3B-PCl_3 , and

* Corresponding author.

E-mail address: okan.koksal@Ugent.be



$\text{Br}_3\text{B}-\text{PBr}_3$, but with a different emphasis and a stricter accuracy target. Rather than focusing on bonding decomposition at a single minimum, we (i) map the local structural landscape in a statistically meaningful manner, (ii) assess how method choice (DFT vs. MP2) affects the identification and ranking of low-energy structures on a potentially shallow, non-Morse-like PES [20], and (iii) connect these structural ensembles to high-accuracy reference energetics at the CCSD(T) level [24, 25]. To enable this, we generated extensive structure libraries for each complex and evaluated them at both levels of theory, comprising 5497 (MP2) and 5493 (DFT) geometries for $\text{F}_3\text{B}-\text{PF}_3$, 5557 (MP2) and 5556 (DFT) geometries for $\text{Cl}_3\text{B}-\text{PCl}_3$, and 5714 (MP2) and 5713 (DFT) geometries for $\text{Br}_3\text{B}-\text{PBr}_3$. This dense sampling allows identification of competing minima and near-degenerate motifs, quantification of structural variability close to the global minimum, and a controlled comparison of DFT and MP2 predictions for weak B–P donor–acceptor contacts. To connect the observed structural diversity with a physically transparent picture of the B–P interaction, we further analyze selected B–P separations using energy decomposition concepts [26–29].

More broadly, machine learned representations of potential-energy surfaces have advanced substantially over the past decade, with widely used approaches including permutationally invariant polynomials, neural networks, and Gaussian-process-based models [30–34]. In the present work, however, we do not attempt a full global PES fit. Instead, we employ a descriptor-based Δ -learning strategy for local low-energy screening, using handcrafted geometric/energetic descriptors and cross-checking the conclusions with rotationally and translationally invariant SOAP representations (see the Supplementary Information). Related orbital-based machine-learning strategies have also been developed, most notably molecular-orbital-based machine learning (MOBML), in which post-Hartree–Fock correlation energies are learned from Hartree–Fock molecular-orbital information rather than primarily from geometric descriptors [35, 36]. Such approaches aim at improved transferability across chemical systems by exploiting compact orbital-based representations. Extending such strategies to larger molecular systems while retaining high-level electronic-structure accuracy remains challenging. For shallow donor–acceptor potential-energy surfaces, reliable energetic ranking requires an accurate treatment of electron correlation. In this context, coupled-cluster methods provide particularly important reference data, and CCSD(T) is widely regarded as a benchmark or “gold-standard” approach for molecular energetics. However, the steep computational scaling of conventional CC methods, culminating in the formal $O(N^7)$ scaling of CCSD(T), makes direct application to large structure libraries impractical [37–39]. Indeed, even for systems of only moderate size, high-level PES construction can require very large coupled-cluster datasets, e.g., Bowman and co-workers reported a 10-atom formic-acid-

dimer PES based on 13 475 CCSD(T)-F12/triple- ζ energy points [40].

A practical strategy to overcome this bottleneck is Δ -machine learning, in which a model is trained to learn the difference between a low-cost baseline and a higher-level reference, thereby correcting large low-level datasets toward coupled-cluster accuracy [41]. In the context of potential-energy surfaces, this idea is commonly expressed as a correction to a low-level PES, $V_{\text{LL}\rightarrow\text{CC}} = V_{\text{LL}} + \Delta V_{\text{CC-LL}}$, where the smoother correction term can often be learned from a comparatively small number of high-level data points [42]. Such Δ -learning strategies have been successfully used to accelerate high-level energetics and PES construction for systems ranging from small molecules to medium-sized species [43, 44]. Motivated by these developments and by the non-trivial PES features reported for phosphine haloboranes [20], we employ a CCSD-trained Δ -learning strategy to screen and rank low-energy regions of the PES and subsequently refine a targeted subset of ML-selected structures at the CCSD(T) level, thereby approaching CCSD(T)-quality energetics without exhaustive CCSD(T) sampling.

The paper is organized as follows. Section II details the electronic-structure methods, the structure-library generation protocol, and the Δ -learning workflow used to approach CCSD accuracy and to guide the subsequent CCSD(T) refinement of selected low-energy candidates. Section III compares DFT and MP2 structural ensembles, identifies the dominant low-energy motifs, and analyzes how high-level corrections reshape the energetic ranking. Finally, Section IV summarizes the principal conclusions and outlines practical guidance for computational studies of shallow donor–acceptor PESs in halogenated phosphane–haloborane complexes.

II. METHODS

A. Structure-library generation

Starting from each equilibrium geometry, we generated randomly distorted structures by applying small, isotropic atomic displacements within a sphere of radius R centered at the original atomic position. For each atom, a displacement length r was sampled uniformly in volume by drawing $r' \in [0, 1]$ and setting:

$$r = R \sqrt[3]{r'}. \quad (1)$$

Two angles $\phi \in [0, 2\pi)$ and $\theta \in [0, \pi]$ were generated to define the displacement direction in spherical coordinates. The corresponding Cartesian components (d_x, d_y, d_z) were added to the original atomic coordinates to yield the perturbed geometry. Because these displacements were applied directly to the Cartesian coordinates of the atoms, not only the intermolecular B \cdots P separation but also the intrafragment B–X and P–X



bond lengths and associated angles could vary in the distorted structures. No subsequent geometry relaxation was performed for these library geometries and they were evaluated directly by single-point calculations. In the present work we employed $R = 0.001$ Å to generate near-equilibrium distortions and $R = 0.09$ Å to sample more strongly distorted configurations. Consequently, a larger number of structures was generated at larger R to ensure sufficient coverage of the higher-variance regions of the configuration space. As shown in Fig. 1, randomly distorted donor–acceptor geometries were generated by applying small stochastic Cartesian displacements to the atomic coordinates of the reference structure, where the displacement direction for each atom was sampled in spherical coordinates through the angles (θ, ϕ) and the displacement length was sampled within a sphere of radius R . The resulting structures therefore sample variations not only in the intermolecular $B \cdots P$ separation, but also in the associated intra- and intermolecular geometric descriptors. Subsequently, we train regression models to learn the correction $\Delta E = E_{\text{CCSD}} - E_{\text{low}}$. The trained model is then used to predict CCSD-quality energies for F_3B-PF_3 , Cl_3B-PCl_3 , and Br_3B-PBr_3 from MP2/DFT inputs across the sampled configurational space.

In this work we focus on the homohalogenated donor–acceptor adducts F_3B-PF_3 , Cl_3B-PCl_3 , and Br_3B-PBr_3 and combine large-scale structure sampling with correlated reference data and energy-decomposition analysis. All electronic-structure calculations were performed with GAMESS [45]. Equilibrium geometries were optimized at the MP2/aug-cc-pVTZ level and, independently, at the DFT (B3LYP-D3)/aug-cc-pVTZ level. The MP2 structure libraries were generated by applying random spherical displacements to the MP2-optimized geometries and evaluating MP2 single-point energies for each sampled structure. The same distorted geometries were then re-evaluated at the DFT (B3LYP-D3)/aug-cc-pVTZ level to construct the corresponding DFT energy libraries. Coupled-cluster reference data were obtained at the CCSD/aug-cc-pVTZ level for a labeled subset of structures and used to train Δ -learning models for the correction $\Delta E = E_{\text{CCSD}} - E_{\text{low}}$, where E_{low} denotes either MP2 [46, 47] or DFT (B3LYP-D3) [48, 49]. To assess the impact of perturbative triples in the low-energy region without exhaustive CCSD(T) sampling, targeted CCSD(T)/aug-cc-pVTZ single-point calculations were carried out for a compact ML-selected refinement set for each complex (including the baseline-optimized minimum and the lowest-energy CCSD training-set structure). Hence, these CCSD(T) calculations are best viewed as targeted single-point refinements of an ML-selected low-energy subset, not as an ML-driven search on a full CCSD(T) potential-energy surface. Unless noted otherwise, correlated calculations employed the frozen-core approximation, which reduces computational cost while retaining the dominant valence correlation contribution to bonding [50]. Basis-set superposition er-

ror (BSSE) was estimated using the Boys–Bernardi counterpoise (CP) procedure [51].

III. RESULTS AND DISCUSSION

A. Geometries and formation energies

Table I compares formation energies for the three complexes at MP2, DFT/B3LYP-D3, CCSD (Δ -ML), and CCSD(T) (subset) levels. In triel-bonded systems, the $X-B-P$ angle serves as a useful structural measure of the extent of pyramidalization at the Lewis-acid center, where stronger donor–acceptor interactions are commonly associated with a larger departure from trigonal-planar geometry and α values approaching 90° for the boron fragment [53, 54]. Furthermore, for F_3B-PF_3 we report a CCSD_{opt} geometry obtained with numerical gradients in GAMESS, yielding $R_{\text{ab}} = 3.26410$ Å and averaged angles $\langle \alpha \rangle = 90.641^\circ$ and $\langle \beta \rangle = 119.427^\circ$ (cf. Table I). Across the halogen series, MP2 predicts a strongly increasing stabilization (F: -2.05 , Cl: -9.30 , Br: -16.54 kcal mol $^{-1}$), whereas CCSD(T) (subset) yields a much more moderate trend (F: -2.19 , Cl: -3.82 , Br: -9.25 kcal mol $^{-1}$, cf. Table I). This indicates that MP2 systematically produces more negative ΔE_{form} values than CCSD(T), with the discrepancy growing substantially from F to Br. In contrast, DFT/B3LYP-D3 gives only slightly negative formation energies for F (-1.57 kcal mol $^{-1}$) and Br (-3.41 kcal mol $^{-1}$), but predicts a positive formation energy for the chlorine adduct ($+1.71$ kcal mol $^{-1}$, see Table I). In addition, BSSE corrections reduce the magnitude of the MP2 stabilization across the series (F: $-2.05 \rightarrow -1.17$, Cl: $-9.30 \rightarrow -5.42$, Br: $-16.54 \rightarrow -11.90$ kcal mol $^{-1}$), which is consistent with increased BSSE for the heavier, more polarizable complexes (cf. Table I). For DFT/B3LYP-D3, CP corrections are comparatively small for F and Br (F: $-1.57 \rightarrow -1.31$, Br: $-3.41 \rightarrow -2.87$ kcal mol $^{-1}$), but further increase the positive formation energy for Cl_3B-PCl_3 ($+1.71 \rightarrow +2.20$ kcal mol $^{-1}$). At the CCSD(Δ -ML) level, the BSSE correction is decisive for the chlorine adduct: the uncorrected formation energy remains slightly negative (-1.52 kcal mol $^{-1}$), whereas the CP-corrected value becomes positive ($+1.96$ kcal mol $^{-1}$). At the CCSD(T) level, the corresponding CP-corrected formation energy is essentially zero (-0.03 kcal mol $^{-1}$). Thus, Cl_3B-PCl_3 emerges as a borderline case whose energetic classification depends sensitively on the level of electron correlation, geometry relaxation, and BSSE correction. This conclusion is sharpened by the ML-assisted optimization, which enabled access to the relevant higher-level minimum and thereby revealed the near-threshold character of the chlorinated adduct much more clearly than would be possible from lower-level structures alone. By contrast, the brominated adduct remains clearly bound even after CP correction at both CCSD and CCSD(T), although the correction substan-



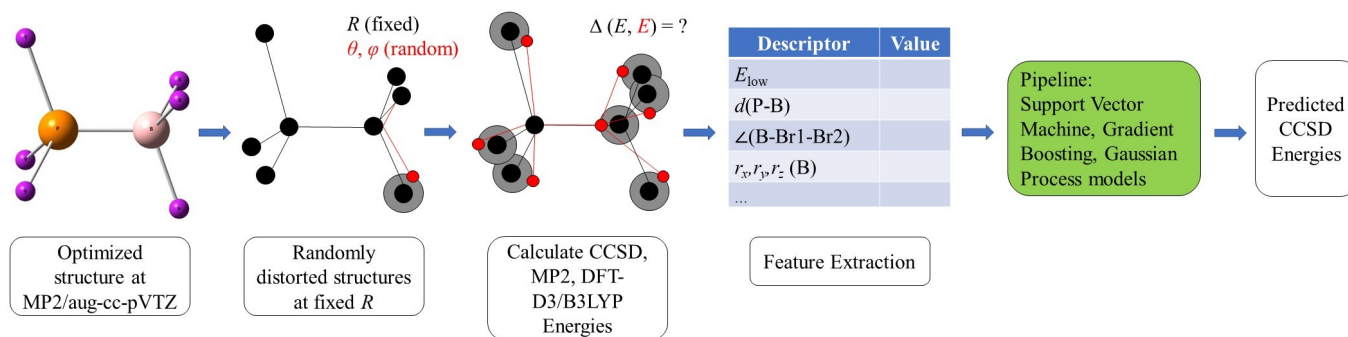


FIG. 1: Δ -machine-learning workflow for predicting CCSD-quality energies of the donor-acceptor complexes F_3B-PF_3 , Cl_3B-PCl_3 , and Br_3B-PBr_3 . Randomly distorted geometries are generated by applying stochastic atomic displacements, where the displacement direction is sampled through the angular coordinates (θ, ϕ) and the displacement length is sampled within a sphere of radius R . For each distorted structure, geometric descriptors and the corresponding low-level reference energy E_{low} are computed, where E_{low} denotes either E_{MP2} or E_{DFT} , depending on the baseline model. Regression models are then trained to learn the correction $\Delta E = E_{CCSD} - E_{low}$. The trained model is subsequently used to predict CCSD-quality energies for unseen structures from low-level inputs alone.

TABLE I: Key structural parameters and formation energies in units of kcal mol⁻¹ for F_3B-PF_3 , Cl_3B-PCl_3 , and Br_3B-PBr_3 at representative levels. R_{ab} denotes the B-P separation. The angles α ($X-B-P$) and β ($B-P-Y$) are averages over the three symmetry-related substituents. CCSD(T) values refer to the ML-guided refined subset (including the optimized geometry at MP2 level of theory and the lowest-CCSD training-set structure). The CCSD_{opt} geometry for F_3B-PF_3 was obtained with numerical gradients in GAMESS. The CP-corrected values are reported as ΔE_{form}^{CP} , where available.

Complex	Level	R_{ab} (Å)	$\langle\alpha\rangle$ (deg)	$\langle\beta\rangle$ (deg)	ΔE_{form}	ΔE_{form}^{CP}
F_3B-PF_3	MP2	3.23715	90.733	119.279	-2.05	-1.17
	DFT/B3LYP-D3	3.39430	90.394	119.192	-1.57	-1.31
	CCSD _{opt}	3.26410	90.641	119.427	-1.91	— ^a
	CCSD (Δ -ML)	3.24436	90.718	119.283	-1.88	— ^a
	CCSD(T) (subset)	3.23715	90.733	119.279	-2.19	— ^a
Cl_3B-PCl_3	MP2	2.01380	102.880	113.441	-9.30	-5.42
	DFT/B3LYP-D3	2.10874	102.351	113.617	+1.71	+2.20
	CCSD (Δ -ML)	2.01383	102.874	113.450	-1.52	+1.96
	CCSD(T) (subset)	2.01381	102.873	113.450	-3.82	-0.03
Br_3B-PBr_3	MP2	1.99482	103.204	112.998	-16.54	-11.90
	DFT/B3LYP-D3	2.07699	103.123	113.015	-3.41	-2.87
	CCSD (Δ -ML)	1.99514	103.213	113.001	-6.50	-2.31
	CCSD(T) (subset)	1.99514	103.213	113.001	-9.25	-4.63

^a CP-corrected CCSD and CCSD(T) formation energies could not be obtained for this complex because the corresponding BSSE-style coupled-cluster treatment did not permit a reliable determination of the CP-corrected energy.

tially reduces the magnitude of the stabilization. At the coupled-cluster level, CP-corrected values could not be obtained for F_3B-PF_3 . The plain supermolecular CCSD and CCSD(T) single-point calculations for this complex converged successfully and displayed a well-behaved single-reference diagnostic, whereas the corresponding BSSE-style coupled-cluster calculations did not converge and yielded a much larger T_1 diagnostic. This indicates that the fragment-in-full-basis reference underlying the counterpoise-like decomposition constitutes a substantially more difficult electronic-structure problem than the

ordinary supermolecular calculation. For Cl_3B-PCl_3 , B3LYP-D3 appears to overestimate the destabilization of the adduct, since it predicts a clearly positive formation energy, whereas the correlated CCSD and CCSD(T) results indicate a borderline case with formation energies close to zero once BSSE effects are taken into account. CCSD (Δ -ML) yields negative uncorrected formation energies for all systems and thus recovers the same sign pattern as CCSD(T) before CP correction. Quantitatively, CCSD (Δ -ML) is systematically less negative than CCSD(T), with deviations of $\Delta E = +0.31$ (F),



+2.30 (Cl), and +2.75 (Br) kcal mol⁻¹. The increasing gap from F to Br suggests that correlation contributions beyond CCSD become progressively more important for the heavier, more polarizable halogen complexes. For F₃B–PF₃, the explicitly optimized CCSD_{opt} formation energy (–1.91 kcal mol⁻¹) is close to the CCSD(Δ -ML) value (–1.88 kcal mol⁻¹) and less negative than CCSD(T) (subset) (–2.19 kcal mol⁻¹), indicating that perturbative triples provide additional stabilization even for the weakly bound fluorinated adduct. Importantly, the B–P separations predicted by MP2 and CCSD (Δ -ML) are nearly identical for Cl and Br (Cl: 2.01380 vs. 2.01383 Å, Br: 1.99482 vs. 1.99514 Å, cf. Table I), indicating that the large differences in ΔE_{form} are not primarily driven by the optimized geometry but instead by the underlying electronic structure treatment. Taken together, Table I shows that CCSD (Δ -ML) substantially improves upon MP2 and B3LYP-D3 for these adducts, while the residual shift toward less negative ΔE_{form} relative to CCSD(T) highlights the growing role of perturbative triples (and/or higher-order) effects for Cl and Br. **Computational cost and motivation for ML-guided sampling:** All CCSD single-point calculations were carried out in parallel on 8 CPU cores (Intel Xeon E5-2683 v4 @ 2.10 GHz). A representative CCSD/aug-cc-pVTZ single-point evaluation required on the order of ~10 h wall time for the F₃B–PF₃ complex, which is the smallest system considered here (8 atoms, 74 electrons), with 115 basis-set shells and 450 Cartesian Gaussian basis functions, corresponding to 377 molecular orbitals (37 occupied) in the CCSD calculation (with 12 frozen-core orbitals and no frozen virtual orbitals). The heavier Cl₃B–PCl₃ and Br₃B–PBr₃ systems contain 122 and 240 electrons, respectively, and are therefore computationally more demanding.

In total, approximately 1000 CCSD single-point calculations were required in this study, comprising the labeled subsets used for training and for held-out evaluation together with additional low-energy candidates identified by the ML screening and subsequently verified at the CCSD level (see Sec. III C), corresponding to ~10,000 h (~14 months) of serial wall time, or ~80,000 core-hours. These CCSD calculations comprised both the labeled subset used for Δ -ML training/validation/testing and an additional set of ML-selected low-energy candidates that were recomputed explicitly at the CCSD level after screening of the full low-level library. In practice, these single-point calculations can be executed as large job arrays on HPC resources (e.g., tens of jobs concurrently), substantially reducing the elapsed wall time. In contrast, a fully relaxed CCSD geometry optimization of F₃B–PF₃ using numerical gradients (NUMGRD=.T. in GAMESS) required nearly three months on 8 CPU cores. Such an optimization is inherently iterative and far less amenable to the same degree of batching because successive optimization steps depend on the preceding geometry. This stark difference in cost between CCSD single-point evaluations and CCSD geometry optimization

underscores the practical importance of the ML-guided workflow, which concentrates high-level calculations on the most informative regions of configuration space while avoiding prohibitively expensive CCSD optimizations for large structure sets.

B. Structure libraries and CCSD-referenced low-energy ordering

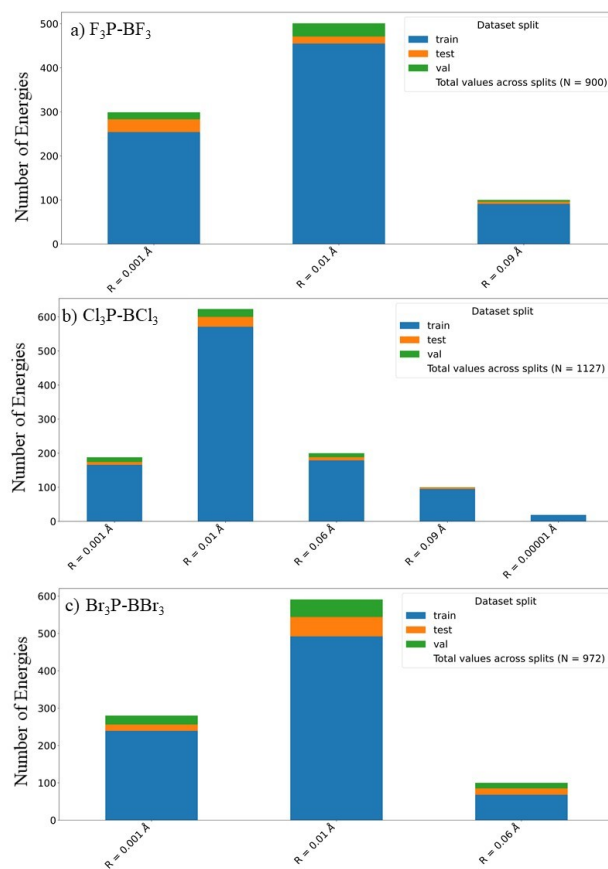


FIG. 2: Distribution of sampled geometries across subsets defined by the displacement-radius parameter R for a) F₃B–PF₃, b) Cl₃B–PCl₃, and c) Br₃B–PBr₃. Stacked bars denote the number of structures assigned to the training, test, and validation splits for each R -subset (energies grouped by the corresponding R -label files). The legend additionally lists the total number of entries across splits.

Fig. 2 shows the data for F₃B–PF₃, Cl₃B–PCl₃, and Br₃B–PBr₃ organized into subsets corresponding to different values of the displacement-radius parameter R , which controls the magnitude of the random atomic perturbations applied during structure generation. Each subset is further partitioned into training, test, and validation splits. The stacked-bar representation highlights how the sampling density varies with R for each complex while preserving a consistent split assignment within each



R-group. Although CCSD single-point energies at the equilibrium MP2 geometries are included in the training/validation/test datasets, they are intentionally omitted from the plotted categories and only contribute to the reported totals.

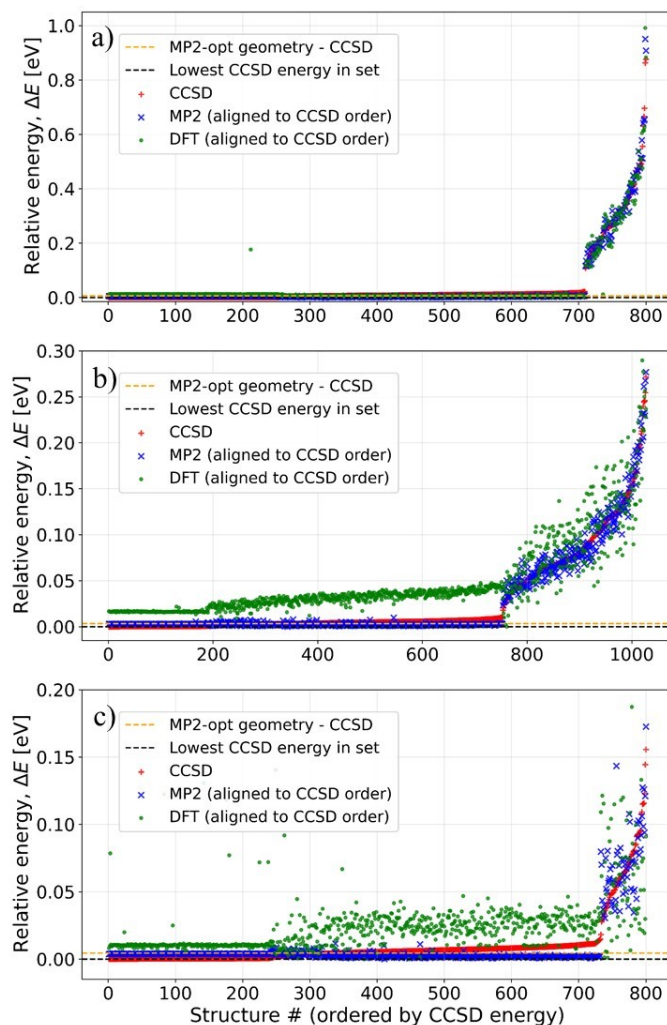


FIG. 3: Energy ordering of the sampled structures for a) F_3B-PF_3 , b) Cl_3B-PCl_3 , and c) Br_3B-PBr_3 . For each complex, structures are sorted by increasing CCSD energy, and the corresponding relative energies $\Delta E = E_i - E_{\min}$ are shown in eV for CCSD (red, left axis), MP2 (blue crosses, left axis), and DFT (green circles, right axis) in the same CCSD-imposed order. The orange dashed line marks the CCSD single-point energy evaluated at the MP2-optimized geometry, referenced to the minimum CCSD energy, while the black dashed line indicates the lowest CCSD energy present in the training set.

Figure 3 contrasts the CCSD relative energies with the associated MP2 and DFT relative energies for the matched CCSD-labeled subset after sorting all structures by increasing CCSD energy. To improve interpretability, the energies are shown relative to the minimum and

expressed in eV rather than in absolute atomic units. Plotting MP2 and DFT in the CCSD-imposed order highlights how well these lower-cost methods preserve the CCSD energetic ranking within this evaluated subset, and it also exposes the high-energy tail associated with more strongly distorted configurations. The CCSD relative energy at the MP2-optimized geometry (orange dashed line) and the lowest CCSD energy included in the training set (black dashed line) define reference points for the equilibrium region sampled by the dataset. Complementary histogram analyses of the relative-energy distributions and of the corresponding relative-energy differences for the matched CCSD-labeled subsets are provided in the Supplementary Information.

C. Δ -ML toward CCSD: descriptors, validation, and screening

To clarify the input representations used in the different machine-learning workflows, Table II summarizes the descriptor sets and preprocessing steps for the models considered in this work. For the descriptor-based workflows, the same handcrafted feature set was used across SVR, GBR, and GPR, so that the difference between these models lies in the regression algorithm rather than in the descriptor definition itself. In addition, a SOAP-based representation combined with PCA and SVR was evaluated as an alternative structural descriptor in the Supplementary Information.

The handcrafted descriptor vector comprised selected Cartesian coordinate components, key intermolecular and intramolecular distances, bond angles, and the corresponding low-level reference energy E_{low} . Moreover, representative descriptor examples and SOAP hyperparameters are reported (cf. SI).

The Δ -ML models were implemented within a pipeline that included support vector regression (SVR), gradient boosting regression (GBR), and Gaussian process regression (GPR). Hyperparameters were optimized with Optuna using 5-fold cross-validation on the training set. For each trial, the model was trained to predict the energy correction $\Delta E = E_{\text{CCSD}} - E_{\text{low}}$ on four folds and evaluated on the remaining fold, whereas the mean absolute error (MAE) averaged over the five validation folds was used as the optimization objective. Among the tested regressors, SVR yielded the most robust balance between cross-validated MAE and low-energy screening selectivity, and was therefore adopted for the final workflow. Figure 4 summarizes the descriptor relationships and their predictive relevance for the three donor-acceptor complexes. The correlation heatmaps in panels a-c) quantify linear dependencies among Cartesian coordinate components, selected interatomic distances and angles, and the low-level energy E_{MP2} , together with the CCSD energies available for the labeled subset. Complementary, the permutation-importance analysis in panels d-f), obtained by shuffling individual features in a trained



TABLE II: Overview of the machine-learning workflows used in this work. The descriptor-based SVR, GBR, and GPR models employ the same handcrafted input representation, whereas the SOAP-based workflow uses a structural descriptor followed by PCA and SVR. In all cases, the learned quantity is the correction $\Delta E = E_{\text{CCSD}} - E_{\text{low}}$.

Workflow	Input representation	Preprocessing	Role in this work
Descriptor + SVR	Handcrafted geometric descriptors + E_{low}	Standardization	Final production model
Descriptor + GBR	Same as above	Standardization	Benchmarked alternative
Descriptor + GPR	Same as above	Standardization	Benchmarked alternative
SOAP + SVR	SOAP descriptor + E_{low}	Standardization + PCA	Robustness check (SI)

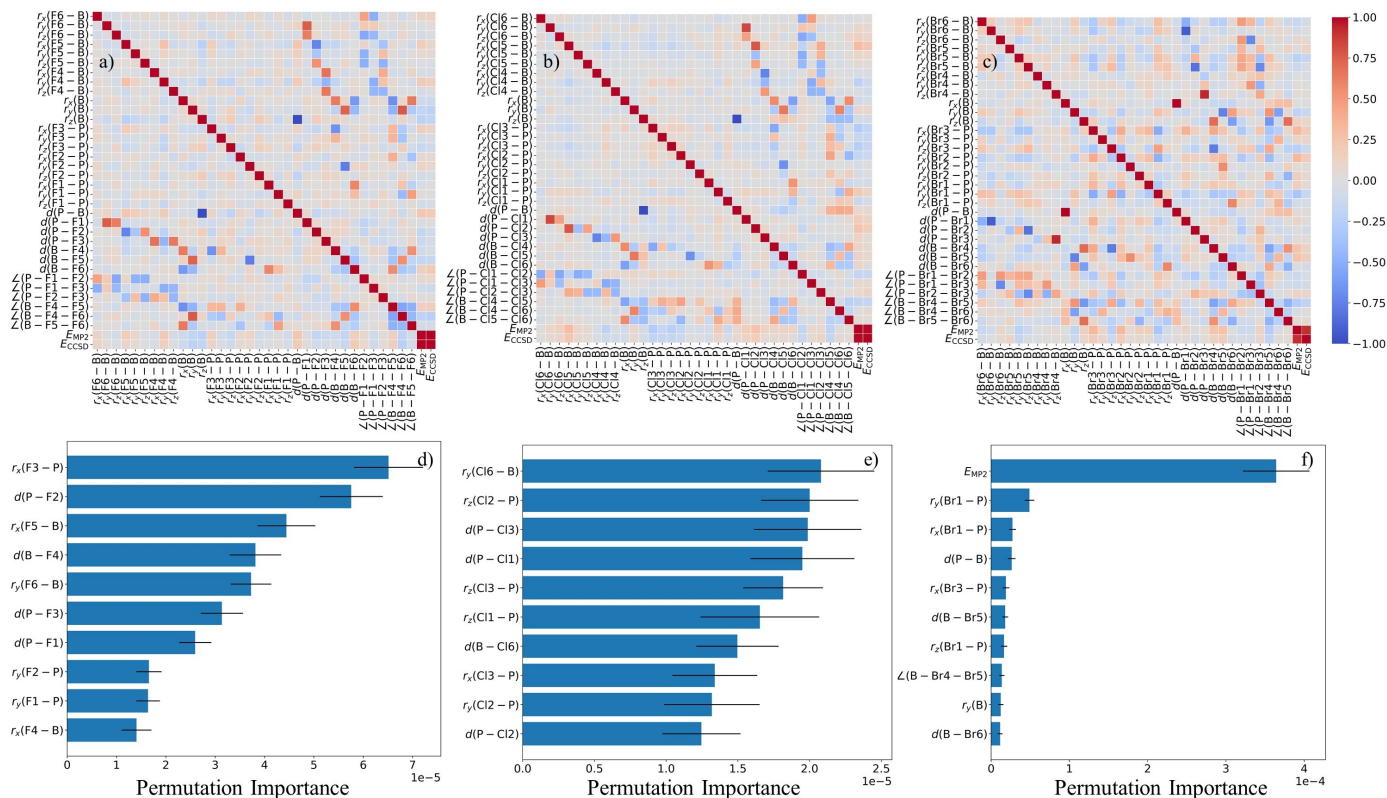


FIG. 4: Feature analysis for Δ -ML models of a,d) $\text{F}_3\text{B}-\text{PF}_3$, b,e) $\text{Cl}_3\text{B}-\text{PCl}_3$, and c,f) $\text{Br}_3\text{B}-\text{PBr}_3$. Panels a–c) show Pearson correlation matrices computed from the full descriptor set (selected Cartesian components $r_{x,y,z}$, key interatomic distances d , bond angles \angle , and the low-level energy), including correlations with E_{MP2} and the available E_{CCSD} values. Panels d–f) present the top-ranked permutation importances (mean \pm s.d.) for an SVR-based Δ -learning pipeline trained to predict $\Delta E = E_{\text{CCSD}} - E_{\text{MP2}}$, highlighting which geometric descriptors and the inclusion of E_{MP2} contribute most strongly to the predictive performance for each complex.

SVR Δ -learning model, identifies the variables that most strongly affect prediction of $\Delta E = E_{\text{CCSD}} - E_{\text{MP2}}$ (and thus E_{CCSD}). Notably, E_{MP2} plays a dominant, non-redundant role for $\text{Br}_3\text{B}-\text{PBr}_3$, in agreement with the stronger statistical dependence between ΔE and E_{MP2} in the brominated system, whereas for $\text{F}_3\text{B}-\text{PF}_3$ and $\text{Cl}_3\text{B}-\text{PCl}_3$ the correction can be captured largely by geometric descriptors with only minor added benefit from including E_{MP2} explicitly as an input feature. This behavior can be understood in light of the energetic hierarchy discussed above: $\text{F}_3\text{B}-\text{PF}_3$ is weakly bound and $\text{Cl}_3\text{B}-\text{PCl}_3$ lies near the threshold of binding once

higher-level correlation and BSSE effects are taken into account, whereas $\text{Br}_3\text{B}-\text{PBr}_3$ remains clearly bound. Accordingly, geometric descriptors are largely sufficient for the fluorinated and chlorinated systems, while the low-level energy E_{MP2} becomes more important for the more robustly bound brominated adduct. This system dependence further suggests that orbital-level descriptors, as used in MOB-ML-type frameworks, may provide an interesting alternative route toward a more uniform representation across the halogen series [35, 36].

In addition to the correlation and permutation-importance analyses shown in Fig. 4, we evaluated



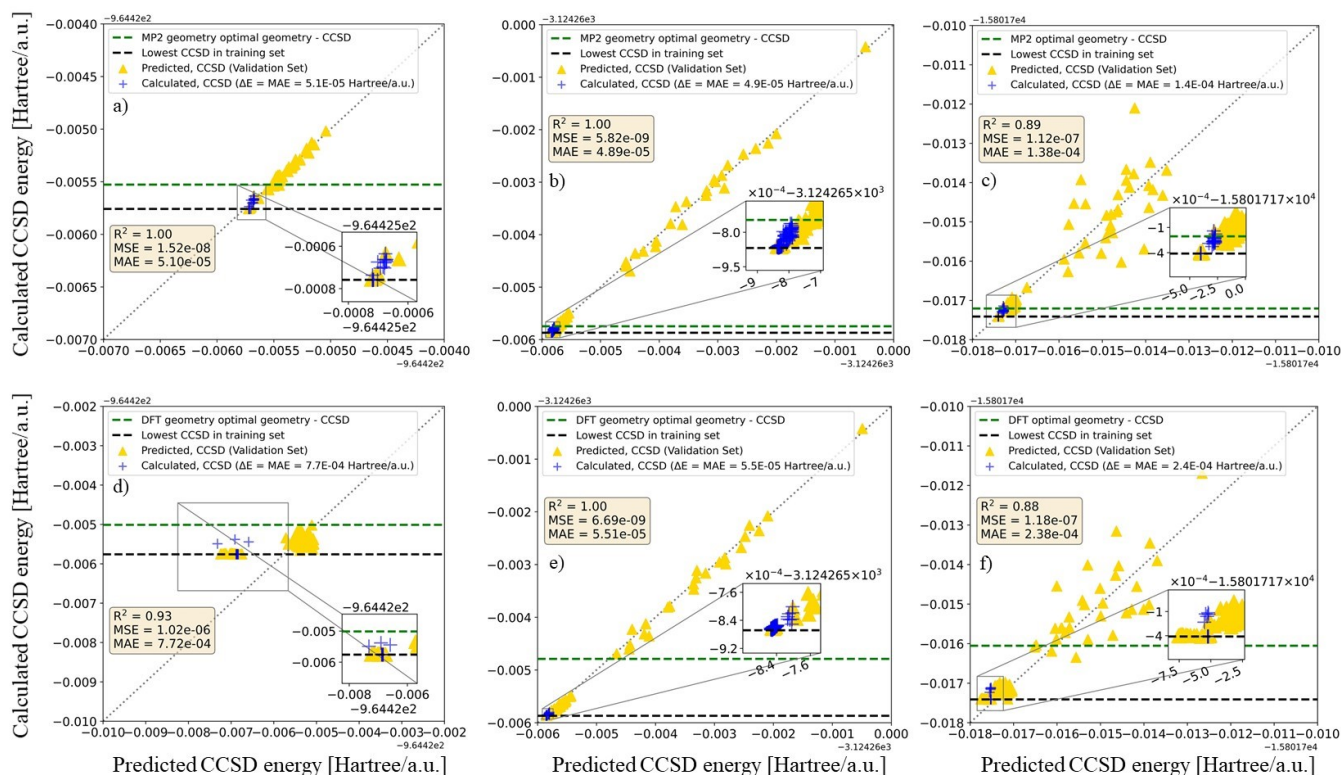


FIG. 5: Parity plots for the held-out (validation and test) CCSD-labeled subsets of a,d) F_3B-PF_3 , b,e) Cl_3B-PCl_3 , and c,f) Br_3B-PBr_3 , showing calculated CCSD energies as a function of the corresponding Δ -ML predicted CCSD energies. Top row a-c): models using MP2 as the low-level baseline and bottom row d-f) illustrate models using DFT as the baseline. Yellow triangles denote predicted CCSD energies for the held-out structures, while blue symbols indicate the corresponding CCSD reference values. Horizontal dashed lines mark the CCSD single-point energy at the optimized (MP2 or DFT) equilibrium geometry (green) and the lowest CCSD energy present in the training set (black). Insets zoom into the near-equilibrium region. Model performance is quantified by the corresponding R^2 and MAE values.

the role of including E_{MP2} as an input feature in the Δ -learning models. The corresponding permutation-importance results (with and without E_{MP2} appended to the geometric descriptor vector) are reported in the Supplementary Information. In line with the system-dependent behavior seen in Fig. 4d-f), removing E_{MP2} from the input features leads to a pronounced degradation in predictive performance for Br_3B-PBr_3 (e.g., $R^2 = 0.72$ and $MAE = 3.09 \times 10^{-4}$ Eh for the MP2-based SVR model without E_{MP2} , see SI), whereas F_3B-PF_3 and Cl_3B-PCl_3 remain largely unaffected (e.g., $R^2 = 0.997$ and $MAE = 1.71 \times 10^{-4}$ Eh for F_3B-PF_3 , and $R^2 = 0.989$ and $MAE = 1.05 \times 10^{-4}$ Eh for Cl_3B-PCl_3 , not shown here). This behavior is in agreement with the stronger dependence of the correction $\Delta E = E_{CCSD} - E_{MP2}$ on E_{MP2} for the brominated complex, as expressed by the larger magnitude of $\rho(\Delta E, E_{MP2})$ compared with the fluorinated and chlorinated systems. As an additional robustness check on the descriptor representation, we also evaluated an alternative machine learning pipeline based on smooth overlap of atomic positions (SOAP) descrip-

tors generated with the Dscribe package [55], followed by principal component analysis (PCA) and SVR regression. The resulting SOAP+PCA models and their predictive performance are described in the Supplementary Information. Importantly, the SOAP-based workflow identifies the same number of low-energy candidate structures for all three complexes as the handcrafted-descriptor models discussed here, supporting the robustness of the ML-guided screening conclusions with respect to the descriptor choice.

Figure 5 assesses the accuracy of the Δ -ML models by comparing predicted and calculated CCSD energies on the held-out non-training CCSD-labeled subsets. For F_3B-PF_3 and Br_3B-PBr_3 , 800 CCSD-labeled structures were used for training, whereas for Cl_3B-PCl_3 1027 CCSD-labeled structures were used (cf. Fig. 2). The remaining labeled structures were kept outside the training process and partitioned into validation and test subsets during model development. The parity plots are shown for the combined held-out subset in order to provide a compact visual summary of predictive agreement



beyond the training data. The MP2-based models (cf. Fig. 5a–c) and DFT-based models (see Fig. 5d–f) are shown separately to highlight the effect of the baseline method on the learned correction. The clustering of points along the parity line demonstrates that CCSD-quality energies are recovered with small errors, particularly in the near-equilibrium region highlighted in the insets. The achieved errors on the held-out non-training subsets are thus appropriate for low-energy screening and energetic ranking, but they are not intended to match the microhartree-level accuracy often required for dynamics- or spectroscopy-grade global PES applications [31]. The green and black reference lines mark additional benchmarks corresponding to the optimized equilibrium geometry and the lowest-energy structure included in the training set, respectively. Building on this validated accuracy, we next apply the CCSD-trained models to screen the full structure libraries and to define targeted CCSD(T) refinement sets.

Using the MP2 baseline, this screening step selected 111/156/122 low-energy structures for F_3B-PF_3 , Cl_3B-PCl_3 , and Br_3B-PBr_3 , respectively, with comparable candidate counts obtained from the DFT-baseline models. The screening workflow involves several nested subsets. First, only a labeled subset of the full several-thousand-structure DFT/MP2 libraries was evaluated at the CCSD level for training, validation, and testing of the Δ -ML model. The trained model was then applied to the remaining structures in the low-level libraries to predict CCSD-quality energies and identify a low-energy basin. Candidate structures falling within an MAE-based energy window around the predicted low-energy minimum were subsequently recomputed explicitly at the CCSD level to verify the ML-guided screening, with the MAE obtained on the held-out non-training subset serving as a practical tolerance for candidate selection. Finally, a still smaller subset of the most relevant low-energy candidates was refined further by single-point CCSD(T) calculations. To assess the impact of perturbative triples on the energetic ordering in the low-energy region, we then carried out targeted CCSD(T) single-point calculations for a refined subset that includes (i) the baseline-optimized equilibrium structure and (ii) the lowest-energy CCSD structure present in the training data, together with additional ML-selected candidates. This yields CCSD(T) refinement sets of 113/158/124 structures for F_3B-PF_3 , Cl_3B-PCl_3 , and Br_3B-PBr_3 (MP2 baseline), enabling a focused validation of CCSD-level screening against CCSD(T) energetics without exhaustive CCSD(T) sampling. Accordingly, the present workflow is intended to identify and refine the relevant low-energy basin rather than to guarantee direct recovery of the exact CCSD(T) minimum from the learned model alone.

The predictive quality of the trained models as a function of training-set size is illustrated in the learning curves shown in Fig. 6. Panel a) reports the mean absolute error of the predicted CCSD energies, while panel b) shows the corresponding R^2 values for the SVR

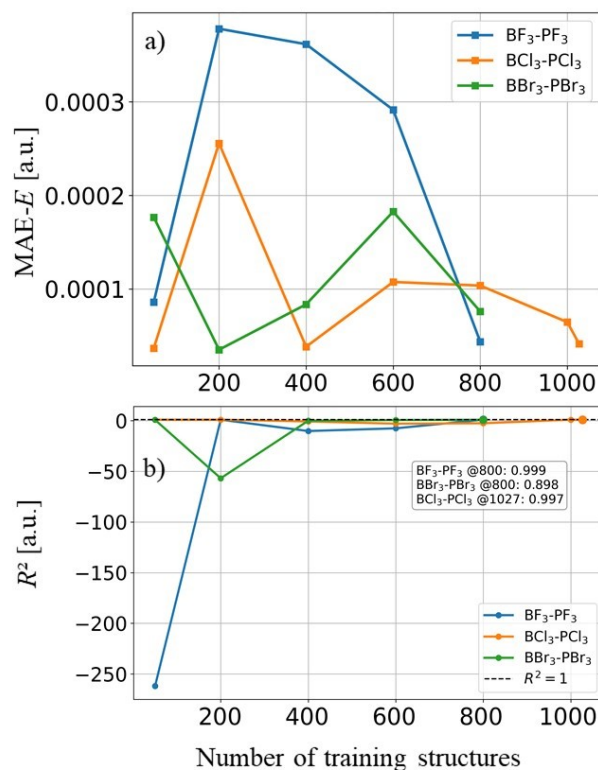


FIG. 6: Learning curves for the SVR Δ -ML models trained to reproduce CCSD energies using MP2 as the baseline method for F_3B-PF_3 , Cl_3B-PCl_3 , and Br_3B-PBr_3 . a) Mean absolute error (MAE-E) of the predicted CCSD energies as a function of the number of training structures. b) Corresponding coefficient of determination (R^2) values obtained on the evaluation sets. The dashed horizontal line marks the ideal limit $R^2 = 1$.

Δ -ML models trained on increasing numbers of structures. Across all three donor-acceptor complexes, the MAE decreases rapidly with increasing training-set size and reaches values on the order of 10^{-4} a.u., indicating that only a few hundred CCSD reference points are sufficient to achieve near chemical accuracy relative to the underlying CCSD reference energies. As the training set grows, the R^2 values approach unity, demonstrating that the models capture the overall energetic trends of the sampled configuration space with high fidelity. Among the three systems, F_3B-PF_3 exhibits the fastest convergence, reaching $R^2 = 0.999$ already at 800 training structures. The Cl_3B-PCl_3 complex shows similarly strong predictive performance and achieves $R^2 = 0.997$ when trained on 1027 structures. For the heavier Br_3B-PBr_3 system the convergence is somewhat slower, which can be attributed to its increased configurational flexibility and broader energy distribution of the sampled structures. Nevertheless, even in this case the model reaches $R^2 = 0.898$ with 800 training structures. These learning curves demonstrate that the Δ -ML framework efficiently



transfers CCSD accuracy across the configurational landscape and rationalize the high predictive quality observed in the parity plots displayed in Fig. 5.

D. Energy decomposition analysis

Although the relaxed B–P scans and the associated energy-decomposition analysis are not part of the Δ -ML training workflow, they provide a complementary physical interpretation of the low-energy PES region targeted by the ML-guided screening. In particular, this analysis serves two purposes in the present study. First, it allows us to assess whether the homohalogenated systems considered here display unusual bond-stretching behavior of the type previously reported for related phosphine haloboranes by Almas and Pearson [20]. Second, it helps rationalize the system-dependent energetic trends and bonding regimes underlying the different screening outcomes for F_3B-PF_3 , Cl_3B-PCl_3 , and Br_3B-PBr_3 .

Energy decomposition analysis (EDA) was used to obtain a physically transparent partitioning of the interaction along selected B–P separations. In the spirit of classical partitioning schemes such as the Morokuma approach [26] and the related DFT-based framework of Ziegler and Rauk [27], the binding between two fragments is separated into deformation and interaction contributions and the latter is further resolved into physically interpretable terms [28]. Here we employ the EDA implementation by Su *et al.* [29] in GAMESS [45], which decomposes the interaction energy as:

$$\Delta E^{\text{int}} = \Delta E^{\text{ele}} + \Delta E^{\text{ex}} + \Delta E^{\text{rep}} + \Delta E^{\text{pol}} + \Delta E^{\text{disp}}. \quad (2)$$

Such decompositions have proven particularly informative for trihaloboranes, where the stability of Lewis adducts reflects a competition between donor \rightarrow acceptor charge transfer, fragment deformation (pyramidalization), and halogen-dependent polarization/back-donation effects [52]. Accordingly, distance-resolved PES scans and EDA evaluations at selected B–P separations are used here to provide a chemically transparent interpretation of the bonding regimes and local PES features relevant to the screened low-energy region.

The Morokuma energy decomposition analysis (EDA) affords a chemically transparent interpretation of the donor–acceptor interaction by separating the total interaction energy into electrostatic (ΔE^{ele}), exchange (ΔE^{ex}), Pauli repulsion (ΔE^{rep}), polarization (ΔE^{pol}), and dispersion (ΔE^{disp}) contributions (cf. Table III). For all three complexes, the strongly compressed region ($r \approx 1.5 \text{ \AA}$) is dominated by a steep repulsive wall, as reflected by the very large positive ΔE^{rep} values (+718, +849, and +868 kcal mol⁻¹ for F_3B-PF_3 , Cl_3B-PCl_3 , and Br_3B-PBr_3 , respectively). This term originates from Pauli exclusion between like-spin electrons when fragment densities overlap strongly.

Although ΔE^{ele} , ΔE^{ex} , and ΔE^{pol} are highly stabilizing at such short separations, their sum cannot compensate the extreme Pauli repulsion, yielding net repulsive interaction energies at $r = 1.5 \text{ \AA}$ in each system.

Upon increasing the B \cdots P separation toward chemically relevant distances, ΔE^{rep} decreases rapidly and the balance between attractive and repulsive contributions becomes favorable. In the F_3B-PF_3 complex, the minimum occurs at a relatively long separation ($r_{\text{eq}} = 3.23715 \text{ \AA}$), where the interaction is weak ($\Delta E^{\text{int}} \approx -1.29 \text{ kcal mol}^{-1}$). At this distance, the orbital-interaction signatures are modest, with only small exchange and polarization contributions ($\Delta E^{\text{ex}} \approx -6.00$ and $\Delta E^{\text{pol}} \approx -1.21 \text{ kcal mol}^{-1}$), which points to limited donor–acceptor mixing at the minimum. Within the conceptual framework employed in our earlier work, such small ΔE^{pol} values are characteristic of binding dominated by long-range electrostatics and dispersion, with comparatively weak covalent/dative contributions. Consequently, the F_3B-PF_3 adduct reaches its most favorable interaction energy in a regime where Pauli repulsion is strongly reduced ($\Delta E^{\text{rep}} = +10.98 \text{ kcal mol}^{-1}$) and strong short-range orbital effects remain suppressed.

In contrast, the chlorinated and brominated systems exhibit a qualitatively different energetic signature: their minima occur at short separations ($r_{\text{eq}} \approx 2.0 \text{ \AA}$), i.e., in a region of substantial density overlap. For Cl_3B-PCl_3 , the equilibrium interaction energy is strongly stabilizing ($\Delta E^{\text{int}} \approx -27.50 \text{ kcal mol}^{-1}$) and is accompanied by large exchange and polarization terms ($\Delta E^{\text{ex}} \approx -156.48$ and $\Delta E^{\text{pol}} \approx -91.44 \text{ kcal mol}^{-1}$). For Br_3B-PBr_3 , the stabilization is even greater ($\Delta E^{\text{int}} \approx -31.85 \text{ kcal mol}^{-1}$) with correspondingly stronger exchange and polarization contributions ($\Delta E^{\text{ex}} \approx -173.69$ and $\Delta E^{\text{pol}} \approx -102.98 \text{ kcal mol}^{-1}$). Following the interpretation used in our previous study, pronounced negative ΔE^{pol} values indicate substantial reshaping of fragment orbitals upon complex formation, which is characteristic of charge transfer and donor–acceptor orbital mixing, while strongly negative ΔE^{ex} values reflect significant overlap-driven stabilization in the bonded region. Together, these features point to a markedly increased covalent/dative component for the BCl_3 and BBr_3 adducts relative to BF_3 . A key observation is that the short equilibrium distances in the Cl and Br cases are achieved despite *very large* Pauli repulsion at the minima: $\Delta E^{\text{rep}} = +317.37 \text{ kcal mol}^{-1}$ for Cl_3B-PCl_3 and $\Delta E^{\text{rep}} = +349.74 \text{ kcal mol}^{-1}$ for Br_3B-PBr_3 as listed in Table III. Rather than indicating an anomalous result, these values provide a diagnostic fingerprint of a short-range donor–acceptor bond: the system pays a substantial Pauli penalty in order to access geometries where the stabilizing orbital-interaction contributions (exchange and polarization) are even larger in magnitude. In other words, the net attraction at r_{eq} for the Cl/Br adducts arises from a compensation mechanism in which strong overlap-driven stabilization outweighs



TABLE III: Results of the the energy decomposition components (kcal mol⁻¹) at MP2/aug-cc-pVTZ level of theory for the relaxed scans.

Complex	r (Å)	$\Delta E^{(ele)}$	$\Delta E^{(ex)}$	$\Delta E^{(rep)}$	$\Delta E^{(pol)}$	$\Delta E^{(disp)}$	$\Delta E^{(int)}$
(1) F ₃ B–PF ₃	1.5	-160.37	-304.73	718.30	-191.15	-8.87	53.18
	2.0	-68.12	-113.84	241.86	-67.46	-4.55	-12.10
	2.5	-21.60	-39.34	77.87	-13.97	-3.94	-0.98
	3.0	-6.25	-11.54	21.60	-2.53	-2.30	-1.03
	3.5	-1.74	-2.79	5.00	-0.57	-1.12	-1.21
	r_{eq}	-3.41	-6.00	10.98	-1.21	-1.65	-1.29
(2) Cl ₃ B–PCl ₃	1.5	-183.43	-375.04	848.83	-231.41	-25.00	33.96
	2.0	-82.30	-160.46	326.24	-94.32	-16.99	-27.83
	2.5	-27.83	-63.62	120.06	-22.70	-13.15	-7.23
	3.0	-8.86	-23.65	41.94	-4.14	-8.25	-2.96
	3.5	-2.52	-7.22	12.13	-0.83	-4.51	-2.96
	r_{eq}	-80.07	-156.48	317.37	-91.44	-16.89	-27.50
(3) Br ₃ B–PBr ₃	1.5	-179.12	-386.07	868.29	-245.00	-30.12	27.98
	2.0	-82.34	-172.14	346.29	-101.83	-21.72	-31.74
	2.5	-29.18	-71.32	132.97	-26.37	-17.24	-11.15
	3.0	-9.82	-28.24	49.53	-5.23	-11.20	-4.96
	3.5	-2.89	-9.23	15.32	-1.11	-6.40	-4.30
	r_{eq}	-83.16	-173.69	349.74	-102.98	-21.77	-31.85

strong Pauli repulsion. In contrast, F₃B–PF₃ avoids this overlap-dominated regime by adopting a larger equilibrium separation, simultaneously reducing $\Delta E^{(rep)}$ and suppressing the magnitude of exchange and polarization stabilization. The dispersion contribution introduces an additional systematic stabilization across the halogen series. At equilibrium, $\Delta E^{(disp)}$ increases in magnitude from BF₃ (≈ -1.65 kcal mol⁻¹) to BCl₃ (≈ -16.89 kcal mol⁻¹) and remains large for BBr₃ (≈ -21.77 kcal mol⁻¹). This trend follows the increasing polarizability of the heavier halogenated fragments and helps explain the more negative $\Delta E^{(int)}$ values for the chlorinated and brominated systems. While dispersion alone does not enforce the short equilibrium distances, it contributes substantially to the overall stabilization in the Cl/Br cases and amplifies the net effect of electrostatics and orbital interactions.

Finally, the distance dependence of the EDA components provides a unified bonding picture for the three adducts. At larger separations (e.g., $r = 3.0$ – 3.5 Å), exchange and polarization contributions diminish markedly, as orbital overlap decreases and the interaction becomes increasingly dominated by electrostatics and dispersion. The crucial distinction is that the F₃B–PF₃ complex reaches its minimum in this weak-interaction regime, whereas Cl₃B–PCl₃ and Br₃B–PBr₃ attain their minima at shorter distances where the interaction energy is governed by a balance between very large Pauli repulsion and even larger overlap- and polarization-driven stabilization. Consequently, the Morokuma EDA suggests that F₃B–PF₃ is best described as a weak donor–acceptor complex with largely electrostatic/dispersion character, whereas the chlorinated and brominated analogues exhibit substantially

stronger donor–acceptor bonding with pronounced covalent contributions, as evidenced by their short equilibrium B–P separations and more negative $\Delta E^{(int)}$ values.

Figure 7 depicts the relaxed B–P bond-stretching scans for the three homohalogenated trihaloborane–trihalophosphine complexes. For each system, the constrained scan was performed along the B–P distance r using relaxed geometries at each fixed separation, and the resulting energies were decomposed into the total energy E_{total} , the fragment energies E_{acid} (BX₃) and E_{base} (PX₃) evaluated at the complex geometry, and the interaction energy:

$$E_{int}(r) = E_{total}(r) - (E_{acid}(r) + E_{base}(r)). \quad (3)$$

All curves are shown as shifted energies, $E(r) - E(r_{eq})$, in units of kcal mol⁻¹, to facilitate a direct comparison of the local PES shapes across the three adducts. A key qualitative outcome of Fig. 7 is that none of the three systems exhibits the bimodal (two-minima) dissociation profile reported by Almas and Pearson for Cl₃B–PH₃ [20]. Instead, each complex shows a single-well dissociation curve: following the equilibrium region, $E_{total}(r) - E_{total}(r_{eq})$ increases smoothly toward the separated-fragment limit without developing a second outer minimum. In all three panels, the decomposition reveals the same underlying balance of physical effects. At very short separations, the steep repulsive wall in E_{total} arises from (i) substantial fragment deformation, most prominently in the Lewis acid term E_{acid} (pyramidalization/distortion of BX₃), together with (ii) strongly repulsive E_{int} (Pauli/steric repulsion at overly compressed contacts). Upon increasing r , the deformation penalty decreases rapidly, while E_{int} transitions



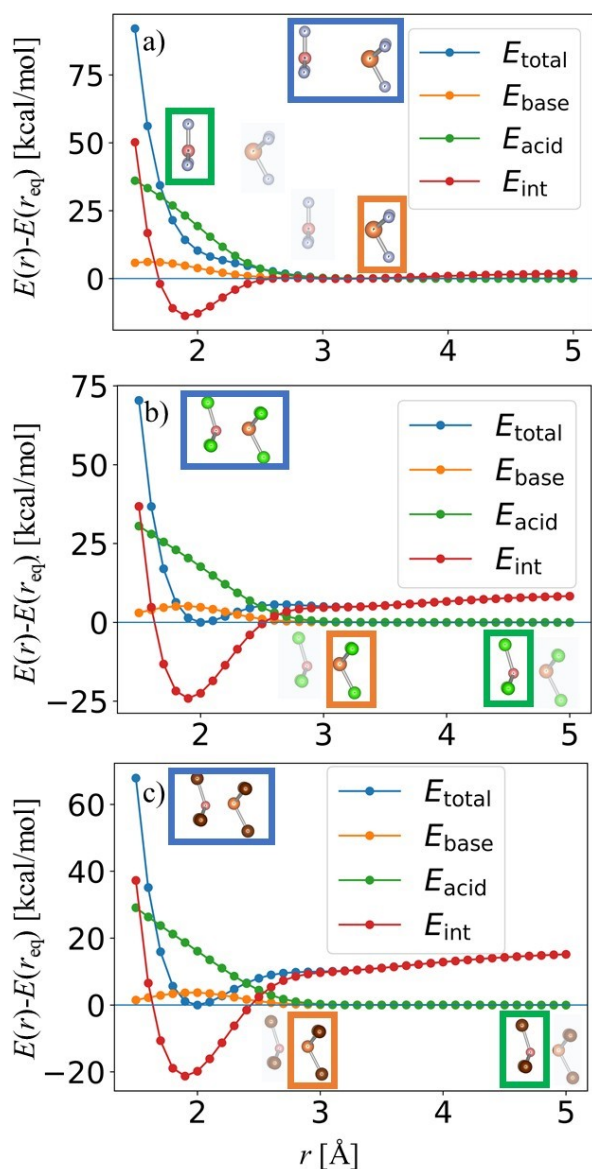


FIG. 7: Relaxed B–P bond-stretching scans with a decomposition of the total energy into fragment and interaction contributions for the homohalogenated trihaloborane–trihalophosphine complexes a) $\text{F}_3\text{B}-\text{PF}_3$, b) $\text{Cl}_3\text{B}-\text{PCl}_3$, and c) $\text{Br}_3\text{B}-\text{PBr}_3$. The shifted energies $E(r) - E(r_{\text{eq}})$ (kcal mol^{-1}) are plotted as a function of the constrained B–P distance r (\AA): the total energy E_{total} , the fragment energies E_{acid} and E_{base} are evaluated at the complex geometry, and the interaction energy $E_{\text{int}} = E_{\text{total}} - (E_{\text{acid}} + E_{\text{base}})$. The inset structures illustrate representative geometries along the scan: the colored frames indicate which quantity is being highlighted, namely the full complex (E_{total} , blue), the isolated base fragment at the complex geometry (E_{base} , orange), and the isolated acid fragment at the complex geometry (E_{acid} , green).

from repulsive to attractive (negative), producing a single global minimum where the stabilization gained from interaction is no longer worth additional deformation of the fragments. Beyond the minimum, E_{int} weakens and the total energy rises monotonically, indicating that any long-range outer-well stabilization is insufficient to generate a second minimum for these homohalogenated pairs within the present scan protocol and level of theory.

The three systems share this common topology but differ in magnitude and range of stabilization. The fluorinated adduct $\text{F}_3\text{B}-\text{PF}_3$ (see Fig. 7a) displays the weakest long-range binding signature (a comparatively small energetic rise above the minimum by $r = 5 \text{ \AA}$), in accordance with a shallow donor–acceptor association at relatively long equilibrium distance. In contrast, the chloro- and bromo-substituted systems (cf. Fig. 7b, c) show markedly stronger short-range stabilization (more negative E_{int} near the minimum) accompanied by a larger deformation cost at compressed distances, consistent with a more pronounced competition between orbital/charge-transfer stabilization and BX_3 pyramidalization in the vicinity of the minimum. Guided by these decompositions and in analogy to the distance-resolved analyses in Ref. [20], the subsequent energy-decomposition analysis (EDA) calculations are performed at a small set of representative separations sampling distinct PES regimes: a compressed geometry ($r \approx r_{\text{eq}} - 0.6 \text{ \AA}$), the equilibrium geometry ($r = r_{\text{eq}}$), and one to two post-minimum separations ($r \approx r_{\text{eq}} + 0.6 \text{ \AA}$ and $r \approx r_{\text{eq}} + 1.2 \text{ \AA}$), using the fully relaxed constrained-scan geometries at each chosen r . This selection cleanly separates the repulsive/distortion-dominated region from the near-equilibrium bonding regime and the long-range tail, enabling a physically transparent interpretation of which EDA terms govern the shape of the local PES in each complex.

IV. SUMMARY

This study introduces a CCSD-referenced Δ -machine learning framework for coupled-cluster-quality screening of low-energy structures in homohalogenated borane–phosphine adducts, with $\text{F}_3\text{B}-\text{PF}_3$, $\text{Cl}_3\text{B}-\text{PCl}_3$, and $\text{Br}_3\text{B}-\text{PBr}_3$ as representative cases. Large geometry libraries generated at MP2 and DFT (B3LYP-D3) levels were used to train regression models for the correction $\Delta E = E_{\text{CCSD}} - E_{\text{low}}$, enabling CCSD-quality energies across thousands of structures at greatly reduced cost. Analysis of the held-out non-training subsets demonstrates that the resulting Δ -ML models reproduce CCSD energies with low errors and preserve the energetic ordering in the near-equilibrium region, offering a practical route for ranking low-energy structures on shallow potential-energy landscapes. Using the CCSD-trained predictors, we screened the complete MP2/DFT libraries and identified compact low-energy candidate sets for further refinement.



Targeted CCSD(T) calculations on the ML-selected subsets show that MP2 substantially overbinds for the chlorinated and brominated adducts, whereas B3LYP-D3 can even predict the wrong sign of the formation energy for $\text{Cl}_3\text{B}-\text{PCl}_3$ by placing this borderline system too far on the unfavorable side. In contrast, the CCSD (Δ -ML) screening recovers the correct uncorrected stability sign for all three adducts and yields formation energies that track the CCSD(T) reference trends substantially better than the underlying low-level methods. Inclusion of BSSE corrections further sharpens the picture, with $\text{Cl}_3\text{B}-\text{PCl}_3$ emerging as a near-threshold adduct and $\text{Br}_3\text{B}-\text{PBr}_3$ remaining clearly bound, whereas for $\text{F}_3\text{B}-\text{PF}_3$ the supermolecular CCSD and CCSD(T) calculations support weak binding, but CP-corrected coupled-cluster values could not be established because the corresponding BSSE treatment proved numerically unstable.

Distance-resolved relaxed scans and Morokuma-type EDA together establish a coherent physical interpretation of the different binding regimes. $\text{F}_3\text{B}-\text{PF}_3$ stabilizes at a longer B...P separation in a weak-interaction regime with comparatively small polarization/orbital contributions, whereas the chlorinated and brominated adducts adopt shorter equilibrium distances where large Pauli repulsion is overcompensated by strong exchange

and polarization, with dispersion increasing systematically from F to Cl to Br. None of the three homohalogenated systems exhibits a second, outer minimum in the relaxed B-P stretching profile under the present scan protocol.

Overall, the proposed Δ -ML strategy bridges the gap between large-scale structure sampling and correlated wave-function accuracy by enabling CCSD-level screening and selective CCSD(T) validation in the low-energy region. The results also show that for shallow donor-acceptor complexes, BSSE correction, higher-level geometry refinement, and selective high-level validation are all essential for a reliable energetic classification. This workflow is expected to be transferable to other weak donor-acceptor complexes for which correlated reference data are required but exhaustive high-level exploration is computationally prohibitive.

ACKNOWLEDGMENTS

I gratefully acknowledge the computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, FWO and the Flemish Government - department EWI.

-
- [1] A. Staubitz, A. P. M. Robertson, M. E. Sloan, I. Manners, *Chem. Rev.* **110**, 4023–4078 (2010).
- [2] D. R. Martin, *Chem. Rev.* **34**, 461–473 (1944).
- [3] H. Booth, D. Martin, *Boron Trifluoride and Its Derivatives*, John Wiley & Sons, New York, NY, (1949), Chap. 4.
- [4] D. Martin, R. E. Dial, *J. Am. Chem. Soc.* **72**, 852–856 (1950).
- [5] V. Jonas, G. Frenking, M. T. Reetz, *J. Am. Chem. Soc.* **116**, 8741–8753 (1994).
- [6] F. Bessac, G. Frenking, *Inorg. Chem.* **42**, 7990–7994 (2003).
- [7] F. Bessac, G. Frenking, *Inorg. Chem.* **45**, 6956–6964 (2006).
- [8] T. A. Holme, T. N. Truong, *Chem. Phys. Lett.* **215**, 53–57 (1993).
- [9] J. O. Jensen, *J. Mol. Struct.: THEOCHEM* **686**, 165–172 (2004).
- [10] T. L. Allen, W. H. Fink, *Inorg. Chem.* **32**, 4230–4234 (1993).
- [11] J. R. Durig, Z. Shen, *J. Mol. Struct.: THEOCHEM* **397**, 179–190 (1997).
- [12] J. A. Plumley, J. D. Evanseck, *J. Phys. Chem. A* **111**, 13472–13483 (2007).
- [13] J. A. Plumley, J. D. Evanseck, *J. Chem. Theory Comput.* **4**, 1249–1253 (2008).
- [14] J. A. Plumley, J. D. Evanseck, *J. Phys. Chem. A* **113**, 5985–5992 (2009).
- [15] B. G. Janesko, *J. Chem. Theory Comput.* **6**, 1825–1833 (2010).
- [16] F. Huang, J. Jiang, M. Wen, Z.-X. Wang, *J. Theor. Comput. Chem.* **13**, 1350074 (2014).
- [17] C. F. Pupim, N. H. Morgon, A. Lopez-Castillo, *J. Braz. Chem. Soc.* **26**, 1648–1655 (2015).
- [18] T. Leyssens, D. Peeters, A. G. Orpen, J. N. Harvey, *New J. Chem.* **29**, 1424–1430 (2005).
- [19] H. Anane, S. El Houssame, A. El Guerraze, A. Jarid, A. Boutalib, I. Nebot-Gil, F. Tomas, *J. Mol. Struct.: THEOCHEM* **709**, 103–107 (2004).
- [20] Q. L. Almas, J. K. Pearson, *ACS Omega* **3**, 608–614 (2018).
- [21] P. S. Bryan, R. L. Kuczkowski, *Inorg. Chem.* **11**, 553–559 (1972).
- [22] O. Köksal, *Comput. Theor. Chem.* **1253**, 115398 (2025).
- [23] T. H. Dunning Jr., *J. Chem. Phys.* **90**, 1007 (1989).
- [24] J. Cizek, *J. Chem. Phys.* **45**, 4256 (1966).
- [25] O. Christiansen, H. Koch, P. Jørgensen, *Chem. Phys. Lett.* **243**, 409–418 (1995).
- [26] K. Morokuma, *J. Chem. Phys.* **55**, 1236 (1971).
- [27] T. Ziegler, A. Rauk, *Theor. Chim. Acta* **46**, 1 (1977).
- [28] A. E. Reed, L. A. Curtiss, F. Weinhold, *Chem. Rev.* **88**, 899 (1988).
- [29] P. Su, H. Li, *J. Chem. Phys.* **131**, 014102 (2009).
- [30] B. J. Braams, J. M. Bowman, *Int. Rev. Phys. Chem.* **28**, 577–606 (2009).
- [31] C. Qu, Q. Yu, J. M. Bowman, *Annu. Rev. Phys. Chem.* **69**, 151–175 (2018).
- [32] J. Behler, *Phys. Chem. Chem. Phys.* **13**, 17930–17955 (2011).
- [33] J. Behler, *Int. J. Quantum Chem.* **115**, 1032–1050 (2015).



- [34] J. Cui, R. V. Krems, *J. Phys. B: At. Mol. Opt. Phys.* **49**, 224001 (2016).
- [35] M. Welborn, L. Cheng, and T. F. Miller III, *J. Chem. Theory Comput.* **14**, 4772–4779 (2018).
- [36] L. Cheng, M. Welborn, A. S. Christensen, and T. F. Miller III, *J. Chem. Phys.* **150**, 131103 (2019).
- [37] T. D. Crawford, H. F. Schaefer III, *Rev. Comput. Chem.* **14**, 33–136 (1999).
- [38] R. J. Bartlett, M. Musiał, *Rev. Mod. Phys.* **79**, 291–352 (2007).
- [39] K. Raghavachari, G. W. Trucks, J. A. Pople, M. Head-Gordon, *Chem. Phys. Lett.* **157**, 479–483 (1989).
- [40] C. Qu, J. M. Bowman, *Phys. Chem. Chem. Phys.* **18**, 24835–24840 (2016).
- [41] R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. von Lilienfeld, *J. Chem. Theory Comput.* **11**, 2087 (2015).
- [42] A. Nandi, B. Qu, R. C. Conte, J. M. Bowman, *J. Chem. Phys.* **154**, 051102 (2021).
- [43] J. M. Bowman, C. Qu, R. Conte, A. Nandi, P. L. Houston, and Q. Yu, *J. Chem. Theory Comput.* **19**, 1–17 (2023).
- [44] S. P. Khire, R. N. Gurav, S. J. Gadre, A. Nandi, *J. Phys. Chem. A* **126**, 1458 (2022).
- [45] G. M. J. Barca, C. Bertoni, L. Carrington, D. Datta, N. De Silva, J. E. Deustua, D. G. Fedorov, J. R. Gour, A. O. Gunina, E. Guidez, T. Harville, S. Irle, J. Ivanic, K. Kowalski, L. Karol, S. Sarom, H. Li, W. Li, J. J. Lutz, I. Magoulas, J. Mato, V. Mironov, H. Nakata, B. Q. Pham, P. Piecuch, D. Poole, S. R. Pruitt, A. P. Rendell, L. B. Roskop, K. Ruedenberg, T. Sattasathuchana, M. W. Schmidt, J. Shen, L. Slipchenko, M. Sosonkina, V. Sundriyal, A. Tiwari, J. L. Galvez Vallejo, B. Westheimer, M. Wloch, P. Xu, F. Zahariev, M. S. Gordon, *J. Chem. Phys.* **152**, 154102 (2020).
- [46] C. Møller, M. S. Plesset, *Phys. Rev.* **46**, 618 (1934).
- [47] M. Head-Gordon, J. A. Pople, M. J. Frisch, *Chem. Phys. Lett.* **153**, 503–506 (1988).
- [48] A. D. Becke, *J. Chem. Phys.* **98**, 1372–1377 (1993).
- [49] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **132**, 154104, (2010).
- [50] Warren J. Hehre, Leo Radom, Paul v. R. Schleyer, John A. Pople, *Ab Initio Molecular Orbital Theory*, John Wiley & Sons, New York, (1986).
- [51] S. B. Boys, F. Bernardi, *Mol. Phys.* **19**, 553 (1970).
- [52] D. R. Silva, L. A. Santos, M. P. Freitas, C. F. Guerra, T. A. Hamlin, *Chem. Asian J.* **15**, 4043 (2020).
- [53] S. J. Grabowski, *Molecules* **25**, 4668 (2020).
- [54] S. J. Grabowski, *Molecules* **25**, 2703 (2020).
- [55] L. Himanen, M. O. J. Jäger, E. V. Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke, and A. S. Foster, *Comput. Phys. Commun.* **247**, 106949 (2020).

