

Integrated Study on Methane Activation: Exploring Main Group Frustrated Lewis Pairs through Density Functional Theory, Machine Learning, and Machine-Learned Force Fields

Ignacio Migliaro and Thomas R. Cundari*

Cite This: *J. Chem. Theory Comput.* 2024, 20, 6388–6401

Read Online

ACCESS |



Metrics & More

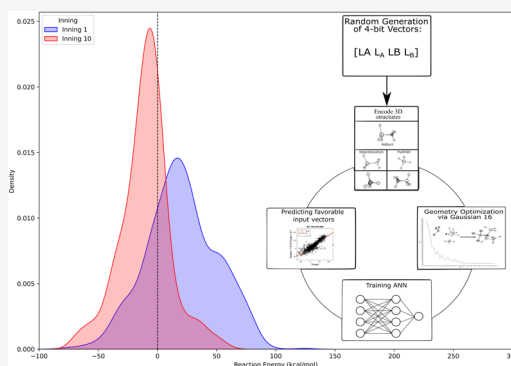


Article Recommendations



Supporting Information

ABSTRACT: Frustrated Lewis Pairs (FLP) are an important advance in metal-free catalysis due to their ability to activate a variety of small molecules. Many studies have focused on a very limited sample of Lewis acids and bases. Herein, we disclose an automated exploration algorithm using density functional methods, artificial neural networks (ANNs), and a molecule builder that incentivizes the exploration of favorable FLP space for the activation of methane *via* two mechanisms: deprotonation and hydride abstraction. The exploration algorithm creates FLPs with different Lewis acids (LA), Lewis bases (LB), and their substituents (L_A/L_B), which proved successful in quickly converging in the favorable chemical space, suggesting chemically sound structures, and generating thousands of potential candidates for methane activating FLPs. By modeling thousands of reactions, an FLP database of methane activation was created, allowing one to data mine properties, *e.g.*, adduct bond length, highest occupied molecular orbital–lowest-unoccupied molecular orbital (HOMO–LUMO) gap, global electrophilicity index, favored Lewis acids/bases/substituents, and substituent steric volume. These properties not only successfully narrow the FLP chemical space but also provide meaningful insight into the chemical nature of competent methane activators. The machine learning discovery strategy disclosed here is general enough to be applicable to many chemical optimization tasks. This study also investigates the efficacy of a Machine-Learned Force Field (MLFF) in predicting the formation energies of Frustrated Lewis Pairs (FLPs). Our model, exhibiting a test error of ± 10 kcal/mol, highlighted impressive computational efficiency by enabling the calculation of all possible FLP permutations within our chemical space. The MLFF demonstrated proficiency in predicting energies, providing a significant acceleration compared to quantum mechanics methods. However, challenges emerged in accurately capturing forces, necessitating recourse to classical force fields for reliable structure relaxation. The present study sheds light on the MLFF's potential as a tool for rapid energy predictions, emphasizing the need for further refinement to enhance its accuracy, particularly in force predictions, to expand its utility in chemical simulations.



INTRODUCTION

Since the discovery of Frustrated Lewis Pairs (FLPs) by Stephan et al., these compounds have provided an efficient route to activation of dihydrogen and activated C–H bonds,^{1–4} as well as other small molecules such as CO,⁵ N₂O, NO, SO₂, and SO.^{5–9} Methane is an abundant chemical compound that has the potential as a hydrocarbon feedstock for multiple industrial applications; however, no FLP-mediated activation of this alkane has been reported.¹⁰ Frustrated Lewis Pairs provide an alternative to conventional transition metal catalysis. FLPs exploit the motif of sterically hindered Lewis acids and Lewis bases, which forestalls the formation of a stable adduct, to activate small molecules with the use of cheaper and Earth-abundant elements, most typically those of the main group. Different studies have focused on a limited array of bulky (often alkyl) substituents, Lewis acids, and/or Lewis bases. In our previous research on modeling trihalides and pnictogen pentahalides as Lewis acids with an ammonia Lewis

base pair, heavier Lewis acids (those based on Bi, Sb, and Tl) were found to have more favorable activation and reaction free energies as compared to Lewis acids centered around lighter main group elements.¹¹ Moreover, as compared to many transition metal intermediates for alkane C–H activation, these reactions were reasonably close to thermoneutral, which should aid in subsequent functionalization steps in a catalytic cycle.

In a previous study of FLP-mediated methane activation,¹¹ it proved difficult to elucidate the chemical factors that most greatly impact methane activation free energies using statistical

Received: March 20, 2024

Revised: June 13, 2024

Accepted: June 14, 2024

Published: June 28, 2024



regression approaches. Therefore, an extension of the work is desirable to increase the chemical space of the potential FLP constituents evaluated. Using machine learning (ML) methods, a protocol is described here that automates the discovery of p-block FLPs for methane activation, modeling a robust database of Lewis acids, Lewis bases, and different substituents. Although specifically applied to the main group FLPs herein, the protocols developed are general enough to be applicable to the ML-directed discovery of other novel catalyst leads.

Searching the chemical space for FLPs for promising methane activators is infeasible even by using cutting-edge density functional theory (DFT) given the myriads of permutations one must generate and then evaluate. Machine learning addresses a key issue of dealing with the immensity of searchable chemical space by providing an effective nonlinear predictor to connect key structural properties (frontier orbital energies, steric parameters, *etc.*) with desired catalyst outcomes (*e.g.*, a lower methane activation barrier or greater selectivity for a particular mechanism). Researchers have sought to automate the process of molecule discovery with the use of ML. For example, Aspuru-Guzik and collaborators have used artificial neural networks (ANNs) for the prediction of organic chemical reactions by building a system which, given a set of reagents and reactants, predicts the likely product.¹² Aspuru-Guzik's team have also developed an efficient exploration of chemical space that converts discrete representations to a multidimensional continuous representation, allowing one to automatically generate novel chemical structures by performing simple operations in latent space.¹³ Other approaches by this team used NNs to study the chemical space of Vaska-type iridium complexes that are competent for dihydrogen activation.¹⁴ Roitberg et al. have used ML methods to create an extensible neural network potential of organic molecules with the use of deep neural networks, having DFT accuracy with force field computational costs. The applicability of ML methods in science and computational chemistry has shown great versatility and potential. In our earlier studies, ML regressors were successfully used to find a correlation between ground state properties such as bond dissociation energies, global electrophilicity index, and reaction energies with the activation energies of methane activation by FLPs.¹¹

ML approaches to automate the exploration of chemical space have also been applied to *in situ* robotic experiments. Cronin and collaborators utilized Support Vector Machines, an ML algorithm, to develop a robot capable of exploring novel reactions by using spectroscopic tools and a combination of reactants. The machine can predict favorable reactions and create a feedback loop that allows the robot to synthesize the favored reactants.¹⁵ Our present study is in a sense a digital analogue of the robot presented by Cronin et al. A fixed database of different permutations of Lewis acids, Lewis bases, and their substituents will be tested to allow for a more efficient search of the favored chemical space to identify novel leads for FLP-based methane activators.

MOLECULE BUILDER

Generation of three-dimensional (3D) models of FLPs was achieved by developing a molecule builder that took as input a 4-bit feature vector, which represents a Lewis acid (LA), Lewis acid substituents (L_A), Lewis base (LB), and Lewis base substituents (L_B), assigning arbitrary numbers that represent atoms for LA/LB and chemical moieties for the substituents.

As a simple example, if ammonia borane was the target FLP, $LA = B$, $LB = N$, and $L_A = L_B = H$. As shown in Figure 1, the

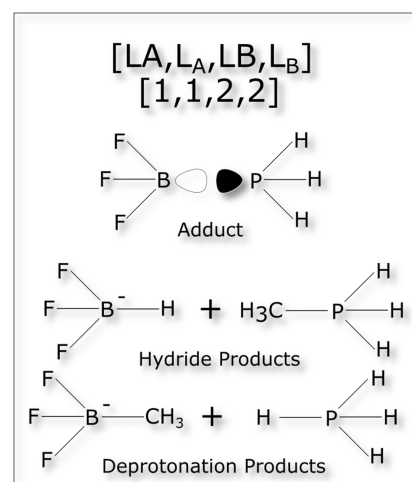


Figure 1. Input vector encoder. 4-bit input vector (Lewis acid (LA), Lewis acid substituent (L_A), Lewis base (LB), Lewis base substituent (L_B)) uniquely defines the FLP; for example, in the present case, $[1, 1, 2, 2]$ corresponds to F_3BF-PH_3 . The input vector also defines the two reactions of interest for FLP-mediated methane activation: hydride abstraction (top) and deprotonation (bottom). From these, the reaction energies (ΔE) are computed.

encoder takes the input vector, builds the FLP defined, and then quantum mechanically optimizes the Cartesian coordinates of the reactants and products of methane activation. Two different pathways are modeled herein, generated by hydride abstraction and deprotonation mechanisms, between methane and an FLP. In previous work, it was predicted that the deprotonation mechanism is favored for triel trihalides and pnictogen pentahalides.¹¹ The scope of this research is to not only use ML to efficiently explore chemical space but also to probe the chemical factors that favor these mechanisms.

The FLP generated by a specific 4-bit input vector is automatically built, input files are generated, and then these stationary points are modeled using Gaussian16.¹⁶ The automated molecule builder takes the four previously optimized subunits—LA, LB, L_A , L_B —combines them to create new FLPs, and then generates and models the FLP adduct, the products of hydridic methane activation, as well as the products generated by methane deprotonation, Figure 1. The molecule builder obtains the LA, LB, L_A , and L_B moieties from a fixed, curated structural library. Taking into consideration the results of our previous¹¹ and current study, for the present research, triels, pnictogens, and alkaline earth metals were chosen to form the basis of the Lewis acid (LA), while the Lewis bases (LB) were limited to nitrogen and phosphorus bases. The substituents were taken from a review, whereby different univalent substituents of varying Hammett constant (specifically σ_m) are reported.¹⁷ Methane activation reaction energies have been calculated at the $\omega B97X-D^{18}/def2-SVP^{19}$ level of theory, which was previously calibrated versus high-accuracy composite *ab initio* techniques. Geometries were optimized in a polar aprotic solvent, acetonitrile, utilizing the SMD model.²⁰ All electronic energies are reported in kcal mol⁻¹.

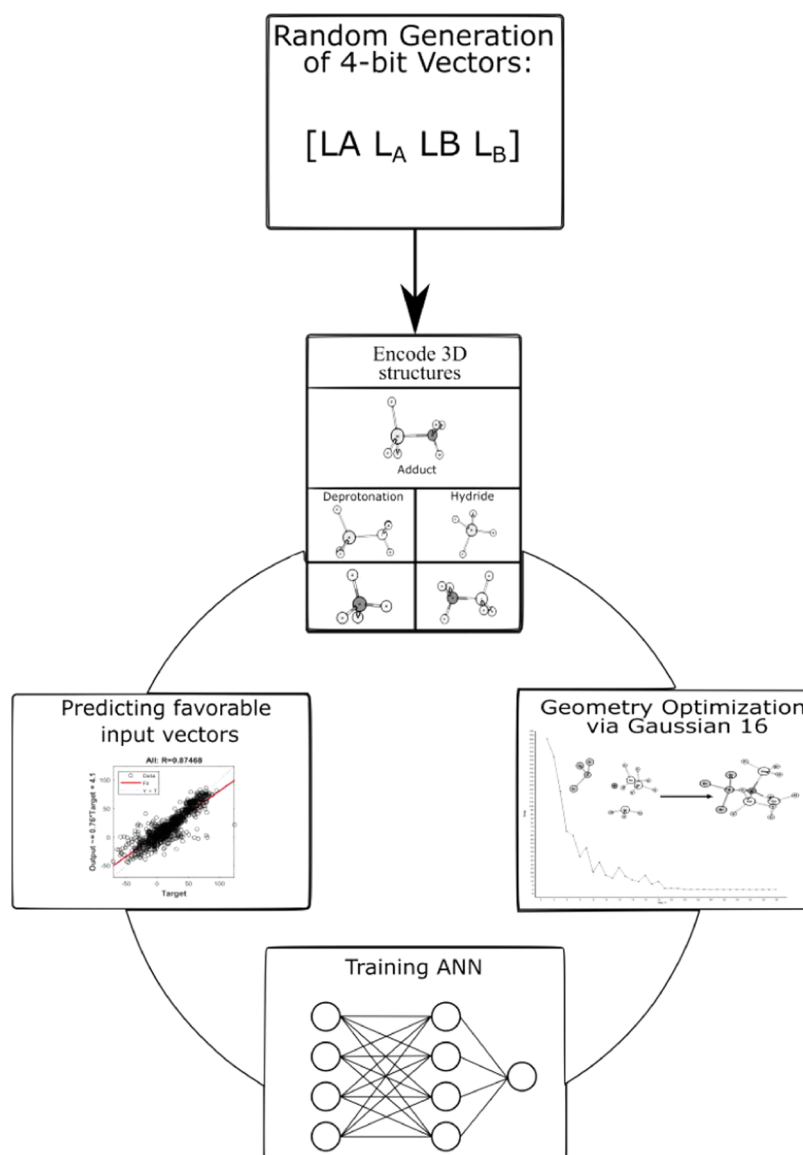


Figure 2. ML exploration protocol guides the modeling of FLP-based methane activators focused on identifying favorable chemical spaces. Starts with the generation of random input vectors, geometry is then optimized for both the adduct and the products to obtain the reactions energies, then these reaction energies and 4-bit feature vectors are fed into the ANN, finally the top 100 reactions that were predicted are fed into the cycle thus completing an inning.

■ EXPLORATION OF CHEMICAL SPACE PROTOCOL

The molecule builder allows one to automate the process of modeling molecular structures but must be guided by an algorithm that optimizes the search for promising methane activators or whatever the target reaction may be. ML has proven to be successful for nonlinear regression; by creating an ML surrogate model, we seek to expedite the exploration of very large chemical spaces by efficiently and effectively identifying chemical systems that give rise to favorable reactions.^{21–24} Paired with the automated molecule builder just described, it is possible to create a feedback loop that allows the search to focus on molecules that are predicted to be the most favored for the specific chemical application.

In 2, the exploration algorithm is depicted. The first step is to generate random input vectors that uniquely define an assortment of chemically diverse FLPs, which will then be modeled to determine their reaction energies with methane; these energies will be calculated with Gaussian16. The

electronic energies generated from the 4-bit input vectors will be automatically parsed from the G16 output, and these data are then fed to an ML algorithm. Due to the vast quantity of possible FLPs that could be explored, it was decided to only consider the electronic energies of the molecules and reactions, as doing additional vibrational frequency calculations would slow the process down. Also, once suitable promising leads are found *via* ML, one would further refine the best candidates at higher levels of theory, including effects such as solvent, dispersion, *etc.*

From the initial reaction energies of randomly selected FLPs (using a baseball analogy, this was termed inning #1), the ML algorithm predicts favored input vectors for a second “inning” of the search. These favored input vectors are determined by ranking the predicted reaction energies from more exothermic to less exothermic ($-\Delta E$ to $+\Delta E$) and selecting the top 100 exothermic candidates. These input vectors predicted by the ML algorithm are in turn fed back into the molecule encoder

and thus continue the search process in subsequent innings. The algorithm will terminate once the calculated and modeled molecules are the same or suitably close enough, *i.e.*, the search has converged. In this study, multiple innings were modeled to discover how ML-guided exploration benefits the examination of an exceptionally large chemical space. The goal is to assess how ML can identify promising portions of chemical space for emphasis, and unpromising areas for de-emphasis, without having to generate and assess all possible combinations, *i.e.*, an Edisonian search in normal parlance.

Sophisticated representations of molecules are often used in ML and cheminformatics, such as SMILES or molecular fingerprints. Although these representations have shown remarkable results,^{14,25–27} oftentimes they are represented by 1024- or 2048-bit vectors that require more sophisticated deep learning methodologies. Herein, it is shown that it is possible to use a simple 4-bit vector with a relatively small artificial neural network (ANN) to provide comparable results and identify novel candidates for FLP-mediated methane activation. This naïve representation of the molecular structure, known as one-hot encoding, has allowed us to have a chemical representation of our system based on chemical intuition of the chemistry of FLPs, assigning categorical variables to each chemical moiety. Although great generalizability is not expected, the current approach can be a great tool for simple surrogate models for specific chemical systems, where molecular fingerprints fail to provide meaningful regressions. The ANN used in this work comes from the Neural Network Toolbox from Matlab,²⁸ a 1 hidden layer, 10 hidden neurons with sigmoid activation function, and a linear output layer, was paired with Bayesian regularization, which is good for small or noisy data sets. The ANN was cross-validated to find the best hyperparameters. The training and validation split was set to 70/30%, and the neural network was retrained after every inning. Moreover, as discussed below, it is possible to use a first round of ML-guided analysis to both focus and diversify subsequent rounds of discovery.

MACHINE-LEARNED FORCE FIELDS

Machine learning (ML) has pervaded various facets of chemical research, integrating tasks from data analysis to the development of force fields for energy estimation. ML stands as a formidable tool, but its efficacy is contingent upon several factors. These include the necessity for accurate training data, exploration of a diverse parameter space, and meticulous selection of appropriate hyperparameters during model training. Machine-learned force fields (MLFF), particularly those exemplified by the work of Roitberg et al.,²⁹ have garnered widespread popularity, as they promise the accuracy of quantum mechanics with the speed of molecular mechanics. Their approach utilizes a descriptor-based neural network potential first introduced *via* Behler–Parrinello symmetry functions,³⁰ as demonstrated in the development of the ANI atomic neural network potential for organic systems. This MLFF has undergone training across various levels of theory, from ω B97X-D to CCSD(T), incorporating millions of generated molecules into its training data set. Noteworthy is the emphasis of ANI on a select group of atoms, C, H, O, N, P, and S, and halogens that serve as the focal point for organic chemistry. ANI has proven to be extremely competent in simulating organic molecule systems with high accuracy and speed. It is noteworthy that there is a noticeable absence of

similar efforts in the realm of inorganic chemistry and even less so in main group chemistry.

RESULTS AND DISCUSSION

Game 1. The first search of FLP space (*i.e.*, the protocol in Figure 2 is run until convergence and is termed a “Game”) gave interesting predictions and identified relevant issues in ML-directed chemical exploration. In this research, convergence was defined when the predicted favored structures are the same as the calculated structures by DFT. This allowed both further expanding and focusing a subsequent search in finding promising main group FLPs for methane activation. In this work, two independent games are presented with different considerations for the chemical space probe involving LA, LB, and their respective substituents. The first game consisted of combining triel/pnictogen Lewis acids and pnictogen/tetrel Lewis bases, which helped elucidate the underlying chemistry of FLP-mediated methane activation without a priori chemical intuition as well as the ML-pertinent specifics parameters that dictate the search. Table S1 in the Supporting Information presents the Lewis acids, Lewis bases, and their respective substituents. The 10 Lewis acid substituents (L_A) chosen are dependent on the Lewis acid chosen, *e.g.*, if LA is a triel or a pnictogen, the central atom will have a valency of three or five, respectively, thus dictating the number of univalent substituents attached to the central Lewis acid atom. The 13 substituents for the Lewis acids have been chosen to be both electron-donating and electron-withdrawing. For the Lewis bases, trivalent pnictogen atoms were chosen, also with varying electron demand on LB being imparted by the various substituents. In total, for Game 1, 4875 total possible FLPs are possible within the chemical space. For each angle, quantum chemical calculations must be performed and analyzed by the ANN.

Chemical intuition would suggest that Lewis acids in FLPs favor electron-withdrawing groups, while LB in FLPs will favor electron-donating groups. The aim of this study is to have the ANN determine such rules with no or limited previous chemical intuition, as this is the most likely scenario in the general case of catalyst discovery. One of the main challenges to automating the discovery of chemical reactions through computational chemistry is the abundance of mistakes in forming the initial structure that could lead to, for example, failed convergence of either the self-consistent field equations or geometry optimization routines. Earlier versions of the molecule builder employed here did not consider the steric overlap of atoms in the initial structure, which led to a higher percentage of failed calculations. For example, for the first optimization cycle (inning), out of 449 calculations, 369 calculations terminated successfully (82% success rate); failed structures were not taken into consideration by the ML algorithm. Manually fixing chemical structures would defeat the purpose of automation, in addition to being too time-consuming.

Each subsequent step calculates 100 new FLP methane activation reactions based on predictions from the ML surrogate model. In the first inning, 100 random 4-bit feature vectors were generated and fed into the algorithm. These vectors were used to predict FLPs with promising energetics for methane activation. The results from these predictions are then used to update the ML algorithm, which refines its predictions for the following innings. For the first game, only the methane deprotonation mechanism was taken into

consideration, as previous studies¹¹ concluded that it was thermodynamically favored versus hydride abstraction reactions. For the following game, both reaction mechanisms were taken into consideration. For game one, a Random Bit Forest (RBF) algorithm³¹ was also evaluated. This technique uses random forests, ANNs, and boosting for a more robust ML regressor. Interestingly, while RBFs proved to be a good regressor for the limited first game,³¹ they proved less capable for the larger second game. ANNs were also tested (Supporting Information) and proved to have less error and better R^2 for the larger, more diverse exploration of the FLP space for the second game.

From the four innings before convergence was achieved in the first game, the best FLP candidates for methane activation are presented in Table S2. These FLPs were found to have favorable (*i.e.*, negative or close to thermoneutral) reaction energies for methane deprotonation. These FLPs are thus potential candidates for further study as methane activators. An aggregate of 30 (30) promising FLPs were found in the optimization cycles—or innings—from among 664 molecules investigated in Game 1. The number of possible 4-bit combinations for the first game is 4875. The ML-directed optimization thus converged at only 13.6% of the total FLP population size!

In Figure 3, we can see the average reaction energy of all of the FLPs per optimization cycle (inning) in more limited game

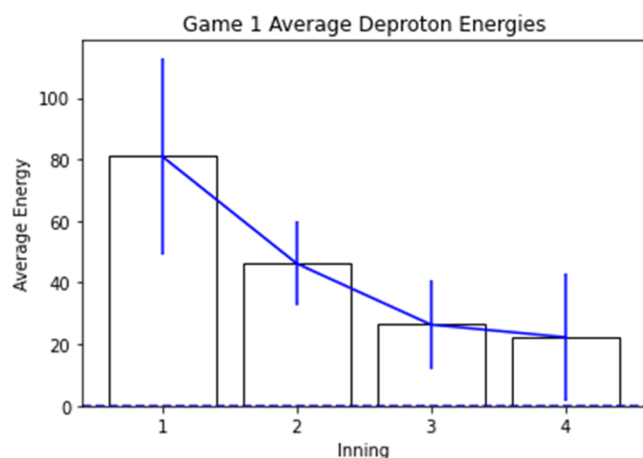


Figure 3. Game 1. Average methane activation reaction energy (kcal mol^{−1})—deprotonation mechanism—per inning with standard deviation depicted by the blue bar. After a random initial assessment of the FLP space (inning 1), the ML-guided search increasingly suggests more favorable reactions.

1. The first inning is the highest in energy as the algorithm initially chooses random input vectors to widely sample chemical space. As optimization progresses the ML algorithm is trained, and increasingly better FLP structures are suggested and evaluated. The following innings follow the trend of having lower average methane reaction energies, but the overall energies tended to be endothermic. The limited choice of substituents (L_A and L_B) in this first survey played a role in the overall chemical space being dominated by endothermic methane activation reactions.

The first instance of running the exploration with an ML surrogate model code provided valuable insight into its capabilities to predict chemical properties and efficiently map. If one analyzes the top FLP candidates from Game 1,

Figure S1, all have N as the Lewis base; all of the other Lewis base choices yielded endothermic reaction energies for methane deprotonation. In Figures S2 and S5, the frequency with which each constituent appears in each of the four features is graphed. The common trend is that in the first inning, there is a wider variety of LA/LB and their substituents being tried. For the Lewis acids, the variety of atoms being suggested and tested by the ML is varied, with all except N being suggested, and As being the most favored. This is consistent with our previous study,¹¹ as nitrogen is not expected to form stable hypervalent species, and it was reported that pentavalent pnictogens make for the most potent Lewis acids for methane activation. The Lewis acids sampled tended to be more diverse than the LA , but favored the heavier triels and pnictogens, specifically focusing on N and P. By the third inning of Game 1, the code was exclusively focusing on N-based Lewis bases. This outcome is also consistent with basic chemical principles, as nitrogen forms the strongest Lewis bases among the pnictogens sampled.

For the substituents of each, there is also unique insight obtained from the ML algorithm. The LA substituents (L_A) favored a more electron-withdrawing group while the Lewis base substituents (L_B) preferred alkyl groups which are electron-donating. These insights are eminently reasonable and were taken to catalyze a second optimization run (or game) with the intention of expanding and focusing the search to a more viable chemical space: focusing on N and P Lewis bases and expanding the Lewis acids and their substituents. Moreover, it is to be emphasized that these leads were arrived at without *a priori* guiding of the search, thus highlighting the robustness of the search protocol and the value of a multitiered ML search strategy for catalyst leads.

Game 2. With the important findings of Game 1, the chemical space explored was modified thusly: LA = Group 2, Group 13, or Group 15 elements and only N or P for LB . It was decided to investigate a wider variety of substituents, both L_A and L_B , with electron-donating properties for the Lewis base and electron-withdrawing properties for the Lewis acid. Substituents were chosen by consulting the review by Hansch et al.¹⁷ Also, given the definition of an FLP, substituents were chosen with increased steric profile.¹⁷ The complete list of the atoms and moieties used in modeling can be found in the Supporting Information. In Figure 4, the energy distribution of all calculated reactions is plotted for the deprotonation mechanism of methane activation; 10 innings were modeled for this study which resulted in 1985 successfully modeled reactions and 1181 reactions (59%) that are exothermic or close to thermoneutral ($\Delta E < 10$ kcal mol^{−1}). A first important observation for Game 2 is that the energy distribution of all modeled reactions skews toward exothermic a favored chemical space region for deprotonation mechanism.

The highest percentage of exothermic reactions modeled can be found in innings 2–5 due to the ML initially suggesting the best candidates; the following innings are modeling the next best reactions, refining already good predictions in favorable portions of FLP space. The FLP search algorithm was able to focus on the promising chemical space for the majority of the innings past the first random inning as can be seen in Figure 5. However, there is a decrease in exothermic reactions found in the later innings as “good” portions of FLP space are exhausted.

Figures S8 and S9 reveal similar analyses for the hydridic activation mechanism. The energy distribution for hydridic

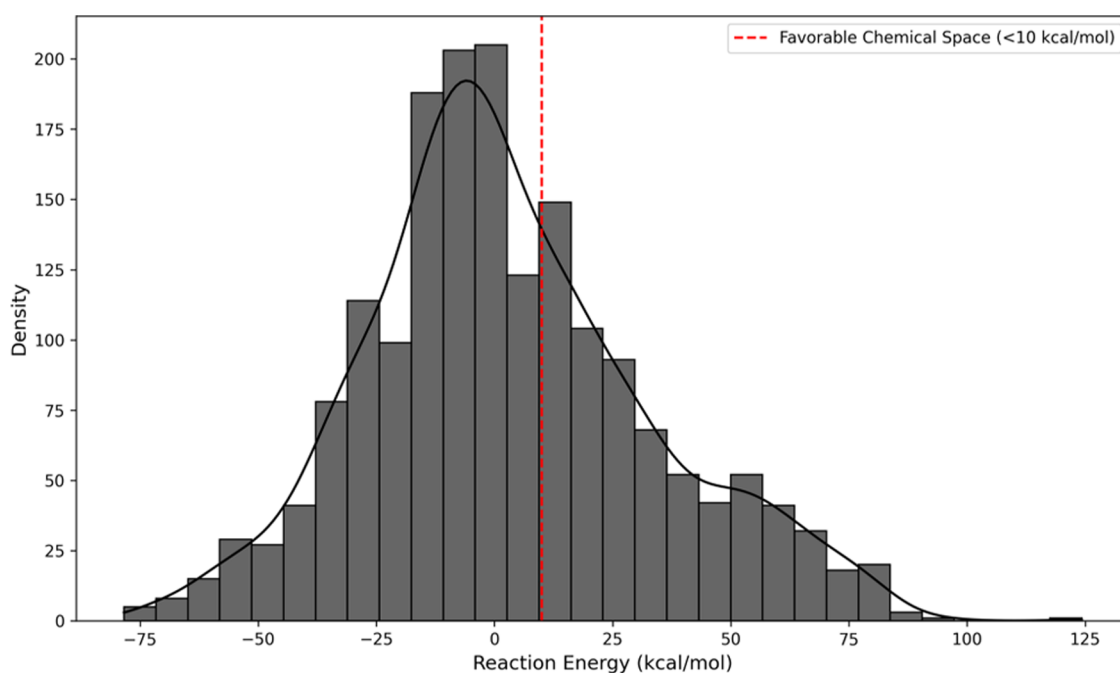


Figure 4. Reaction energy distribution (kcal mol^{-1}) of all modeled reactions for the deprotonation mechanism of methane activation utilizing the exploration protocol. Reaction energies computed at the $\omega\text{B97X-D/def2-TZVPP/SMD-MeCN}$ level of theory. The red vertical line marks 10 kcal/mol anything below is defined as favorable chemical space. The ML-guided algorithm is able to skew the search toward negative—or favorable—methane activation reactions.

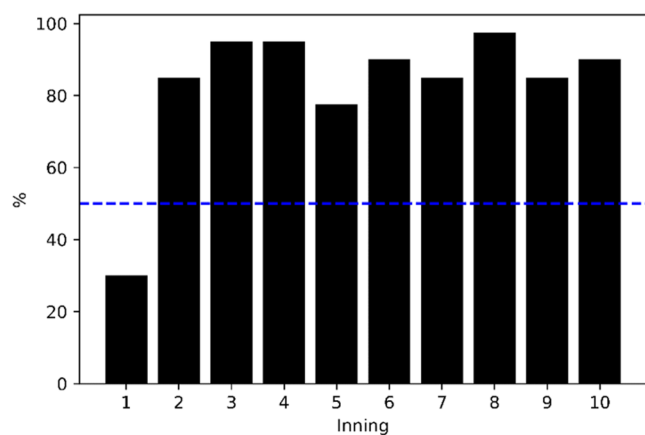


Figure 5. Percentage of methane deprotonation reactions modeled with $\Delta E < 10 \text{ kcal mol}^{-1}$ as a function of inning. Energies were calculated at the $\omega\text{B97X-D/def2-TZVPP/SMD-MeCN}$ level of theory. After a random sampling of FLPs in inning 1, the ML-guided algorithm quickly focused on favorable portions of chemical space.

methane activation is centered in favorable chemical space and most of the reactions targeted have $\Delta E < 10 \text{ kcal mol}^{-1}$. If the percentage of reactions that are $<10 \text{ kcal mol}^{-1}$ are plotted, similar observations to those just discussed for the deprotonation mechanism are seen, whereby thermodynamically favored reactions are in majority in the innings after the randomly selected FLPs of the first inning. Figure S10 plots the cumulative count of different methane activation energies as a function of inning; different innings are represented by different colored lines. There is a stark difference between the first inning (dark blue line in Figure S10) with randomly selected FLPs versus the subsequent innings that exploit ML to guide the identification of more favorable chemical spaces. Another interesting observation is the precipitous drop from

inning 1 to inning 2, whereupon the improvement as far as the ΔE metric is concerned is more moderated. The last two innings, green and yellow lines in Figure S10, have a lower count of favorable reaction energies. The algorithm quickly converged in the correct chemical space in the first few innings and identified the “best” leads. As the innings progress, the ML algorithm suggests the remaining next most favorable FLP-based methane activators. By the final innings, the favorable portions of chemical space have been depleted, thus providing a signal that the ML search has converged.

Comparing the first and last innings, it can be observed how the different energies are distributed, Figure 6. The first, random inning of ML optimization has a wider dispersion and is skewed to the right ($\Delta E > 0$); this indicates that the algorithm is not initially focused on a specific chemical space, which is desirable, as one desires to first search the available chemical space most broadly. Thus, a large proportion of the initially sampled reactions are endothermic for methane activation. In contrast, the last inning has a narrower peak to the left of thermoneutral ($\Delta E = 0$) indicating the ML has quickly and efficiently focused on the exothermic chemical space and has the majority of molecules being exothermic for methane deprotonation.

In summary, the results indicate that the machine learning code can efficiently and diversely explore FLPs, quickly narrowing the search to favorable portions of chemical space, i.e., those for which methane is exothermic and close to thermoneutral. Hence, numerous potential candidates were obtained for methane activation with the exploration algorithm, while avoiding the majority of infeasible and unfavorable FLPs. The total number of reactions that can be modeled for the selected chemical space of Game 2 is 7448 deprotonations and 7448 hydride abstractions. For the deprotonation mechanism, convergence was observed around Inning 8, which is 1601 molecules of the total 7448,

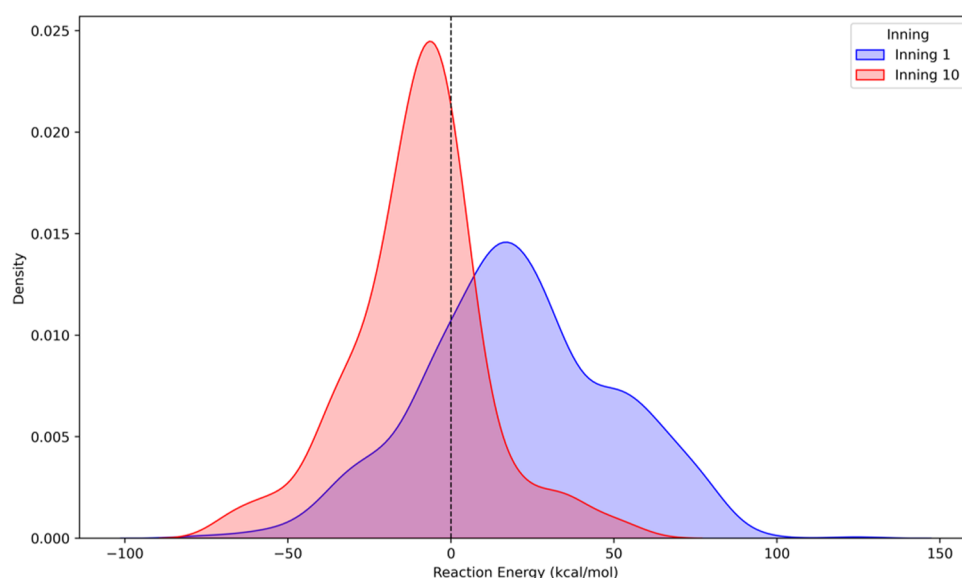


Figure 6. Kernel distribution estimation plot of Inning 1 vs Inning 10 for FLP-mediated methane deprotonation (kcal mol^{-1}); reaction energies (ΔE) calculated at the $\omega\text{B97X-D/def2-TZVPP/SMD-MeCN}$ level of theory. Note the ML algorithm guides the search from unfavorable ($\Delta E > 0$) to favorable ($\Delta E < 0$), and the distribution is narrower in the final inning of the search protocol.

corresponding to 21% of the chemical space; similar observations can be seen for the hydride mechanism; see the [Supporting Information](#).

Training and Testing Machine-Learned Force Fields from Our Data. The machine-learned force field was trained using Schnetpack³² for 1309 epochs. The training process was halted when no substantial improvement in validation was observed for 200 consecutive epochs. For further details, including the precise hyperparameter yaml file and the model itself, refer to the GitHub repository at <http://github.com/IgnacioMigliaro>. The data underwent meticulous preprocessing to eliminate any potential outliers in both energy and geometry optimization. The geometry optimization was assessed through manual inspection of each structure, ensuring proper convergence to a chemically sound structure. For energy elimination, the quartile range from Q2 to Q3 was employed, effectively removing extreme outliers. Despite the leftward skewness evident in the energy distribution, [Figure 7](#), these steps proved crucial in refining the data set for subsequent training of the MLFF.

The training mean squared error (MSE) for energy stood at 1.06 eV, while for forces, it registered at 0.13 eV/Å. Notably, the initial stages of training were marked by substantial noise, evident in the energy training graph, [Figure 8](#). However, as the training progressed, a stabilization phase ensued, leading to convergence. Similar observations were made during validation, where initial noise gradually diminished, reaching a point where no discernible improvement was observed in the training process. The MSE for validation energy settled at 1.25 eV, and for forces, it recorded a value of 1.08 eV/Å. Additional figures, including those depicting other aspects of the training process, can be found in the [Supporting Information](#). These results underscore the convergence and stabilization achieved during the training process despite the initial challenges posed by noise.

The test mean absolute error (MAE) for energy was recorded at 0.44 eV, while the MAE for forces reached 0.15 eV/Å. These values correspond to an average absolute error of 10.0 kcal/mol for energy and 3.47 kcal/mol/Å for forces. It is

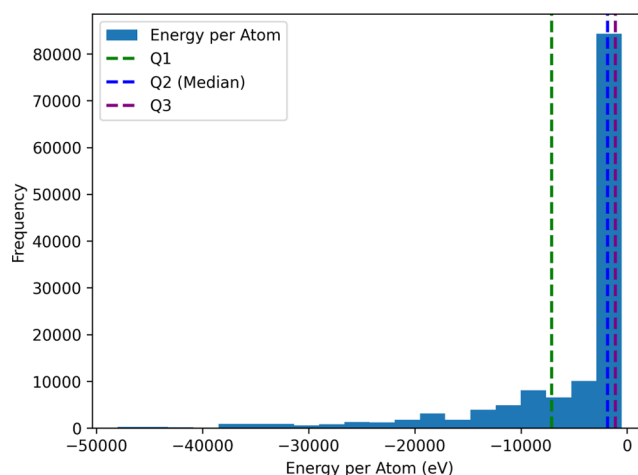


Figure 7. Histogram showing the distribution of energy per atom (eV) for our data generated using DFT, at the $\omega\text{B97X-D/def2-TZVPP/SMD-MeCN}$ level of theory.

noteworthy that the initial expectation was for forces to exhibit a considerably lower error, potentially several magnitudes smaller than achieved. The results indicate that our model excels in predicting energies but demonstrates limitations in accurately predicting forces. This discrepancy suggests that while the model may perform well in providing inference values for energy, its effectiveness in geometry optimizing structures is diminished.

Data Analysis of Methane Activation Reactions. The goal of this research is not to only find the perfect molecule or “needle” but rather to narrow the chemical space so that subsequent computational, synthetic, spectroscopic, and other efforts are focused on the further evaluation of the most promising catalyst candidates. Colloquially, the purpose of the ML algorithm is not to find a needle in a haystack, but rather to reduce the size of the haystack quickly and effectively. As such, after the ML exploration algorithm has run, there is not only a prioritized set of catalyst leads for further scrutiny—

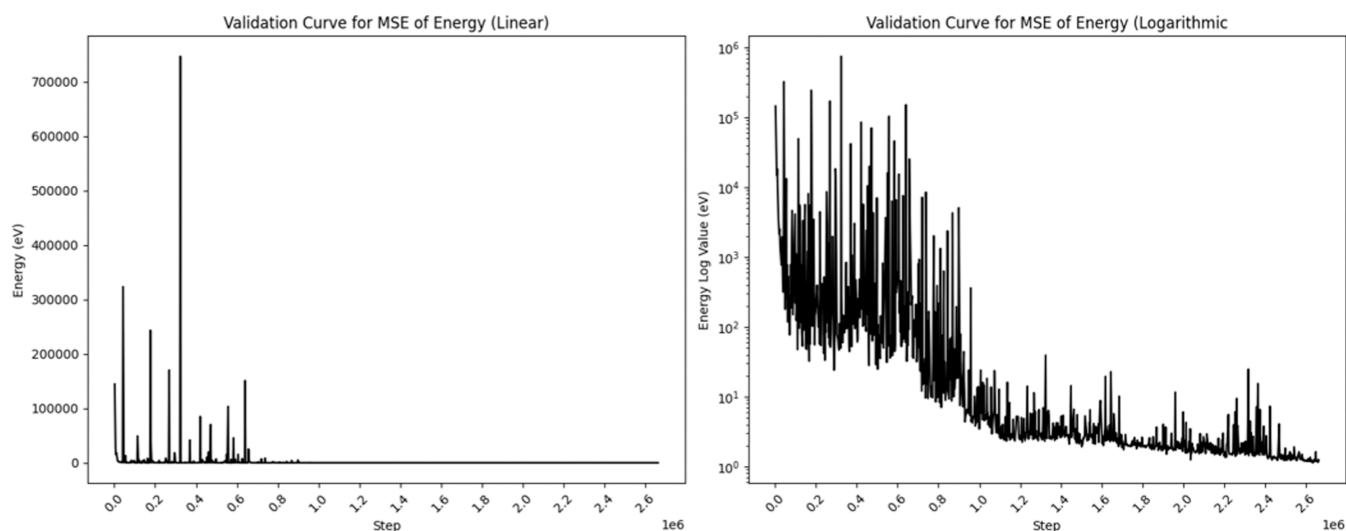


Figure 8. Validation curve for mean square error for energy (eV): linear scale (left) and logarithmic scale (right).

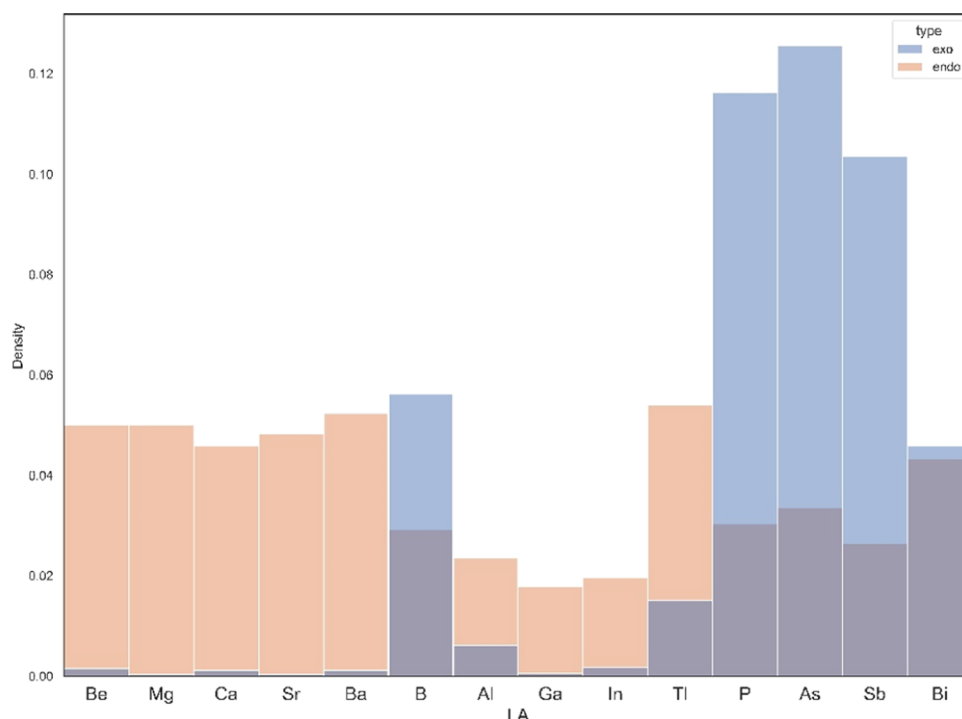


Figure 9. Histogram showing the frequency distribution of Lewis acids (LA) in deprotonation reactions for methane activation: endothermic (orange) and exothermic (blue) reactions. The ML-derived FLP database indicates that favorable chemical space ($\Delta E < 0$) is populated by B-, P-, As-, Sb-, and Bi-containing Lewis acids.

computational or experimental—but also an FLP database for methane activation *via* two different mechanisms. This database can then be mined to yield insight into the underlying reasons for why particular systems are favorable for the target reaction. Density functional theory calculations are, of course, very rich in information, which can then be data mined to find specific trends in the behavior of these reactions. Information such as orbital energies, highest occupied molecular orbital–lowest-unoccupied molecular orbital (HOMO–LUMO) gap, reaction energies, and geometric properties such as bond length, angles, steric volume, *etc.* can be analyzed to see how the descriptors impact, or do not impact, the specific target reaction.

The histogram with frequencies of Lewis acids suggested can be found in Figure 9; the different hues represent endothermic (orange) and exothermic (gray) reactions. A trend is apparent in that certain Lewis acids are more prevalent in exothermic reactions and *vice versa*. These more thermodynamically favored reactions are FLPs containing LA = B, P, As, Sb, or Bi. In previous research,¹¹ it was concluded that heavier LA, and heavier pnictogens, are favored for methane activation. This arrived at in a traditional Edisonian approach of setting up a search matrix in chemical space and methodically going through all combinations. Herein, even while examining a much wider chemical space, the ML search protocol independently arrived at the same conclusion in a much more efficient manner. Importantly, the ML search breaks with

preconceived notions one may have inferred from the literature that light Lewis acids such as B and Al are “better” for FLPs for small-molecule activation.^{6,33}

Figure 10 shows the frequency histogram for Lewis acid substituents (L_A) for the methane deprotonation mechanism.

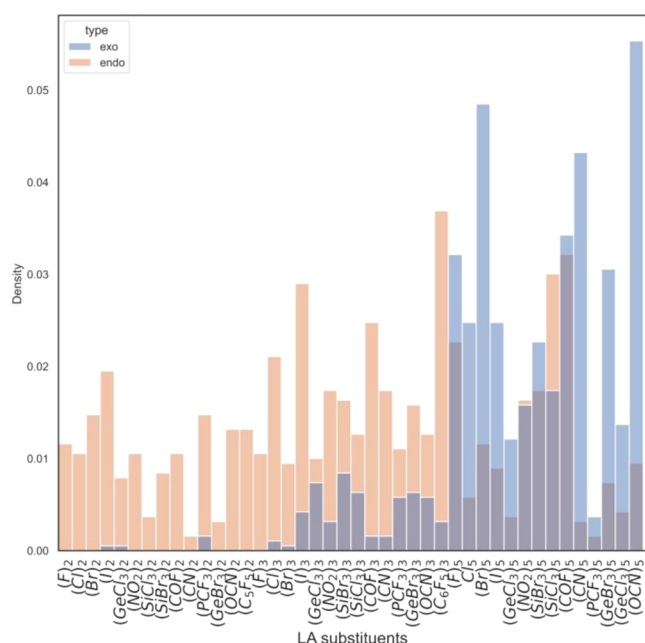


Figure 10. Histogram showing the frequency distribution of Lewis acid substituents (L_A) in the deprotonation mechanism for FLP-mediated methane activation. The favorable chemical space is dominated by the substituents for pentavalent (pnictogen) Lewis acids. Some halides are heavily favored such as F and Br. Other favored L_A species are CN, GeBr_3 , and OCN.

Although substituent choice is correlated with the type of Lewis acid (some Lewis acids are divalent, trivalent, or pentavalent), the plot provides insight into which substituents are favored for each valency. One could mine the data more deeply to see if the preferred LA choice correlates with different L_A or *vice versa*; similar considerations are expected to apply to L_B and L_B although with some caveats, as discussed below. The main take-home message extracted from the ML-guided search is that the majority of favored FLP-mediated methane activation reactions are found for Lewis acids with pentavalent substituents, which again mirrors previous conclusions from a more traditional search of FLP space that noted pnictogens (PnX_5) are favored over triels (TrX_3).¹¹ For triels, there is a less clear trend as to what substituents are favored, which may reflect the fact that the favorability of the target reactions is primarily dominated by the central Lewis acid, and the attached substituents are a secondary influencing factor. Such information extracted from the ML-derived database is valuable in guiding follow-on efforts as, for example, synthetic viability, availability of necessary reagents, or even sustainability or intellectual property considerations must also factor into the catalyst search protocol.

For substituents of pentavalent Lewis acids, the ML-derived database yields a few substituents that show a clear preference in terms of the objective, *viz.*, exothermic reactions with methane. Pnictogens favor F and Br over I and Cl. Other more atypical FLP substituents that are prevalent in exothermic reactions are CN, GeBr_3 , and OCN. In a more realistic scenario, one would want to give greater weight to first investigate catalyst leads with regard to additional metrics such as ease of synthesis, sustainability parameters, IP concerns, *etc.* Comparable results are found for the hydridic reaction mechanism, plotted in the Supporting Information (Figure S11).

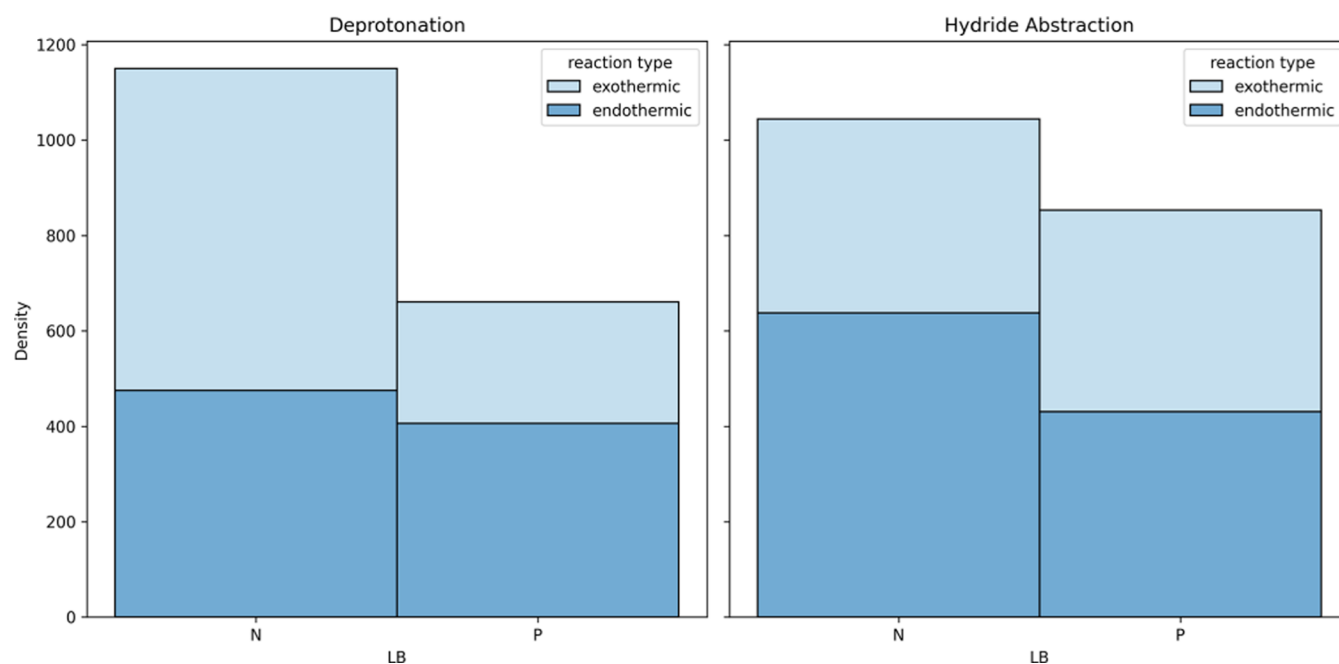


Figure 11. Frequency histogram of Lewis bases (L_B) for methane deprotonation (left) and hydride abstraction (right). Endothermic (light gray) and exothermic (dark blue) reactions are indicated. For the deprotonation mechanisms, N is favored over P for exothermic reactions, but for the hydridic mechanism, both N- and P-derived Lewis bases are favored equally.

The same frequency analysis was performed for the Lewis bases. It is found that for both the hydridic and deprotonation mechanisms, the LA preferred B as well as heavier pnictogens. For LB, however, there is a difference; there is a preference for N for favorable deprotonation mechanisms, but for the hydridic mechanism, N and P have the same frequency in exothermic reactions, Figure 11. Such information could be used when, for example, there is a need to design catalysts that prefer one mechanism over the others, allowing the ML protocol to enhance selectivity in addition to activity.

In Figure 12, the only unfavorable chemical space for L_B (the Lewis base substituents) regards the different amine

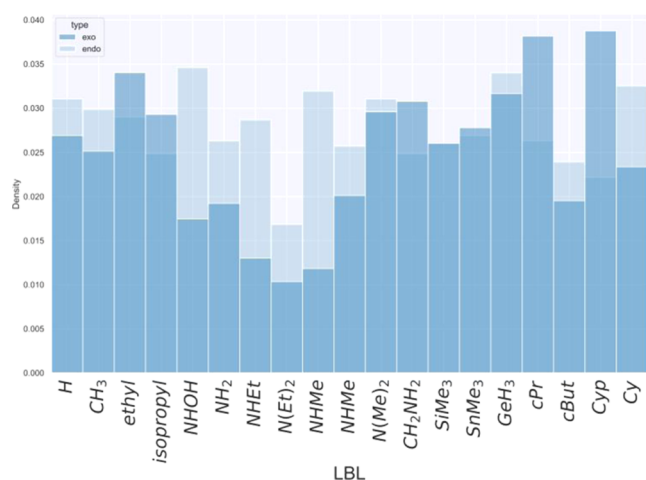


Figure 12. Histogram showing the frequency distribution of Lewis base substituents (L_B) for exothermic and endothermic methane activation reactions *via* the deprotonation mechanism. We see that both endothermic and exothermic reactions are dispersed among the substituents investigated, although amine substituents are less favored than the rest.

substituents ($L = \text{NH}_2, \text{NHEt}, \text{N(Et)}_2$). There is an increased preference for cyclic alkyls, such as cyclopropyl and cyclopentyl. Similar results are observed for the hydridic mechanism found (Figure S11). The information extracted from the ML-generated database on the more favorable FLP components helps narrow the desirable chemical space to a few specific substituents (L_A and L_B) and Lewis acids (LA). Such an

important deduction could help guide synthetic chemists interested in FLPs for methane activation. Also, this insight could be used to guide a third “game” whereby the search is focused with respect to three of the four FLP components, and a wider survey of LB is undertaken. Or perhaps, one might desire to expand LA to include other Lewis acids such as d-block metal complexes.

Not only can the components of the FLP be analyzed in terms of their impact on reaction energetics, but one can also data mine different QM-derived properties like the HOMO–LUMO gap or descriptors such as bond length, Hammett constants for the substituents buried volume, *etc.* In the Supporting Information (Figures S17 and S16), representative descriptors such as LA–LB bond length and L_A/L_B Hammett values are plotted and compared for exothermic and endothermic reactions. It is seen from these properties that there is no difference between the two reactivity subsets, indicating that they are not meaningful discriminating features vis-à-vis exothermic or endothermic methane activation reactions. One example of a property that showed potential in being a good determinant of methane activation favorability is the buried volume, as shown in Figure 13.

Buried volume is defined as the space occupied by a ligand in the first coordination sphere of a metal center; this calculation was done using the SambVca code.³⁴ Both exothermic and endothermic reactions display a right skew for a larger buried volume, i.e., sterically hindered L_A/L_B are preferred. Exothermic reactions have a higher density toward higher buried volume; the average buried volume for exothermic reactions is 74% as compared to 68% for the endothermic reactions. The kernel density estimation plot, Figure 13 (right), further exemplifies the difference between endothermic and exothermic reactions, showing exothermic reactions have a greater skew toward more sterically hindered substituents. This finding is congruent with FLP theory, as the FLP requires sterically hindered substituents to avoid forming a stable, unreactive Lewis acid–Lewis base adduct.

Electronic properties were also data mined, such as HOMO–LUMO gap and the global electrophilicity index of the FLPs; the latter is derived from frontier orbital energies. The frequency plot for the HOMO–LUMO gap of the FLP in Figure 14 shows that exothermic methane deprotonation reactions have FLP adducts with smaller HOMO–LUMO gaps in comparison to endothermic systems. The same results

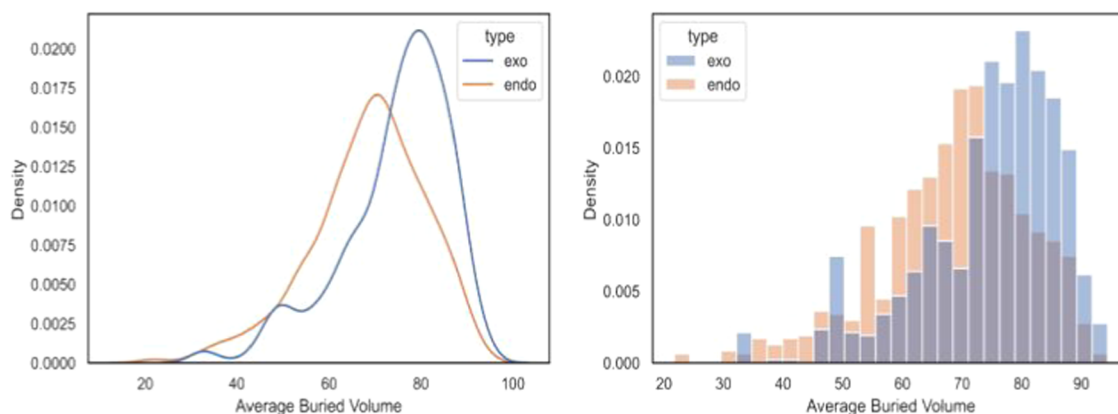


Figure 13. Histogram of substituent (L_A/L_B) buried volume for deprotonation mechanism for exothermic (blue) and endothermic (orange) reactions (left). Kernel distribution estimation plot of the buried volume (right). Exothermic reactions favored a higher buried volume than that of endothermic reactions.

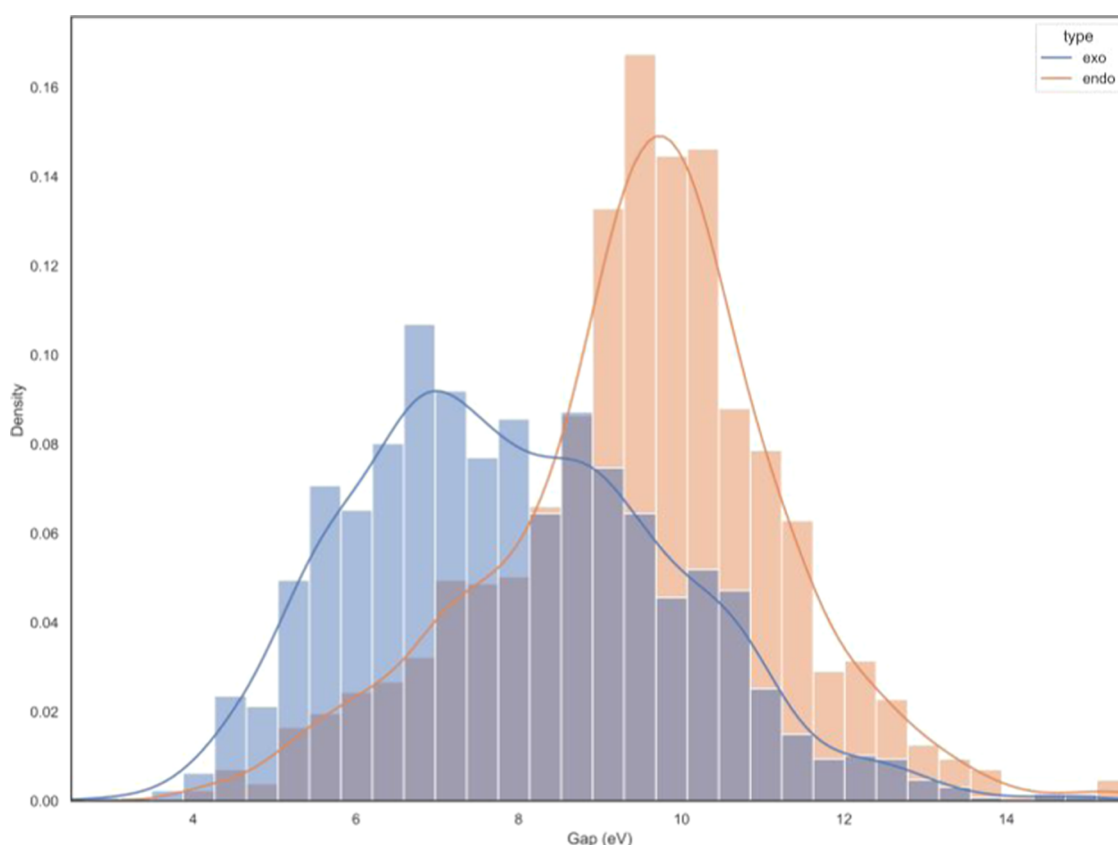


Figure 14. Frequency plot and histogram for the calculated FLP HOMO–LUMO gap (eV) for endothermic (orange) and exothermic (blue) deprotonation reactions for methane activation. Exothermic reactions have red-shifted HOMO–LUMO gaps.

are found for the hydridic activation mechanism, [Supporting Information](#).

Global electrophilicity index is a metric reported by Stephan et al.;³⁵ it is a measure of the electrophilicity of a compound derived from HOMO and LUMO energies.

This metric quantifies how much the molecule of interest desires to obtain an electron. It has been reported that for initial activation of methane, the complex must be a good acid, and base character is more important later in the reaction coordinate for CH activation.³⁶ In support of this proposal, in [Figure 15](#), it is observed that the endothermic deprotonation reactions are heavily focused among FLPs with lower GEIs, meaning that less electrophilic FLPs are less inclined to react with methane. Meanwhile, although more disperse, exothermic reactions skew to the right, indicating that promising catalysts leads have enhanced electrophilicity.

Features such as those discussed above are not only important to understand the nature of FLPs in methane activation but also allow for subsequent rounds of ML-inspired catalyst optimization, if needed, to create more meaningful features. For example, a PCA (principal component analysis) of features could potentially help train an ML classifier to determine if a reaction will be exothermic or endothermic just from ground state properties of the adduct, saving significant calculation time and allowing for high-throughput screening of millions of molecules through either a wider survey of the chemical components or the consideration of multiple optimization objectives.

Analysis of Machine-Learned Force Field Inference and Insights. To assess the capabilities of the machine-learned force field (MLFF) for structural optimization of main

group FLPs, inference was run on structures generated by the molecule builder described previously, which assembles FLPs using preoptimized ligand fragments, Lewis acids, and Lewis base ligands. Surprisingly, the optimized structures exhibited outward expansion rather than converging to chemically reasonable configurations. This outcome aligned with our observations during training, whereby forces converge less well compared to energies.

To further evaluate the newly derived MLFF's performance, a universal force field³⁷ was applied for initial optimization, achieving more chemically plausible FLP geometries. Subsequently, these optimized structures were subjected to MLFF for energy assessment. It is essential to note that the current MLFF is restricted to predicting energies for neutral-charged molecules. Consequently, simulations involving the reaction energies for our mechanism, which includes ionic products, are not feasible. As an alternative, we focused on predicting the formation energies of FLPs, aiming to obtain energy metrics comparable to those from DFT to provide meaningful comparisons. This approach allowed us to navigate the limitations of MLFF but still obtain valuable insights into its performance within the current chemical space.

Exploring the entirety of chemical space and leveraging the newly developed FLP molecule builder to generate all possible permutations, we created a comprehensive set of 7058 Frustrated Lewis Pairs (FLPs). The objective was to evaluate the formation energies of these pairs with the MLFF. During the analysis, extreme outliers in the formation energies were identified and subsequently removed, specifically those falling below -100 kcal/mol. These outliers were indicative of instabilities in the MLFF model, potentially stemming from

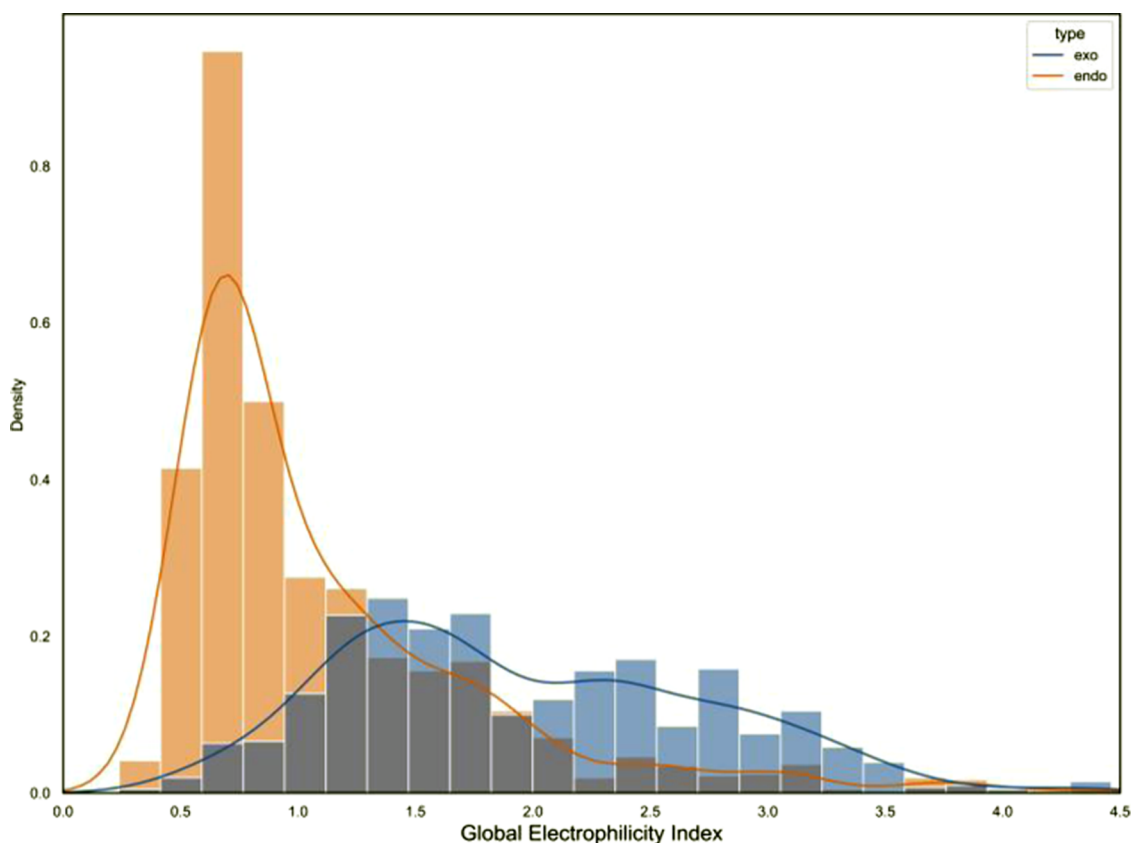


Figure 15. Histogram for Global Electrophilicity Index (GEI, in eV) for exothermic (blue) and endothermic (orange) reactions. More favorable methane activation reactions are characterized by a greater GEI, although the distribution is quite dispersed.

inadequacies in the training data. The decision to eliminate such outliers seeks to refine the data set and ensure a more reliable assessment of the FLP formation energies. This careful curation of data contributes to a more accurate understanding of the capabilities and limitations of the model in the context of FLP predictions. The MLFF has an inference time of less than a second per FLP, an incredible speed up as compared to quantum mechanics. Pairing classical FF with MLFF for energy evaluation thus provided an analysis within minutes. This significantly reduced computational work compared to hundreds of hours typically required in a DFT calculation.

In Figure 16, the average FLP formation energies are organized with respect to the Lewis acid constituent. There is a clear difference in Al-, In- and Tl-based Lewis acids yielding a more stable FLP. The DFT results pointed to the pnictogen group as better at activating methane, by which it may be deduced that a less stable, more strained FLP is needed to provide a more reactive adduct for methane activation.

SUMMARY AND CONCLUSIONS

A novel exploration protocol was developed and proven to be successful in finding exothermic reactions for the activation of methane through two different reaction mechanisms, hydridic activation and deprotonation. In contrast to our previous study,¹¹ which was limited to Group 13 trihalides and Group 15 pentahalides, herein are tested different groups, such as Group 2, Group 13, and Group 15 elements, with 42 different electron-withdrawing groups, leading to over 3000 reactions being modeled with state-of-the-art DFT methods. From the reaction energy distribution graphs, it can be concluded that the reactions being modeled are largely and quickly focused

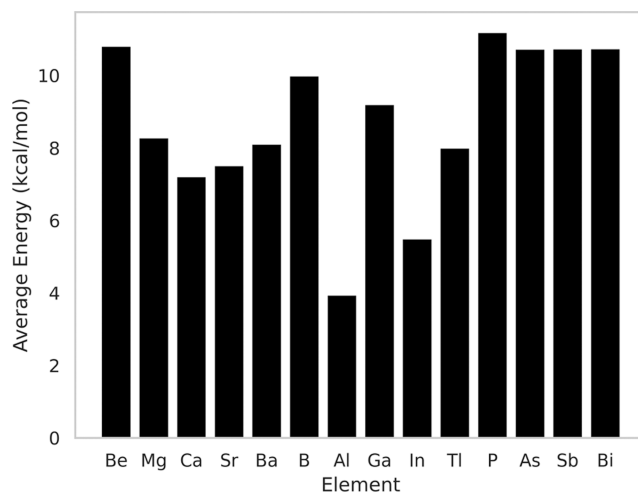


Figure 16. Average FLP formation energy for each Lewis acid calculated by MLFF (kcal/mol).

upon favorable portions of chemical space. The cumulative distribution graphs showed the difference between each optimization cycle (inning). The difference between random input vectors (inning 1) and vectors suggested by the algorithm is clear. The cumulative distribution graphs showed that the code converged in very few innings. Once the best molecules are modeled, they are knocked out of the suggestion pool, so the ML algorithm, which does not model duplicates, suggests the best candidates first, and then the code chooses the next best, next-next best, *etc.* at which point an increase in average reaction energy is observed. Comparing the kernel

distribution graphs between the first and last innings also shows that the algorithm is correctly focusing on favorable chemical space and has much less dispersion than the randomly generated vectors of the first inning. The ML protocol also generates a database of DFT-derived data that is rich in information and which thus allows mining of and chemical characteristics that are important in optimal catalyst design. For example, it was further confirmed without *a priori* guidance to the ML that FLPs with heavier pnictogens are favored for methane activation, as are bulkier substituents, cyclic alkyl substituents. Also, nitrogen bases are favored for the deprotonation mechanism, while N- and P-based Lewis bases make equal contributions to FLPs with favorable thermodynamics for hydridic methane activation. Electronic properties were also mined: lower HOMO–LUMO gaps in the adduct may be an important indicator of whether the reaction will be favored or not. Also, there are trends in the global electrophilicity index where more electrophilic molecules are favored for both reaction mechanisms studied, but the trends were more dispersed, suggesting such a descriptor may be of secondary significance in terms of efficient FLP-mediated C–H activation.

Although promising, additional work would be needed to fully leverage the ML protocol disclosed here. For example, one would wish to connect ground state descriptors with transition state properties such as activation energies—as discussed in ref 11—to further winnow or prioritize the candidate pool. Also, adding metrics, such as a synthetic feasibility score or a measure of the sustainability of the catalyst leads, would be desirable. However, the data provided here not only indicate that ML protocols can provide meaningful insight into the efficient searching of favorable chemical spaces of Frustrated Lewis Pairs for methane activation but highlight that ML can be exploited in a multitiered strategy to search catalyst space in a manner that is more focused and more diverse. Finally, the methods discussed here are general enough to be extensible to other reactions and reagents.

The machine-learning force field (MLFF) was a valuable tool for efficiently analyzing the formation energies of Frustrated Lewis Pairs (FLPs). The MLFF yielded a significant acceleration compared to quantum mechanics methods, with an inference time of under a second per structure, as opposed to the lengthy calculations associated with self-consistent field (SCF)-based quantum methods such as DFT, which take upward of a minute each on a high-performance computing platform. However, this enhanced speed comes at the expense of precision as the model exhibits instability in predicting molecular forces, hindering its ability to effectively relax FLP structures. In response to this limitation, it was necessary to resort to classical force fields, which, while less accurate than density functional theory (DFT), provided a means to obtain structurally reasonable configurations in a very short amount of time. Future research in the use of MFFs for the optimization of inorganic entities will need more careful evaluation of optimization strategies that yield consistent molecular forces.

This study serves as an initial exploration into the potential benefits of training MLFFs for main group compounds, hinting at their role as direct rivals to DFT in the future. Notably, as of the current writing, there are, to our knowledge, no MLFFs capable of simulating ions. The closest approximation is Chgnet,³⁸ which utilizes magnetic moments to provide insights into ions. However, this approach was not feasible for our force field, as specific data could not be incorporated into our

calculations. This work highlights the ongoing advancements and challenges in the field of MLFFs, emphasizing the need for further development to address their limitations and enhance their applicability in diverse chemical scenarios.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.4c00354>.

Detailed combinations of Lewis acids, Lewis bases, and their substituents for FLP deprotonation and hydride mechanisms for methane activation; frequency distributions of reaction components in favorable chemical spaces; reaction energy distributions, and data on the performance and accuracy of the ML-guided exploration protocol (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Thomas R. Cundari – Department of Chemistry, Center of Advanced Scientific Computing and Modeling, University of North Texas, Denton, Texas 76203, United States;

orcid.org/0000-0003-1822-6473;

Email: ignaciomigliaro@my.unt.edu

Author

Ignacio Migliaro – Department of Chemistry, Center of Advanced Scientific Computing and Modeling, University of North Texas, Denton, Texas 76203, United States;

orcid.org/0000-0003-2678-7713

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jctc.4c00354>

Author Contributions

This manuscript was written through contributions from all authors.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Martin Migliaro for his useful conversations and discussions in machine learning and statistics. This research was supported by NSF grant CHE-1953547.

■ REFERENCES

- (1) Welch, G. C.; San Juan, R. R.; Masuda, J. D.; Stephan, D. W. Reversible, Metal-Free Hydrogen Activation. *Science* **2006**, 314 (5802), 1124–1126.
- (2) Dureen, M. A.; Stephan, D. W. Terminal Alkyne Activation by Frustrated and Classical Lewis Acid/Phosphine Pairs. *J. Am. Chem. Soc.* **2009**, 131 (24), 8396–8397.
- (3) Ménard, G.; Stephan, D. W. C–H Activation of Isobutylene Using Frustrated Lewis Pairs: Aluminum and Boron σ -Allyl Complexes. *Angew. Chem., Int. Ed.* **2012**, 51 (18), 4409–4412.
- (4) Stephan, D. W. Frustrated Lewis Pairs: From Concept to Catalysis. *Acc. Chem. Res.* **2015**, 48 (2), 306–316.
- (5) Stephan, D. W.; Erker, G. Frustrated Lewis Pair Chemistry of Carbon, Nitrogen and Sulfur Oxides. *Chem. Sci.* **2014**, 5 (7), 2625–2641.
- (6) Stephan, D. W. The Broadening Reach of Frustrated Lewis Pair Chemistry. *Science* **2016**, 354 (6317), No. aaf7229.

- (7) Neu, R. C.; Otten, E.; Lough, A.; Stephan, D. W. The Synthesis and Exchange Chemistry of Frustrated Lewis Pair-Nitrous Oxide Complexes. *Chem. Sci.* **2011**, *2* (1), 170–176.
- (8) Stephan, D. W. A Perspective on Frustrated Lewis Pairs. *J. Am. Chem. Soc.* **2015**, *137*, 10018–10032.
- (9) Ménard, G.; Stephan, D. W. Room Temperature Reduction of CO₂ to Methanol by Al-Based Frustrated Lewis Pairs and Ammonia Borane. *J. Am. Chem. Soc.* **2010**, *132* (6), 1796–1797.
- (10) Tang, P.; Zhu, Q.; Wu, Z.; Ma, D. Methane Activation: The Past and Future. *Energy Environ. Sci.* **2014**, *7* (8), 2580–2591.
- (11) Migliaro, I.; Cundari, T. R. A Density Functional Study of Methane Activation by Frustrated Lewis Pairs with Group 13 Trihalides and Group 15 Pentahalides and a Machine Learning Analysis of Their Barrier Heights. *J. Chem. Inf. Model.* **2020**, *60*, 4958–4966.
- (12) Wei, J. N.; Duvenaud, D.; Aspuru-Guzik, A. Neural Networks for the Prediction of Organic Chemistry Reactions. *ACS Cent. Sci.* **2016**, *2* (10), 725–732.
- (13) Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **2018**, *4* (2), 268–276.
- (14) Friederich, P.; Dos Passos Gomes, G.; De Bin, R.; Aspuru-Guzik, A.; Balcells, D. Machine Learning Dihydrogen Activation in the Chemical Space Surrounding Vaska's Complex. *Chem. Sci.* **2020**, *11* (18), 4584–4601.
- (15) Donina, L.; Dragone, V.; Long, D.; Cronin, L. Machine Learning to Search for New Reactivity. *Nature* **2018**, *559*, 377–381.
- (16) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, F. D. et al. *Gaussian 09*; Gaussian, Inc.: Wallingford CT, 2009.
- (17) Hansch, C.; Leo, A.; Taft, R. W. A Survey of Hammett Substituent Constants and Resonance and Field Parameters. *Chem. Rev.* **1991**, *91* (2), 165–195.
- (18) Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2012**, *32*, 174–182.
- (19) Hellweg, A.; Rappoport, D. Development of New Auxiliary Basis Functions of the Karlsruhe Segmented Contracted Basis Sets Including Diffuse Basis Functions (Def2-SVPD, Def2-TZVPPD, and Def2-QVPPD) for RI-MP2 and RI-CC Calculations. *Phys. Chem. Chem. Phys.* **2015**, *17* (2), 1010–1017.
- (20) Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic Interaction of a Solute with a Continuum. A Direct Utilization of AB Initio Molecular Potentials for the Prediction of Solvent Effects. *Chem. Phys.* **1981**, *55* (1), 117–129.
- (21) Ayodele, T. O. Types of Machine Learning Algorithms. In *New Advances in Machine Learning*; IntechOpen, 2010; pp 19–48.
- (22) Bishop, C. M. *Pattern Recognition and Machine Learning* 2006.
- (23) Hilario, M.; Kalousis, A.; Müller, M.; Pellegrini, C. Machine Learning Approaches to Lung Cancer. *Proteomics* **2003**, *3*, 1716–1719.
- (24) Dral, P. O. Quantum Chemistry in the Age of Machine Learning. *J. Phys. Chem. Lett.* **2020**, *11* (6), 2336–2347.
- (25) Pattanaik, L.; Coley, C. W. Molecular Representation: Going Long on Fingerprints. *Chem* **2020**, *6* (8), 1204–1207.
- (26) Scott, I. M.; Vermeer, C. P.; Liakata, M.; Corol, D. I.; Ward, J. L.; Lin, W.; Johnson, H. E.; Whitehead, L.; Kular, B.; Baker, J. M.; Walsh, S.; Dave, A.; Larson, T. R.; Graham, I. A.; Wang, T. L.; King, R. D.; Draper, J.; Beale, M. H. Enhancement of Plant Metabolite Fingerprinting By Machine Learning. *Plant Physiol.* **2010**, *153*, 1506–1520.
- (27) Capecchi, A.; Probst, D.; Reymond, J. L. One Molecular Fingerprint to Rule Them All: Drugs, Biomolecules, and the Metabolome. *J. Cheminf.* **2020**, *12* (1), No. 43.
- (28) Mathworks Matlab Neural Fitting Toolbox 2020.
- (29) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8* (4), 3192–3203.
- (30) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98* (14), No. 146401.
- (31) Wang, Y.; Li, Y.; Pu, W.; Wen, K.; Shugart, Y. Y.; Xiong, M.; Jin, L. Random Bits Forest: A Strong Classifier/Regressor for Big Data. *Sci. Rep.* **2016**, *6*, No. 30086.
- (32) Schütt, K. T.; Kessel, P.; Gastegger, M.; Nicoli, K. A.; Tkatchenko, A.; Müller, K. R. SchNetPack: A Deep Learning Toolbox for Atomistic Systems. *J. Chem. Theory Comput.* **2019**, *15* (1), 448–455.
- (33) Flynn, S. R.; Wass, D. F. Transition Metal Frustrated Lewis Pairs. *ACS Catal.* **2013**, *3* (11), 2574–2581.
- (34) Poater, A.; Cosenza, B.; Correa, A.; Giudice, S.; Ragone, F.; Scarano, V.; Cavallo, L. SambVca: A Web Application for the Calculation of the Buried Volume of N-Heterocyclic Carbene Ligands. *Eur. J. Inorg. Chem.* **2009**, *2009*, 1759–1766.
- (35) Jupp, A. R.; Johnstone, T. C.; Stephan, D. W. The Global Electrophilicity Index as a Metric for Lewis Acidity. *Dalton Trans.* **2018**, *47* (20), 7029–7035.
- (36) Cundari, T. R. Calculation of a methane carbon-hydrogen oxidative addition trajectory: comparison to experiment and methane activation by high-valent complexes. *J. Am. Chem. Soc.* **1994**, *116*, 340–347.
- (37) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114* (25), 10024–10035.
- (38) Deng, B.; Zhong, P.; Jun, K. J.; Riebesell, J.; Han, K.; Bartel, C. J.; Ceder, G. CHGNet as a Pretrained Universal Neural Network Potential for Charge-Informed Atomistic Modelling. *Nat. Mach. Intell.* **2023**, *5* (9), 1031–1041.