Ligand Many-Body Expansion as a General Approach for Accelerating Transition Metal Complex Discovery

Daniel B. K. Chu[1], David A. González-Narváez[1], Ralf Meyer[1], Aditya Nandy[1,2], and Heather J. Kulik[1,2,*]

[1]*Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139,* USA

[2]*Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139,* USA

ABSTRACT: Methods that accelerate the evaluation of molecular properties are essential for chemical discovery. While some degree of ligand additivity has been established for transition metal complexes, it is underutilized in asymmetric complexes, such as the square pyramidal coordination geometries highly relevant to catalysis. To develop predictive methods beyond simple additivity, we apply a many-body expansion to octahedral and square pyramidal complexes and introduce a correction based on adjacent ligands (i.e., the *cis* interaction model, or *cis* model). We first test the *cis* model on adiabatic spin-splitting energies of octahedral Fe(II) complexes, predicting DFT-calculated values of unseen binary complexes to within an average of 1.4 kcal/mol. We next show that the *cis* model infers both DFT- and CCSD(T)-calculated model catalytic reaction energies to within 1 kcal/mol on average. The *cis* model predicts low-symmetry complexes with reaction energies outside the range of binary complex reaction energies. We observe that *trans* interactions are unnecessary for most monodentate systems but can be important for some combinations of ligands, such as complexes containing a mixture of bidentate and monodentate ligands. Finally, we demonstrate that the *cis* model may be combined with Δ-learning to predict CCSD(T) reaction energies from exhaustively calculated DFT reaction energies and the same fraction of CCSD(T) reaction energies needed for the *cis* model, achieving around 30% of the error from using the CCSD(T) reaction energies in the *cis* model alone.

1

1. Introduction.

Transition-metal complexes (TMCs) present unique opportunities and challenges for applications across many fields, including medicine,[1-3] photonics,[4-6] and catalysis.[7-9] However, the design space of TMCs is quite vast and underexplored. On the order of 10,000 ligands have been structurally characterized in octahedral TMCs,[10] meaning more than 10^{21} theoretical combinations of ligands could be used to generate novel octahedral TMCs. Nevertheless, the number of TMCs that have been characterized is on the order of 100,000.[11] Therefore, methods which can identify key TMCs from a large chemical space, either directly as leads or those that could be leveraged to infer information about many other TMCs, are invaluable.

While virtual high-throughput screening (VHTS) with density functional theory (DFT) and machine learning (ML) has greatly accelerated the *in silico* design of molecules[12-16] and materials,[16-21] discovery of molecular catalysts remains an open area of research,[11,22,23] including for small-molecule activation, such as in the production of hydrogen,[24-26] ammonia,[27-30] and methanol.[31-33] Researchers have carried out extensive computational investigations into ligand effects on catalytic TMCs in various reactions.[34-40] Indeed, both computational[41-43] and experimental[44-46] studies aim to build libraries of ligands[10,47] and understand how they influence the properties of derived TMCs.[48-50] While a number of automated methods exist for ligand design,[43,51-55] and ML can help target design objectives for the derived TMCs,[54-57] it is still intractable to perform explicit calculations for the entirety of the TMC design space.

While DFT is widely employed in large-scale screening, DFT calculations with open-shell TMCs are less precise and more time-intensive than DFT calculations on organic, closed-shell molecules, making them intractable for fully exhaustive exploration of transition metal chemical

space.[58-60] Some improvement in accuracy can be afforded by carefully calibrated[61] black-box wavefunction theory (WFT) methods such as DLPNO-CCSD(T),[62] but most methods are too demanding for large-scale screening. Strategies to infer energetics of many TMCs from a few data points are necessary to mitigate cost limitations. For example, these points can be used to develop Δ-ML[63] or semi-empirical corrections[64,65], but it is nontrivial to determine which TMCs will provide the most information and where ML predictions are erroneous.[66,67] Thus, it is attractive to develop physical models for relating the properties of TMCs and to identify a minimal set of TMC properties that can be used to predict properties of the full space with greater accuracy.

In organic chemistry, advanced group additivity theories, including those that have corrections for functional groups and ring strain,[68] have seen resounding success in predicting thermochemical properties to within 1 kcal/mol.[68,69] In organometallic chemistry, additivity models have been considered in the context of ligand-field theory (LFT) and the contribution of each ligand to the field that the central metal ion experiences.[70] This has typically been applied in its simplest form as homoleptic averaging, which estimates the value of a property for a heteroleptic complex as the stoichiometric average of the properties of corresponding homoleptic TMCs of constituent ligands. Homoleptic averaging has been applied to redox potentials,[71-73] absorption spectra,[74,75] and other properties,[76-78] but it can have errors exceeding 10 kcal/mol.[77] This motivates more complex LFT models, e.g., by considering the additivity of orbital energies.[79,80] One drawback of this more complex LFT is that it requires new equations to be derived for every symmetry class.[70,81,82] Thus, models that can be applied across symmetry classes but with the accuracy that can be achieved only via considering more complex interactions are desirable.

3

Because homoleptic averaging fails for many systems, we desire ligand additivity relationships that incorporate information from heteroleptic TMCs. Beyond additivity, there are also LFTs that include non-additive contributions, i.e., those which cannot be separated into individual ligand contributions.[70,83] We recently presented a model that incorporates heteroleptic information for predictions on low-symmetry octahedral TMCs using information from higher-symmetry octahedral TMCs comprising the same ligand types.[77] The previously presented model only applies to octahedral TMCs, and thus it is beneficial to develop a framework that may be applied to any geometry individually and potentially across multiple geometries. The many-body expansion (MBE), which represents the property of a system via contributions of interacting subsystems with increasing complexity, represents such a framework. MBE is ubiquitous in theoretical chemistry, underpinning many force fields,[84-89] fragmentation/embedding methods,[90-96] and correlated wavefunction theory.[97-100] The MBE has also been applied for estimating metal–ligand bonding by considering subsystems of up to three heavy atoms at a time.[101,102]

In this work, we consider entire ligands as the smallest subsystems within the MBE to derive ligand additivity relationships that describe both the additive and non-additive field contributions to some degree. We apply this framework to TMCs with octahedral coordination environments and square pyramidal asymmetric TMCs, and we provide guidelines for accurate screening of materials and catalysts with minimal data. Finally, we show that the models are highly effective in predicting WFT energetics based on learning the difference between WFT and DFT.

2. Theory

2a. Many-body expansion for ligand additivity

We consider several schemes for estimating the properties of a heteroleptic transition metal complex from combinations of representative complexes. In the simplest scheme, homoleptic averaging estimates a property, Q, of a heteroleptic complex as a stoichiometric weighting of the properties of the corresponding homoleptic (HO) complexes, i.e.

$$Q \approx \frac{1}{n} \sum_{i=1}^{n} n_i Q_i^{\text{HO}} \tag{1}$$

where n_i is the count of ligand type i present in the structure, and n is the coordination number of the geometry (e.g., six for octahedral).

Applying the MBE to ligands, a general quantitative property, Q, of a TMC may be divided into interaction terms, $q^{(l)}$, which involve l ligands. We truncate the MBE at order k as

$$Q \approx Q^{(k)} = \sum_{l=0}^{k} q^{(l)} \tag{2}$$

such that $Q^{(n)}$ is exact for a TMC with coordination number n. Each $q^{(l)}$ may be further decomposed by the identities of the ligands involved and the symmetry of each interaction (e.g., *cis* and *trans* interactions for $q^{(2)}$ in octahedral TMCs). For a given metal center, a TMC may be uniquely defined via the symmetry and ligand identities of the $q^{(n)}$ term, so we may assume that each term depends only upon the identity of the ligands without loss of generality. That is, Jahn–Teller distortion and other asymmetries do not require any special treatment in this formulation. At a given truncation order, we may also select which interactions to include in the approximation. Thus, eq. (2) provides a robust framework to obtain models for ligand additivity that are adjustable in complexity and accuracy.

We apply this ligand MBE to mononuclear octahedral complexes to derive simple models for ligand additivity relationships. The coordination sites in an octahedral geometry are symmetric,

so the first-order term is simply the sum of each ligand's additive contribution, q_i, which is equivalent for each ligand of the same type, L_i. Thus,

$$Q^{(1)} = q^{(0)} + \sum_{i=1}^{N} n_i q_i \qquad (3)$$

where n_i is the count of L_i present in the structure, and N is the number of unique ligand types considered. At first order, the ligand MBE for octahedral complexes is equivalent to homoleptic averaging upon using the homoleptic properties to solve for each q_i. At second order, there are *cis* and *trans* interactions. We expect the *cis* interactions to dominate the *trans* interactions because the ligands involved in *cis* interactions are closer together in space. Additionally, there are 12 *cis* interactions in an octahedral complex but only three *trans* interactions with at least twice as many *cis* interactions between differing ligand types in each binary configuration (Supporting Information Figure S1). For these reasons, we retain only the *cis* interactions to obtain

$$Q^{(2)} \approx Q_c^{(2)} \equiv q^{(0)} + \sum_{i=1}^{N} n_i q_i + \sum_{i=1}^{N} \sum_{j=i}^{N} n_{ij,c} q_{ij,c} \qquad (4)$$

where $q_{ij,c}$ is the contribution due to ligands L_i and L_j in a *cis* orientation, and $n_{ij,c}$ is the count of interactions of this type. The *cis* interaction model, or *cis* model, defined by eq. (4) is a linear model where the q are unknown parameters, and the n are explanatory variables, or the variables which may adjusted via the TMC identity to elicit changes in Q ($n_{ij,c}$ for binary octahedral complexes are shown in Figure 1). The $q_{ij,c}$ values describe ligand–ligand interactions and provide a means to quantify the non-additive contribution to ligand additivity relationships.[70] Including *trans* and higher-order interactions likewise leads to linear models (Supporting Information Text S1).
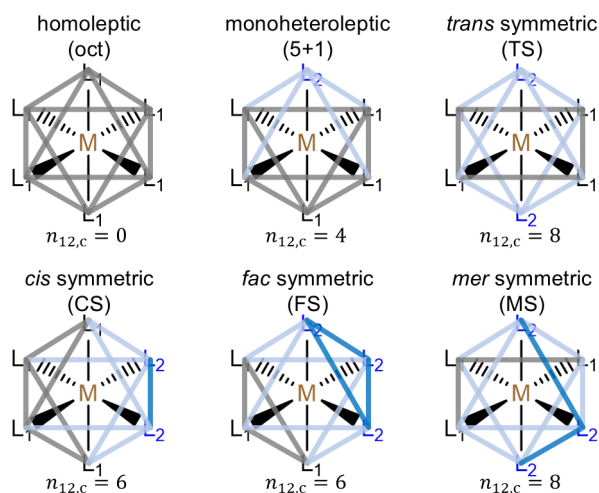
Figure 1. Symmetry classes for octahedral transition metal complexes with up to two unique ligands, L_1 and L_2, are labeled by their name above and corresponding count of the L_1–L_2 interactions ($n_{12,c}$) below. The L_1–L_1, L_1–L_2, and L_2–L_2 *cis* interactions are respectively indicated by gray, light blue, and dark blue lines.

2b. Considerations for application of the ligand MBE models

The ligand MBE results in linear models with highly structured explanatory variables, i.e., that take on discrete integer values, making them amenable to well-studied linear regression methods and applicable to low-symmetry complexes. Due to the finite number of possible TMCs with a given set of ligands, the explanatory variables are not necessarily linearly independent. However, it is sufficient to use a least-squares solution of minimal norm to deal with dependencies between explanatory variables because they are always integers and not subject to measurement error. Namely, the unknown parameters in eq. (4) are at most binary (i.e., involve two ligand types) and may be determined by measurements of Q for the binary complexes (Figure 1). The q parameters evaluated in this way still apply to ternary (i.e., three ligand types) and lower-symmetry complexes. Additionally, eq. (4) applies to complexes that belong to special cases of the octahedral geometry. For example, square pyramidal complexes may be considered within the octahedral model by labeling the empty site as a null "ligand" (Figure 2). This results in separate additive

7

contributions for the ligands adjacent (equatorial) and opposite (axial) to the open site because the number of *cis* interactions with the null ligand is equal to the number of equatorial ligands (Figure 2 and Supporting Information Text S2). The unknown parameters may be shared between octahedral and square pyramidal complexes if the property of interest may be evaluated in both geometries (i.e., not for reaction energies), but square pyramidal complexes require extra parameters to describe the null ligand. Truncated ligand MBE models of this type still lower screening costs by sharing these parameters between octahedral and square pyramidal complexes for a given property. In future work, the model could be similarly extended to predictions for square planar and seesaw complexes using two null ligands within these models.
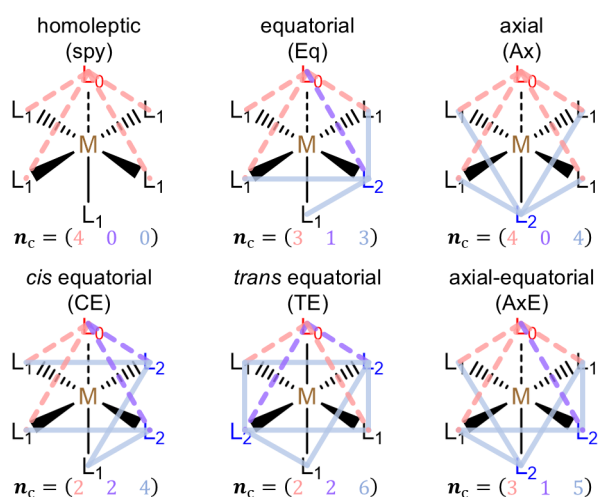

Figure 2. Symmetry classes for square pyramidal transition metal complexes with up to two unique ligands, L_1 and L_2, are labeled by their name above and corresponding interaction counts $\boldsymbol{n}_c = (n_{01,c}\ n_{02,c}\ n_{12,c})$ below. The open coordinating site is illustrated by a null ligand, L_0. The L_0–L_1, L_0–L_2, and L_1–L_2 *cis* interactions are respectively indicated in light red, purple, and light blue. For each pair of ligands, a total of twelve isomers can be obtained because all complexes are unique if the L_1 and L_2 labels are swapped.

To determine the minimum number of measurements required to specify the unknown parameters in our *cis* model for octahedral complexes, we examine known physical constraints. The first constraint is that the sum of the n_i always equals the coordination number (i.e., six in this

case). We also obtain a constraint for each ligand type by counting the number of each ligand involved in the *cis* interactions as

$$4n_i = 2n_{ii,\mathrm{c}} + \sum_{j \neq i} n_{ij,\mathrm{c}} \tag{5}$$

where the multiplicative factor 4 arises because each ligand is involved in 4 *cis* interactions (see Figure 1). We apply these constraints to eliminate $q^{(0)}$ and each $q_{ii,\mathrm{c}}$ from eq. (4) to obtain

$$Q_{\mathrm{c}}^{(2)} = \sum_{i=1}^{N} n_i q_i' + \sum_{i=1}^{N} \sum_{j>i}^{N} n_{ij,\mathrm{c}} q_{ij,\mathrm{c}}' \tag{6}$$

where the q' are the set of independent variables that satisfy eq. (4). Similar constraints may be obtained for any model by counting ligands or interactions (Supporting Information Text S3). Reviewing the form of eq. (6) indicates that the *cis* model has $N(N+1)/2$ independent parameters. Considering a modest set of $N = 12$ unique monodentate ligands, this means that we only require measurements on 78 of the 82,160 possible octahedral structures[77] that the model can describe (i.e., less than 0.1%). The percentage of required measurements decreases further as N increases, scaling as N^{-4}, representing enormous potential in accelerating the exploration of TMC space.

In building a ligand MBE model, there is a choice regarding which set of calculations or measurements to incorporate to inform the model. We suggest a minimal set of measurements to perform by evaluating the expected uncertainty in model predictions for binary complexes. Given that measurement uncertainty (e.g., due to error from the DFT functional) is likely independent of the TMC symmetry, we may assume a known variance for the purpose of minimizing errors and use standard uncertainty propagation for linear regression. We perform an exhaustive evaluation on each set of three out of the ten possible binary octahedral structures and find that the set with minimal total uncertainty comprises the *mer* symmetric (MS) and both octahedral homoleptic (oct) complexes. This result implies that the error in homoleptic averaging should be smaller for the

9

monoheteroleptic (5+1), *cis* symmetric (CS), and *fac* symmetric (FS) structures, in agreement with

our previous findings.[77] A Mathematica[103] script for this calculation is provided on Zenodo.[104]

Using this basis to solve for the unknown parameters in eq. (5) gives,

$$Q_{\mathrm{c}}^{(2)} = \frac{1}{6} \sum_{i=1}^{N} n_i Q_i^{\mathrm{oct}} + \frac{1}{8} \sum_{i=1}^{N} \sum_{j>i}^{N} n_{ij,\mathrm{c}} (Q_{ij}^{\mathrm{MS}} - Q_{ij}^{(1),\mathrm{MS}}) \tag{7}$$

where superscripts on Q indicate the quantity is measured for that symmetry, and subscripts

indicate the ligand composition.

For properties belonging to both octahedral and square pyramidal complexes, it is sufficient

to include the homoleptic square pyramidal structures to fully specify all unknowns in the *cis*

model. Unfortunately, reaction energies cannot be evaluated for an octahedral complex due to the

lack of an open site, so we require a basis entirely comprising binary square pyramidal structures,

which we characterize by the location and orientation of the minority ligands, giving five binary

symmetry classes – equatorial (Eq), axial (Ax), *cis*-equatorial (CE), *trans*-equatorial (TE), and

axial-equatorial (AxE) (Figure 2). Through the same procedure as in the octahedral symmetry

classes, we determine that the set of four square pyramidal structures that minimizes the total

expected uncertainty comprises both square pyramidal homoleptic (spy) and both AxE complexes.

With this, we have now shown how to apply our newly derived ligand additivity models for

octahedral and square pyramidal TMCs to substantially reduce screening costs.

3. Methods.

3a. Data Curation and Reaction Mechanism

We calculate reaction energies for the first two steps of the biologically-inspired radical

rebound mechanism for methane-to-methanol conversion[105] by mononuclear Fe(II) catalysts with

10

monodentate ligands (Scheme 1). We compute the oxo formation energy, $\Delta E(\text{oxo})$, from the N_2O

terminal oxidant as

$$\Delta E(\text{oxo}) = E(\mathbf{2}) + E(N_2) - E(\mathbf{1}) - E(N_2O) \qquad (8)$$

where (**1**) and (**2**) are in the same spin state. The high-valent Fe(IV)=O intermediate then catalyzes

hydrogen atom transfer (HAT) from a methane substrate to form a Fe(III)–OH intermediate (**3**),

leaving a methyl radical. We compute the reaction energy for the HAT step, $\Delta E(\text{HAT})$, as

$$\Delta E(\text{HAT}) = E(\mathbf{3}) + E(\bullet CH_3) - E(\mathbf{2}) - E(CH_4) \qquad (9)$$

We compute the reaction energetics across four spin surfaces: low-spin (LS), intermediate-spin

(IS), high-spin (HS), and ground-state spin-allowed (GSSA). The LS, IS, and HS surfaces assume

a singlet, triplet, and quintet resting state respectively and majority-spin addition for the HAT step

(i.e., to form a doublet, quartet, or sextet). The GSSA surface uses the ground state spin of (**1**) to

compute $\Delta E(\text{oxo})$ and uses the lower-energy spin state of the spin-accessible states of (**3**) formed

either through minority- or majority-spin addition to compute $\Delta E(\text{HAT})$. In cases where the singlet

is the ground state of (**1**), GSSA is the same as the LS surface.

We investigate all binary combinations of eight monodentate ligands spanning a large

range of field strengths: weak-field water (H_2O), methanol (MeOH), and hydrogen sulfide (H_2S);

intermediate-field ammonia (NH_3) and phosphine (PH_3); and strong-field carbonyl (CO),

acetonitrile (MeCN), and methyl isocyanide (misc).

Scheme 1. Studied steps of the radical rebound mechanism for the partial oxidation of methane to methanol on iron(II) transition metal complexes with monodentate ligands. The reaction steps proceed left-to-right from the resting state (**1**) to the metal-oxo intermediate (**2**) formed by two-electron oxidation with N_2O, followed by hydrogen atom transfer (HAT) to form a metal-hydroxo intermediate (**3**). In a complete cycle, the methyl radical ($\bullet CH_3$) rebounds to form a metal-bound methanol intermediate before methanol is released, regenerating (**1**). An iron(II) catalyst is shown with the most general monodentate ligands labeled L_{eq1}, L_{eq2}, L_{eq3}, L_{eq4}, and L_{ax}.
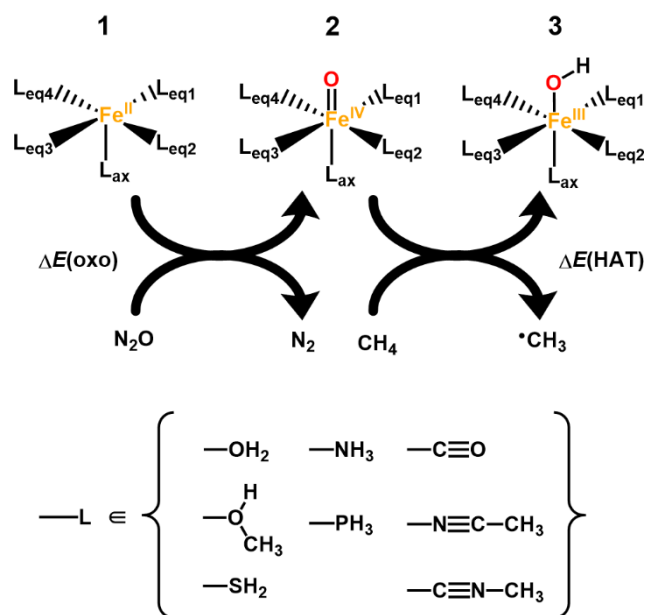
3b. Electronic Structure Calculations.

DFT geometry optimizations were performed using a development version of TeraChem v1.9.[106,107] The B3LYP[108-110] global hybrid functional was employed with the LANL2DZ[111] effective core potential for transition metals and the 6-31G* basis[112] for all other atoms. Calculations on all spin states (i.e., including singlet states) were performed in a spin-unrestricted formalism with level shifting[113] of 0.25 Ha applied to both majority- and minority-spin virtual orbitals to aid self-consistent field convergence. Geometry optimizations were carried out in the gas phase with the translation rotation internal coordinate optimizer[114] using the L-BFGS algorithm. Default tolerances of 4.5×10^{-4} Ha/bohr and 10^{-6} Ha were applied as the convergence criteria for the maximum gradient and energy difference between steps, respectively. DLPNO-CCSD(T) calculations were carried out in ORCA 5.0[115-117] using ORCA-computed B3LYP reference orbitals, iterative triples,[118] and the CPS1/CBS[3:4] extrapolation scheme.[61] When comparing DFT predictions to DLPNO-CCSD(T), we use the DFT energy from the B3LYP reference orbital calculation of the larger basis set from the CPS1/CBS[3:4] extrapolation (i.e.,

ZORA-recontracted[119] versions of the def2 basis sets[120] with ZORA-def2-QZVPP on Fe, ZORA-def2-SVP on H, and ZORA-def2-TZVP on all other atoms) rather than the energy from the TeraChem calculation in the basis set used for geometry optimizations (Supporting Information Table S1).

The initial structures for all species used in spin-splitting energy calculations and all Fe(IV)=O species were constructed using molSimplify v1.7,[121] which uses OpenBabel[122,123] as a backend. Metal-hydroxo initial geometries were generated by adding an H atom to the optimized metal-oxo structure as in prior work.[124,125] All resting state catalyst structures were obtained as single-point energies after removal of the oxygen atom from optimized metal-oxo structures.

Optimized geometries were validated as octahedral following previously developed criteria (Supporting Information Table S2).[126,127] Failed calculations were resubmitted, starting from the converged geometry of another spin state where available, with preference towards a lower-spin state. Structures flagged for ligand dissociation were resubmitted via an intermediate calculation with frozen metal–ligand distances on the initial geometry before full optimization of the coordinates. We checked for other conformations of structures with Jahn–Teller or axial distortions by restarting the geometry optimization with a fixed axial metal–ligand distance and kept the lowest-energy conformer that passed the geometry checks (Supporting Information Figure S2). Failed SCF calculations were reattempted without level shifting and subsequently with level shifting of 1.0 Ha and 0.1 Ha on majority- and minority-spin virtual orbitals respectively. Calculations that failed to pass convergence or geometry checks after these job rescue efforts did not affect conclusions but were eliminated (Supporting Information Tables S3 and S4).

3c. Regression and model evaluation.

13

Although the mean absolute error (MAE) is frequently used to measure performance within computational chemistry, it is a less useful metric when comparing regression models with differing numbers of model parameters, n_p. Instead, we define the scaled root-mean-square error (sRMSE), which includes a scaling factor based on the degrees of freedom in the regression,

$$\text{sRMSE} = \sqrt{\frac{n_m}{n_m - n_p - 1}} \bullet \text{RMSE} \qquad (8)$$

where n_m is the number of measurements or data points of the target variable. This scaling factor is based on that used by Ezekiel in the adjusted-R^2 for the evaluation of models with different numbers of parameters.[128-131] The sRMSE is a better metric for comparing the various orders of the ligand MBE models to see if the higher-order models are worth the increased cost of the extra parameters.

All regression tasks use the Moore–Penrose pseudo-inverse to obtain the least-squares solution of minimum norm as implemented in NumPy[132] v. 1.26. To enable a consistent definition of sRMSE and compare models, regression tasks involving bidentate ligand parameters assume that monodentate ligand parameters are fixed at the values obtained from the minimal basis solution (Supporting Information Table S5).

4. Results and Discussion.

4a. Evaluation of model accuracies on spin state energetics

While homoleptic averaging has been demonstrated for symmetric geometries including octahedral,[71,74,80,133] tetrahedral,[75] and linear/sandwich[73] geometries, it has not been thoroughly explored for asymmetric coordination geometries, such as square pyramidal. We thus first evaluate the performance of homoleptic averaging in predicting spin state energetics for square pyramidal

14

complexes and compare it to the performance for octahedral complexes. We investigate Fe(II) complexes containing up to two unique ligands from a set of four small monodentate ligands that span a large range of ligand field strengths: strong-field methyl isocyanide (misc) and carbonyl (CO), intermediate-field ammonia (NH_3), and weak-field water (H_2O). We compute the B3LYP adiabatic high-spin (HS, i.e., quintet) to low-spin (LS, i.e., singlet) splitting, ΔE_{H-L}, for all such complexes (i.e., 52 octahedral and 64 square pyramidal). The ΔE_{H-L} values for the homoleptic square pyramidal complexes are generally lower (i.e., more high-spin favoring) and span a smaller range than the corresponding homoleptic octahedral complexes (Figure 3 and Supporting Information Figure S3). This observation aligns with an additive interpretation where each ligand adds a contribution to the bare Fe(II) ion, which is strongly high-spin favoring, with the lowest-energy singlet excited state being 87 kcal/mol above the quintet ground state.[134,135]
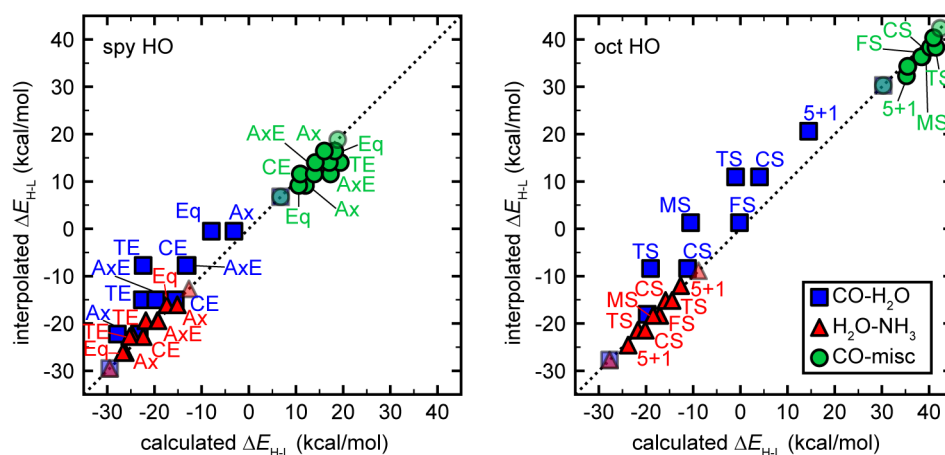

Figure 3. Parity plots for the homoleptic averaging (i.e., interpolated from the spy HO or oct HO sets) ΔE_{H-L} (in kcal/mol) vs. B3LYP computed values for Fe(II) between square pyramidal (left) and octahedral (right) geometries. Points are colored according to the pair of ligands they correspond to: CO–misc (green circles), H_2O–NH_3 (red triangles), and CO–H_2O (blue squares), as indicated in the inset. Key isomers are annotated. Points used for the interpolation are translucent, whereas the remaining points are opaque. In both panels, a black dotted parity line is shown.

An additive interpretation also suggests that the heteroleptic combinations should have $\Delta E_{\text{H-L}}$ values between those of the parent homoleptic complexes, as previously observed in octahedral complexes.[77] This also appears to hold generally for square pyramidal complexes (Figure 3 and Supporting Information Figure S3). The one exception is the TE isomer of $Fe(II)(CO)_2(misc)_3$, with a calculated $\Delta E_{\text{H-L}}$ of 19.4 kcal/mol that is slightly higher than the upper homoleptic limit of 18.8 kcal/mol for $Fe(II)(misc)_5$. In the interpretation of square pyramidal complexes as a special subset of octahedral complexes, the TE $Fe(II)(CO)_2(misc)_3$ complex would still be well within the range of homoleptic values for the octahedral complexes that is spanned by the bare metal ion (i.e., homoleptic octahedral with all null ligands) and $Fe(II)(misc)_6$ at the upper limit. However, no per-ligand additive explanation can account for differences between the TE and CE isomers of any square pyramidal complex, and CE $Fe(II)(CO)_2(misc)_3$ is inside the range of square pyramidal homoleptic values. Still, in predicting the heteroleptic values overall, homoleptic averaging achieves a mean absolute error (MAE) of 3.1 kcal/mol in the square pyramidal complexes that is comparable to the 3.4 kcal/mol MAE observed for octahedral complexes.

While the MAE for square pyramidal complexes is slightly lower and very nearly meets a looser threshold of 3 kcal/mol that has been suggested for transition metal chemical accuracy,[136,137] the magnitude of error depends on both the ligand composition and the symmetry class. Complexes comprising ligands with a greater difference between the parent homoleptic $\Delta E_{\text{H-L}}$ values on average exhibit higher errors from homoleptic averaging (Supporting Information Figure S4). The average magnitude of errors normalized by the difference in homoleptic parent compounds is 1.5 times larger for square pyramidal complexes than for octahedral complexes (0.12 vs. 0.08). Thus, the relative errors from homoleptic averaging may be more significant for square pyramidal

16

complexes than octahedral complexes. We previously observed for octahedral complexes that the TS and MS symmetry classes generally have greater errors.[77] Of the square pyramidal symmetry classes, TE is the most similar to TS and MS because adding a ligand to the open site of TE will result in either TS or MS depending on the ligand added. Indeed, the error for the TE class (MAE of 5.0 kcal/mol) is likewise greater than other square pyramidal symmetry classes (Supporting Information Table S6). Thus, homoleptic averaging performs equally, or worse, for asymmetric geometries compared with the symmetric geometries.

Using the ligand MBE models we introduced that have adjustable accuracy (see Sec. 2b), we now evaluate the tradeoff between cost and accuracy with respect to the expansion order across the set of $\Delta E_{\text{H-L}}$ in monodentate octahedral complexes. To provide unbiased treatment of each model's accuracy, we use linear regression on all 52 available data points rather than a minimal basis. That is, the "noninteracting" model evaluated here is given by eq. (3) and differs from homoleptic averaging in that each q_i is determined via linear regression instead of directly from the homoleptic complexes. While each model (i.e., the non-interacting, *cis* only, and *cis* + *trans*) meets the 3 kcal/mol MAE cutoff, the MAE, scaled RMSE (sRMSE), and maximum absolute error are each approximately halved upon including *cis* interactions, nearly meeting the standard 1 kcal/mol MAE cutoff for main-group chemical accuracy (Figure 4). However, the FS isomer of $Fe(II)(H_2O)_3(CO)_3$ is corrected in the wrong direction in the *cis* model, leading to a major outlier with an error greater than 5 kcal/mol (Figure 4). Accounting for *trans* interactions as well brings the maximum error under 3 kcal/mol but offers only minor improvement to the MAE and sRMSE (Figure 4). For these reasons, we suggest that the *cis* model is likely to yield the best trade-off of accurate predictions with the highest data efficiency. In practice, it is sufficient to use the minimal

17

basis presented in eq. (7), which reduces data required by up to eightfold for the *cis* model but only increases the MAE by 0.1 kcal/mol (Supporting Information Table S6).
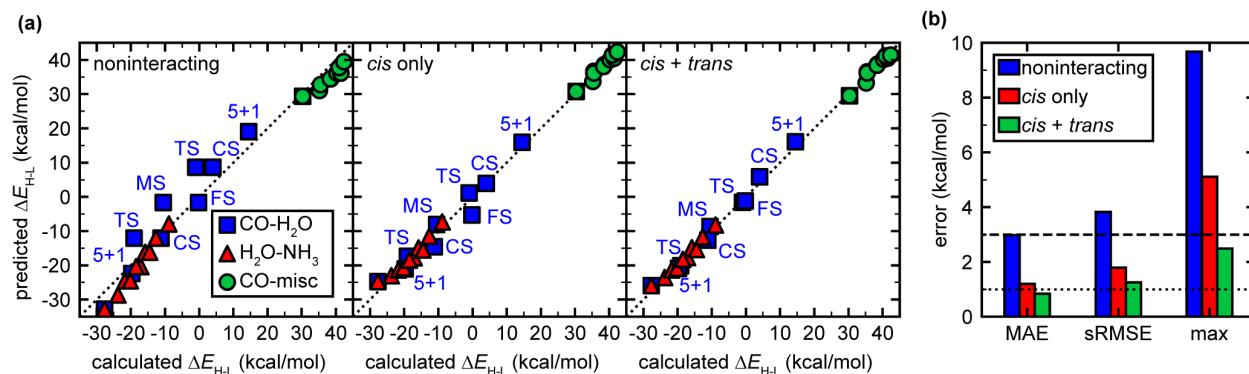

Figure 4. (a) Parity plots and (b) error metrics for the performance of the noninteracting model, the *cis* interaction model, and the *cis + trans* interactions model in predicting ΔE_{H-L} of octahedral Fe(II) complexes containing monodentate ligands. Points are colored according to the pair of ligands they correspond to: CO–misc (green circles), H_2O–NH_3 (red triangles), and CO–H_2O (blue squares), as indicated in the inset. Binary isomers of the CO–H_2O systems are annotated with their symmetry labels. In all panels of the parity plots, a black dotted parity line is shown. The error metrics are computed for all 52 binary and homoleptic systems investigated, including those absent from the parity plots. In the bar chart of error metrics, the transition metal chemical accuracy threshold (3 kcal/mol) is indicated by the dashed black line while main-group chemical accuracy (1 kcal/mol) is indicated by the dotted black line.

While the models thus far consider only monodentate ligands, most realistic transition metal complexes include higher-denticity ligands, motivating consideration of multidentate ligands in ligand additivity models. In previous studies, we have found that frameworks developed for monodentate ligands can work equally well for multidentate ligands.[124,138] Here, we consider bidentate ligands, which introduce different symmetry *cis* interactions (i.e., in-plane and out-of-plane). The new terms are redundant with the *trans* interaction terms (Supporting Information Text S4). Therefore, it is equivalent to use the same model as with monodentate ligands, but it is possible that the *trans* terms are more important for bidentate ligands. We investigate ΔE_{H-L} values in binary and ternary octahedral complexes that contain either 2,2′-bipyridine (bipy) or ethylenediamine (en) along with combinations of NH_3, H_2O, CO, and pyridine (pyr) ligands or alternatively binary

complexes of only the bidentate bipy/en ligands. We observe that homoleptic averaging performs well for the bidentate-only systems, with an MAE of 0.5 kcal/mol on $Fe(II)(bipy)(en)_2$ and $Fe(II)(bipy)_2(en)$ (Figure 5). This is in agreement with past literature observations that homoleptic averaging works well for binary bidentate complexes.[72] However, we find varied performance for the mixed bidentate-monodentate binary systems in both homoleptic averaging and the *cis* interaction model, with MAEs of 5.9 kcal/mol and 5.1 kcal/mol respectively on the three heteroleptic complexes of bipy and CO. Nevertheless, we observe that the inclusion of *trans* interactions nearly eliminates prediction error, reducing the MAE to 0.5 kcal/mol on the three heteroleptic complexes of bipy and CO (Figure 5). The increased number of parameters is balanced by the significant reduction in error in both binary and ternary complexes because the sRMSE value decreases from 2.9 kcal/mol to 1.2 kcal/mol in the binary systems and from 2.6 kcal/mol to 2.4 kcal/mol in the ternary systems with the addition of the *trans* parameters involving bipy (Figure 5 and Supporting Information Figure S5). While it appears that not all pairs of ligands require the *trans* interactions, the need in some cases (i.e., bipy–CO) warrants their inclusion. This implies that the presented ligand MBE models can be applied to multidentate ligands and motivates future study of which ligand pairs require additional terms.
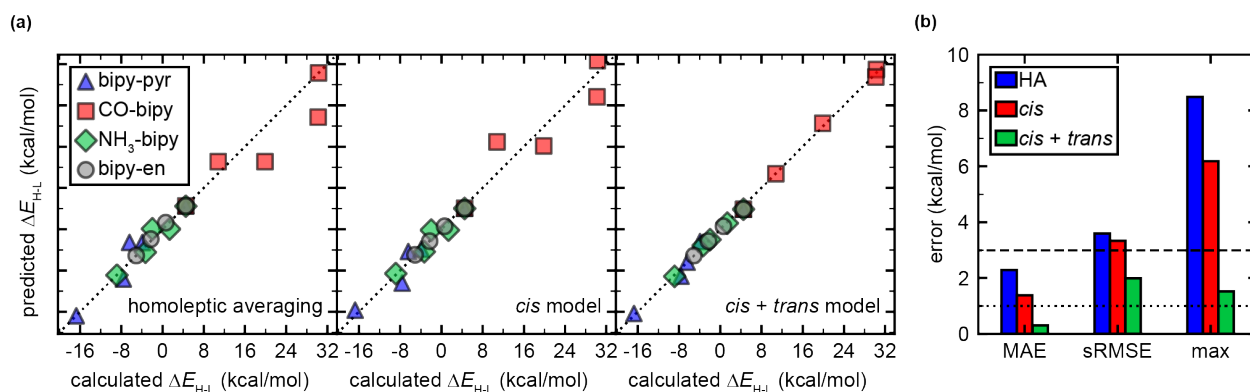

Figure 5. (a) Parity plots and (b) error metrics for the performance of homoleptic averaging (HA), the *cis* model, and the *cis + trans* model in predicting $\Delta E_{H\text{-}L}$ of Fe(II) complexes containing bidentate ligands. Points are colored according to the pair of ligands they correspond to: bipy–pyr (blue triangles), CO–bipy (red squares), NH_3–bipy (green diamonds), and bipy–en (gray circles).

19

In all panels of the parity plots, a black dotted parity line is shown. The error metrics are for all 32 binary systems investigated, including those absent from the parity plots. In the bar chart of error metrics, transition metal chemical accuracy (3 kcal/mol) is indicated by the dashed black line while main-group chemical accuracy (1 kcal/mol) is indicated by the dotted black line.

Having established that the *cis* model addresses most error in octahedral complexes with monodentate ligands, we next investigate the *cis* model performance for square pyramidal complexes. We first consider adiabatic spin-splitting energies because they can be evaluated for both octahedral and square pyramidal complexes, unlike the reaction energies of catalytic steps that are only compatible with a structure with an open metal site. Treating the square pyramidal complexes as octahedral complexes with a "null" ligand, we can directly apply the octahedral *cis* model. We investigate four cases: the *oct HO* case, which includes information only from the octahedral homoleptic complexes including the bare Fe(II) ion as the homoleptic octahedral case of the "null" ligand; the *spy HO* case, which includes information only from the square pyramidal complexes; the *all HO* case, which includes information from all homoleptic complexes (i.e., octahedral and square pyramidal); and *all HO + MS,* which includes information from all of the homoleptic complexes and the MS complexes (Figure 6). While the *all HO* case has very similar overall performance to *spy HO* across square pyramidal complexes, it performs slightly worse on all ligand pairs except H_2O–NH_3 (overall *all HO* MAE of 3.2 kcal/mol vs. 2.9 kcal/mol for *spy HO*, see Figure 3 and Supporting Information Table S7). The *oct HO* case gives noticeably worse predictions than *spy HO* (*oct HO* MAE of 4.8 kcal/mol). Only the fourth case, *all HO + MS*, gives superior performance to *spy HO* with an MAE of 2.4 kcal/mol (Figure 6). This observation shows that the *cis* model can utilize information from the binary heteroleptic octahedral complexes to inform predictions on the binary heteroleptic square pyramidal complexes without the need for calculations on any heteroleptic square pyramidal complexes. This removes the ambiguity of

20

selecting a basis in the square pyramidal complexes and saves on computational cost. It also shows that keeping the model parameters separate between the two geometries achieves slightly better performance within a given geometry, meaning that we do sacrifice some accuracy for the cost savings of sharing information between octahedral and square pyramidal geometries.
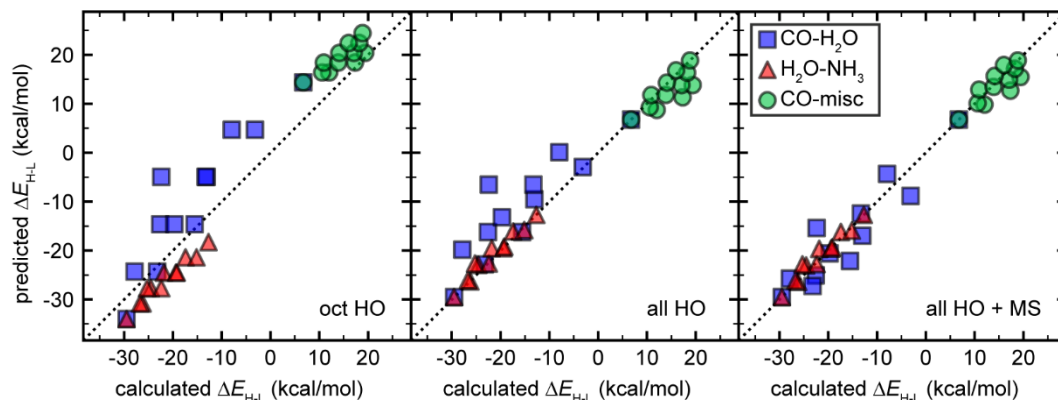

Figure 6. Parity plots of model predictions for $\Delta E_{\text{H-L}}$ (kcal/mol) vs. B3LYP computed values in Fe(II) square pyramidal complexes using models that share information with the octahedral complexes. The models are labeled based upon which complexes were used to obtain the model parameters. Points are colored according to the pair of ligands they correspond to: CO–misc (green circles), H_2O–NH_3 (red triangles), and CO–H_2O (blue squares), as indicated in the inset. In all panels, a black dotted parity line is shown.

4b. Ligand additivity trends in reaction energies

In previous studies, we showed that homogeneous catalysts do not always follow linear scaling relationships as strongly as heterogeneous catalysts do and are additionally affected by the spin state and metal center.[125,139,140] However, these previous studies investigated restricted subsets of binary symmetry classes. Thus, we analyzed the distribution of reaction energies in our data set, which includes all binary symmetry classes, to determine if this set yields trends consistent with prior observations of i) privileged ligand types for optimal reactivity[124] and ii) a lack of strong scaling relationships between reaction energies.[139] Within the homoleptic catalysts, we find that there is weaker correlation between B3LYP ΔE(oxo) and ΔE(HAT) energetics in the IS state than

in the LS or HS states, in agreement with previous findings (Figure 7 and Supporting Information Table S8).[139] The heteroleptic catalysts include values outside the range of homoleptic catalyst ΔE(oxo) and ΔE(HAT) values, leading to an even weaker scaling relationship over the full set of catalysts (Figure 7 and Supporting Information Table S8). This shows that substitutions in square pyramidal complexes can result in perturbations outside the range of homoleptic properties, a phenomenon we observed only once (i.e., TE $Fe(II)(CO)_2(misc)_3$) for spin-splitting energies (see Figure 3). Here, we find that the range of heteroleptic ΔE(oxo) is 25% larger than that of homoleptic ΔE(oxo), while the same difference in ranges is only 4% for ΔE(HAT). As a measure of reaction energetics, we quantify the sum, $\Sigma E=\Delta E$(oxo)$+\Delta E$(HAT), which, for favorable catalysts, should be close to or below zero (Figure 7). Within each spin state, ΣE generally increases with increasing ligand field strength, and weak-field MeOH gives the lowest sum within the homoleptic catalysts (i.e., the LS state shown as a green square or IS state shown as a green triangle in Figure 7). In this context, the most favorable points are IS complexes with weaker-field MeOH or H_2O in the equatorial plane and stronger-field misc or NH_3 in the axial position. This result implies that a balance of weak-field ligands in the equatorial plane with strong-field ligands in the axial position may make for better catalysts. Previous work similarly identified weak-field equatorial ligands as important for favorable HAT energetics.[124] Combining the new observation that strong-field axial ligands make oxo formation more favorable with the past observation that anionic ligands make methanol release more favorable,[124] we suggest future study of cyanide or nitrite as axial ligands.
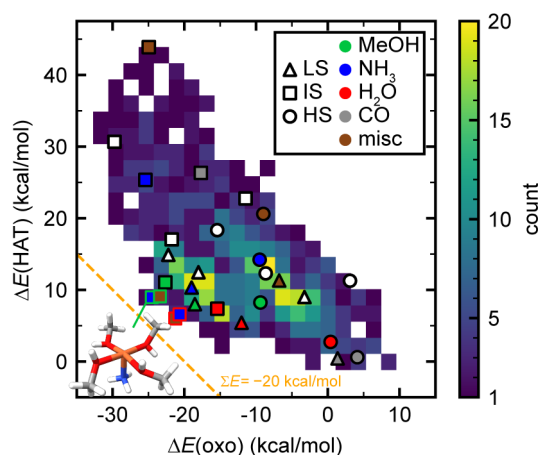
Figure 7. The hydrogen-atom transfer (HAT) reaction energy vs Fe(IV)-oxo formation (oxo) formation energy for all possible binary square pyramidal complexes colored by frequency from low (purple) to high (yellow). Homoleptic species are identified with black-outlined points and select ligands are colored, while the remaining ligands are white-filled points. Reactions on the low-spin (LS), intermediate-spin (IS), and high-spin (HS) surfaces are identified with triangles, squares, and circles respectively. The four points indicating the complexes with the lowest sum of the reaction energies (ΣE) are identified and filled with the color corresponding to the minority ligand and outlined with the color corresponding to the majority ligand – i.e., green and red outlines for majority MeOH and H_2O complexes, respectively, with blue and brown fills for complexes with minority ligands of NH_3 and misc, respectively. The dashed orange line indicates a ΣE value of -20 kcal/mol. The catalyst with the lowest ΣE, Ax Fe(II)(MeOH)$_4$(NH$_3$) in the intermediate-spin state is illustrated in the inset in its resting state.

Although we might expect thermochemistry predictions to be more straightforward than spin-state energetics because DFT reaction energies are typically less method-sensitive,[141] the reaction energy data appears noisier than the spin-splitting energies did. That is, we observe a greater variation in the values of ΔE(oxo) and ΔE(HAT) for heteroleptic complexes compared to the homoleptic values than we did for the $\Delta E_{\text{H-L}}$ property, suggesting additivity-based models may struggle. We first investigated the performance of homoleptic averaging for reaction energetics. We find that homoleptic averaging overall performs slightly better in predicting ΔE(HAT) than ΔE(oxo), with an MAE of 1.7 kcal/mol compared to 2.1 kcal/mol (Figure 8). Both reaction energy estimates from homoleptic averaging exhibit smaller MAEs than for $\Delta E_{\text{H-L}}$ (MAE of 3.1 kcal/mol). However, homoleptic averaging performs especially poorly in predicting the ΔE(oxo) of systems

containing mixtures of CO and H_2O ligands, with a maximum error of 6.2 kcal/mol in the HS state, which exceeds the magnitude of the range of homoleptic values (i.e., 3.7 kcal/mol, Figure 8). This example is indicative of the higher relative errors for ΔE(oxo) prediction than for ΔE(HAT). This aligns with the observation that heteroleptic complexes have a much larger range of ΔE(oxo) relative to the homoleptic range. The range may be larger since the oxo formation step involves the formation of a double bond with the metal center of the complex, leading to a stronger influence of subtle differences in ligand–metal–ligand interactions for oxo formation. Given that ΔE(HAT) is not strictly a metal-centered property, we will further investigate whether models with more interaction terms could improve the predictions to perform as well as those for ΔE(oxo) (see next). Additionally, homoleptic averaging performs notably worse in areas of interest for catalyst design. For the 10 binary complexes with the lowest ΣE values, homoleptic averaging generally overestimates both ΔE(HAT) and ΔE(oxo), respectively, by 5.8 kcal/mol and 2.6 kcal/mol on average (Supporting Information Table S9). In this zone of interest, the homoleptic averaging prediction errors for ΔE(HAT) are much higher than for ΔE(oxo), unlike the trend over the complete dataset. Moreover, homoleptic averaging cannot be used to find better catalysts than those already found because interpolation cannot predict a lower ΣE value than the lowest homoleptic ΣE value found.
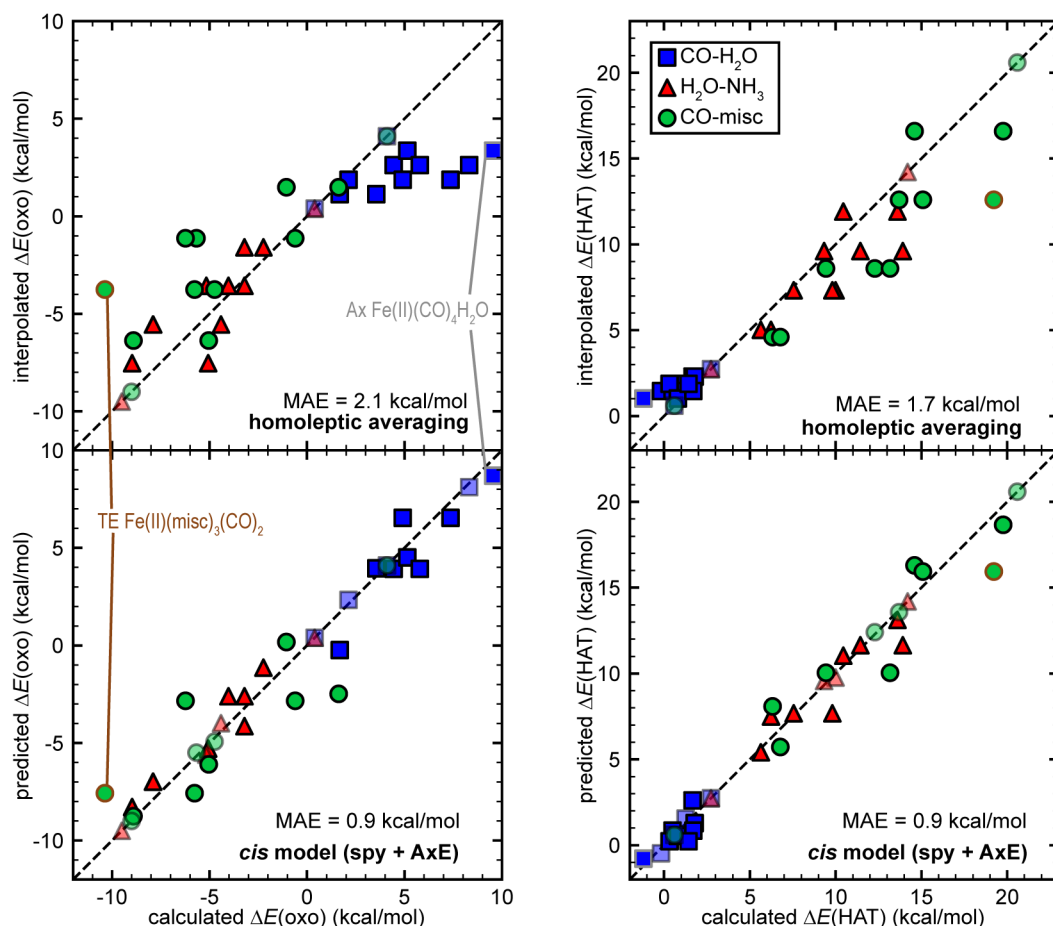
24

Figure 8. Parity plots for the HS reaction energies inferred by homoleptic averaging (top) and the *cis* model using the spy + AxE basis (bottom) vs B3LYP computed values for oxo formation (left) and hydrogen atom transfer (right) steps. Points are colored according to the pair of ligands they correspond to: CO–misc (green circles), H_2O–NH_3 (red triangles), and CO–H_2O (blue squares), as indicated in the inset. In each panel, the mean absolute error (MAE) across all 288 complexes in all three spin states is indicated in the bottom right corner. In all panels, a black dotted parity line is shown. Points seen by the models are translucent. Two complexes are labeled in the inset: TE $Fe(II)(misc)_3(CO)_2$ and Ax $Fe(II)(CO)_4(H_2O)$ with corresponding points outlined in gray and brown respectively.

We next investigated the extent to which including additional terms could improve model predictions of ΔE(oxo) and ΔE(HAT) reaction energetics. We showed (see Sec. 2b) that the statistically optimal basis for the *cis* model comprises the two AxE and the two spy complexes within a set of binary square pyramidal complexes, so we use all AxE and spy complexes to inform the regression for the *cis* model parameters. Across all three spin states and all pairs of ligands, the

25

cis model achieves an MAE of 0.9 kcal/mol for both ΔE(HAT) and ΔE(oxo). Notably, the *cis* model infers both the ΔE(oxo) and ΔE(HAT) of HS Ax Fe(II)(CO)$_4$(H$_2$O) within 0.8 kcal/mol of the ground-truth B3LYP result, whereas homoleptic averaging has errors of 6.2 and 2.2 kcal/mol for ΔE(oxo) and ΔE(HAT), respectively, for the same complex. While homoleptic averaging performed worse for ΔE(HAT), the *cis* model performs equally for both, indicating that the *cis* model performs well even for properties that are not metal-local. Notably, the *cis* model does not have a strong bias in the region of interest and significantly outperforms homoleptic averaging, with MAEs under 2 kcal/mol across the 10 complexes with the lowest ΣE values (Supporting Information Table S9). The *cis* model appears to perform better in regions with greater asymmetry in equatorial metal–ligand distances, which is surprising because our MBE-based models only account for ligand identity (Supporting Information Figure S6). That is, Jahn–Teller distortions are not provided as input to the *cis* model, but the model implicitly distinguishes complexes with these distortions based on data from the spy and AxE structures. Finally, the *cis* model performance appears to be unaffected by asymmetry within a ligand since it performs equally well for MeOH, which coordinates in an asymmetric fashion to the metal center (Supporting Information Table S10). Overall, we have demonstrated the ability of the *cis* model to make accurate predictions on thermochemical properties of transition metal complexes even outside the range of homoleptic values and in structures with asymmetry.

We now aim to demonstrate how this model can also be used to make accurate predictions on lower-symmetry (e.g., ternary) complexes. With only eight monodentate ligands, there are 5,328 possible square pyramidal complexes (Supporting Information Text S5). We thus use the *cis* model to predict if any lower-symmetry complexes would have more negative ΣE values than the IS Ax Fe(II)(MeOH)$_4$(NH$_3$) complex that has the lowest computed ΣE value of −15.6 kcal/mol.

The *cis* model in the spy + AxE basis does predict some ΣE values under -15.6 kcal/mol but also incorrectly predicts the binary complex with the lowest ΣE as the IS Ax Fe(II)(MeOH)$_4$(misc) at -16.9 kcal/mol (vs. a ground-truth value of -14.3 kcal/mol, see Table 1). In this sense, the *cis* model does not predict that there will be a lower ΣE value in the lower-symmetry complexes compared to the binary complexes among the LS, IS, and HS surfaces. Based on overall trends and predicted lowest ΣE values from the *cis* model, we perform DFT calculations on the eight lower-symmetry IS complexes with MeOH and H_2O in the equatorial plane and NH_3 or misc in the axial position (see Figure 7). As expected, none of these eight complexes have ΣE values lower than the lowest binary ΣE values (Table 1). This is indicative of a strong underprediction bias in this zone of interest by the *cis* model. It should be noted that while the *cis* model does not predict any ΣE values outside the range of binary values, it is still possible to predict values outside the range of binary values with the *cis* model (see next). Subsequent studies could improve upon the current performance by using an active-learning approach to refine the model performance in zones of interest and utilize dynamic models that allow for more interaction terms to be accounted for as more data is acquired.

Table 1. The *cis* model (with spy + AxE basis) predictions on complexes with lowest ΣE values. For the binary and ternary cases, the three complexes with the lowest overall ΣE values in any spin state are listed as well as the three lowest ΣE values that occur on the GSSA surface. Additionally, the binary complex with the lowest predicted ΣE value in the GSSA surface is shown since it is not within the three with the lowest calculated values.

	L_{eq1}	L_{eq2}	L_{eq3}	L_{eq4}	L_{ax}	spin surface	predicted ΣE (kcal/mol)	calculated ΣE (kcal/mol)	residual ΣE (kcal/mol)
Binary (overall)	MeOH	MeOH	MeOH	MeOH	NH_3	IS	-16.2	-15.6	-0.6
	H_2O	H_2O	H_2O	H_2O	misc	IS	-16.6	-15.3	-1.4
	MeOH	MeOH	MeOH	MeOH	misc	IS	-16.9	-14.3	-2.6
Binary (GSSA)	PH_3	PH_3	PH_3	PH_3	misc	LS	-1.2	-6.5	$+5.3$
	PH_3	PH_3	PH_3	misc	PH_3	LS	-1.7	-6.2	$+4.5$
	PH_3	PH_3	PH_3	PH_3	CO	LS	-1.8	-6.0	$+4.2$
	PH_3	misc	PH_3	misc	PH_3	LS	-3.9	-3.7	-0.2
Ternary (overall)	H_2O	MeOH	MeOH	MeOH	misc	IS	-16.7	-14.8	-1.8
	H_2O	MeOH	MeOH	MeOH	NH_3	IS	-15.8	-13.8	-2.0
	H_2O	MeOH	H_2O	MeOH	NH_3	IS	-15.5	-13.7	-1.7

27

	PH$_3$	PH$_3$	misc	misc	CO	LS	−4.0	−2.5	−1.5
Ternary (GSSA)	PH$_3$	misc	PH$_3$	misc	CO	LS	−4.0	−2.3	−1.7
	PH$_3$	PH$_3$	PH$_3$	misc	CO	LS	−4.1	−2.1	−2.1

Thus far, we have considered reactions on separate spin surfaces because simulating spin-forbidden reactions require significantly greater computational cost and are not relevant for testing our models. Realistically, many methane oxidation catalysts are believed to exhibit "two-state reactivity," requiring a transition to an excited state spin-surface.[142,143] Ideally, favorable reaction surfaces will be in the ground-state spin-allowed (GSSA) surface such that no change in spin state is required to follow the minimum energy path (see Sec. 3a), but spin state transitions are often invoked in Fe(II) catalysts.[124] In the present work, the lowest ΣE value found on the GSSA surface is −6.5 kcal/mol (i.e., for LS Ax Fe(II)(PH$_3$)$_4$(misc)) compared to −15.6 kcal/mol for any spin surface (i.e., IS Ax Fe(II)(MeOH)$_4$(NH$_3$)). We tested the ability of the *cis* model to predict the GSSA ΣE value of complexes directly and found comparable performance to the predictions across individual spin surfaces, with an MAE of 1.35 kcal/mol compared to an MAE of 1.25 kcal/mol on the IS surface (Supporting Information Table S11). Nevertheless, unlike on the IS surface, predictions of binary ΣE on the GSSA surface have a tendency towards overestimation at the lowest ΣE values (Table 1). Next, we used the *cis* model to predict the GSSA ΣE value of lower-symmetry complexes. While the *cis* model does not predict any GSSA ΣE values below −6.5 kcal/mol, it does predict three ternary complexes as the lowest three values (Table 1). However, the DFT-calculated values for these three ternary complexes were underestimated by the *cis* model and thus are not below −6.5 kcal/mol. These results indicate that it is more difficult to predict values on the GSSA surface, which aligns with previous findings that scaling relations do not hold across spin surfaces.[125] When there is a bias in prediction of the extrema of the binary complexes, this bias is also present in the lower-symmetry complexes, regardless of spin state convention.

28

Overall, the greater number of intermediates involved in estimating reaction energetics makes their prediction more challenging than spin-state energetics, and further improvements would likely benefit from an approach that used more detailed models than the *cis* model.

While accelerated prediction of DFT energetics is useful, it would be preferable to use models to predict energetics for higher-accuracy methods, such as DLPNO-CCSD(T). Indeed, we find that the performance of the *cis* interaction model for reproducing DLPNO-CCSD(T) values of reaction energetics is comparable to that for predicting DFT values, with a combined average MAE of about 1 kcal/mol for the two reaction steps (i.e., ΔE(oxo) and ΔE(HAT), Table 2). Once again, the Ax and TE symmetry classes have higher errors than other symmetry classes, indicating that a more detailed model could be beneficial for predicting these symmetry classes (Supporting Information Table S12). As a potentially even more powerful approach, we explored whether the *cis* model could be used for Δ-learning, where the model employs the same complexes as in the standard *cis* model but instead is used to predict the difference between DLPNO-CCSD(T) and B3LYP energetics for other configurations based on those in the training data. This predicted difference would then be combined with the explicitly calculated B3LYP energetics to infer DLPNO-CCSD(T) values on those remaining symmetries. In line with other "Δ-learning" approaches,[63] the *cis* model performs significantly better in predicting the difference (Δ) between B3LYP and DLPNO-CCSD(T) than in predicting either value independently with an overall MAE of 0.3 kcal/mol. This increased performance is especially evident for ΔE(HAT), where the MAE decreases from 1.3 kcal/mol for both B3LYP and DLPNO-CCSD(T) values to 0.3 kcal/mol for predicting the difference. Thus, high-accuracy combinatorial explorations can be accelerated by performing DLPNO-CCSD(T) calculations for higher-symmetry complexes in two ways: by accelerating direct prediction of DLPNO-CCSD(T) energies with a *cis* interaction model at an

29

accuracy of around 1 kcal/mol, or by carrying out all DFT calculations and the using the Δ-learning model to achieve even higher accuracy (ca. 0.3 kcal/mol with respect to DLPNO-CCSD(T)).

Table 2. Error metrics of the *cis* interaction model (with spy + AxE basis) for predictions of B3LYP, CCSD(T) using the DLPNO approximation, and Δ (the difference between B3LYP and DLPNO-CCSD(T)) for the 20 binary IS catalysts of H_2O/NH_3 and $H_2O/MeCN$ in units of kcal/mol.

	ΔE(HAT)			ΔE(oxo)			combined		
	B3LYP	CCSD(T)	Δ	B3LYP	CCSD(T)	Δ	B3LYP	CCSD(T)	Δ
MAE	1.25	1.26	0.34	0.80	0.90	0.30	1.03	1.08	0.32
RMSE	1.90	1.86	0.50	1.15	1.33	0.42	1.57	1.62	0.42
max	4.98	4.02	1.24	3.14	3.24	1.17	4.98	4.02	1.24

5. Conclusions

We described a physically motivated many-body expansion framework to infer the properties of low-symmetry coordination complexes from higher-symmetry coordination complexes with an adjustable cost–accuracy tradeoff. Using this framework, we derived the *cis* interaction model for mononuclear octahedral complexes. This model expands upon our previous interpolative schemes for octahedral complexes to include square pyramidal complexes and complexes with more than three unique ligand types. We further demonstrated that the model only requires $N(N+1)/2$ parameters to make predictions on the full space of $O(N^6)$ octahedral complexes, where N is the number of unique monodentate ligands in the design space. Uncertainty analysis revealed that the optimal set of measurements to use in making predictions for octahedral complexes comprises all homoleptic structures and the *mer* symmetric isomer for each pair of ligands.

To test the model performance, we computed spin-splitting energies for octahedral and square pyramidal complexes and determined that schemes beyond homoleptic averaging were also

needed in both square pyramidal and octahedral complexes for predictions of spin-state energetics. We found that our *cis* model improved prediction performance over homoleptic averaging in monodentate octahedral systems (MAE of ~1 kcal/mol vs. ~3 kcal/mol). Including *trans* interactions had little effect on the overall MAE, but they were necessary for accurate predictions on some monodentate and most bidentate systems. We also found that the octahedral spin-splitting energies could be used with the *cis* model to improve predictions on square pyramidal complexes.

We demonstrated that the *cis* interaction model predicts reaction energetics within 1 kcal/mol of calculated results on average for key reaction steps in methane-to-methanol conversion on square pyramidal molecular catalysts. Furthermore, we identified trends that lead to more favorable reaction energetics, suggesting that future studies should investigate strong-field ligands with a negative charge in the position *trans* to the reacting moiety. Finally, we showed that our models can be used to predict DLPNO-CCSD(T) energetics, with the lowest errors (0.3 kcal/mol) when the *cis* model is used to predict Δ-learning energies in combination with exhaustive DFT calculations. We expect that the models will perform similarly in predicting other properties of TMCs, including activation energies. Thus, our models present a novel means for data-efficient computational discovery of molecular catalysts from large chemical spaces.

ASSOCIATED CONTENT

Supporting Information.

Illustration and enumeration of *trans* interactions in octahedral complexes; proofs of model properties; example CPS1/CBS[3:4] extrapolation; geometric cutoffs and calculation recovery efforts; errors of homoleptic averaging with range of values; select model parameters; model performances with various properties; model performance for ternary complexes containing bidentate ligands; Jahn–Teller effect on model performance; enumeration of low-symmetry square pyramidal complexes (PDF)

This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*email:hjkulik@mit.edu

Notes

The authors declare no competing financial interest.

Vennelakanti, Gianmarco Terrones, Adam H. Steeves, and MIT's Writing and Communication

Center for feedback in the writing of this manuscript.

REFERENCES

(1)   Singh, V.; Palakkeezhillam, V. N. V.; Manakkadan, V.; Rasin, P.; Valsan, A. K.; Kumar, V. S.; Sreekanth, A. Recent Developments on the Potential Biological Applications of Transition Metal Complexes of Thiosemicarbazone Derivatives. *Polyhedron* **2023,** *245*, 116658.

(2)   Ruggiero, E.; Alonso-de Castro, S.; Habtemariam, A.; Salassa, L. In *Luminescent and Photoactive Transition Metal Complexes as Biomolecular Probes and Cellular Reagents*; Lo, Kenneth Kam-Wing, Ed.; Springer: Berlin, Heidelberg, 2015, DOI:10.1007/430_2014_165 10.1007/430_2014_165,69-107

(3)   Timoshnikov, V. A.; Selyutina, O. Y.; Polyakov, N. E.; Didichenko, V.; Kontoghiorghes, G. J. Mechanistic Insights of Chelator Complexes with Essential Transition Metals: Antioxidant/Pro-Oxidant Activity and Applications in Medicine. *Int. J. Mol. Sci.* **2022,** *23*, 1247.

(4)   Zysman-Colman, E. *Iridium(Iii) in Optoelectronic and Photonics Applications*; John Wiley & Sons, 2017.

(5)   COE, B. J.; CURATI, N. R. M. Metal Complexes for Molecular Electronics and Photonics. *Comments Inorg. Chem.* **2004,** *25*, 147–184.

(6)   Kalyanasundaram, K. Applications of Functionalized Transition Metal Complexes in Photonic and Optoelectronic Devices. *Coord. Chem. Rev.* **1998,** *177*, 347-414.

(7)   Androutsopoulos, A.; Sader, S.; Miliordos, E. Potential of Molecular Catalysts with Electron-Rich Transition Metal Centers for Addressing Long-Standing Chemistry Enigmas. *J. Phys. Chem. A* **2024,** *128*, 4401-4411.

(8)   Savéant, J.-M. Molecular Catalysis of Electrochemical Reactions. Mechanistic Aspects. *Chem. Rev.* **2008,** *108*, 2348-2378.

(9)   Choplin, A.; Quignard, F. From Supported Homogeneous Catalysts to Heterogeneous Molecular Catalysts. *Coord. Chem. Rev.* **1998,** *178-180*, 1679-1702.

(10)   Duan, C.; Ladera, A. J.; Liu, J. C.-L.; Taylor, M. G.; Ariyarathna, I. R.; Kulik, H. J. Exploiting Ligand Additivity for Transferable Machine Learning of Multireference Character across Known Transition Metal Complex Ligands. *J. Chem. Theory Comput.* **2022,** *18*, 4836-4845.

(11)   Nandy, A.; Duan, C.; Taylor, M. G.; Liu, F.; Steeves, A. H.; Kulik, H. J. Computational Discovery of Transition-Metal Complexes: From High-Throughput Screening to Machine Learning. *Chem. Rev.* **2021,** *121*, 9927-10000.

(12)   Shu, Y.; Levine, B. G. Simulated Evolution of Fluorophores for Light Emitting Diodes. *J. Chem. Phys.* **2015,** *142*, 104104.

(13)   Jorner, K.; Tomberg, A.; Bauer, C.; Sköld, C.; Norrby, P.-O. Organic Reactivity from Mechanism to Machine Learning. *Nat. Rev. Chem.* **2021,** *5*, 240-255.

(14)   Gómez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Duvenaud, D.; Maclaurin, D.; Blood-Forsythe, M. A.; Chae, H. S.; Einzinger, M.; Ha, D.-G.; Wu, T.; Markopoulos, G.; Jeon, S.; Kang, H.; Miyazaki, H.; Numata, M.; Kim, S.; Huang, W.; Hong, S. I.; Baldo, M.; Adams, R. P.; Aspuru-Guzik, A. Design of Efficient Molecular Organic Light-Emitting Diodes by a High-Throughput Virtual Screening and Experimental Approach. *Nat. Mater.* **2016,** *15*, 1120-1127.

(15)   Kanal, I. Y.; Owens, S. G.; Bechtel, J. S.; Hutchison, G. R. Efficient Computational Screening of Organic Polymer Photovoltaics. *J. Phys. Chem. Lett.* **2013,** *4*, 1613-1623.

(16)   Keith, J. A.; Vassilev-Galindo, V.; Cheng, B.; Chmiela, S.; Gastegger, M.; Müller, K.-R.; Tkatchenko, A. Combining Machine Learning and Computational Chemistry for Predictive Insights into Chemical Systems. *Chem. Rev.* **2021,** *121*, 9816-9872.

(17)   Curtarolo, S.; Hart, G. L. W.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O. The High-Throughput Highway to Computational Materials Design. *Nat. Mater.* **2013,** *12*, 191-201.

(18)   Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (Pymatgen): A Robust, Open-Source Python Library for Materials Analysis. *Comp Mater Sci* **2013,** *68*, 314-319.

(19)   Westermayr, J.; Gastegger, M.; Schütt, K. T.; Maurer, R. J. Perspective on Integrating Machine Learning into Computational Chemistry and Materials Science. *J. Chem. Phys.* **2021,** *154*, 230903.

(20)   Choudhary, K.; Kalish, I.; Beams, R.; Tavazza, F. High-Throughput Identification and Characterization of Two-Dimensional Materials Using Density Functional Theory. *Sci. Rep.* **2017,** *7*, 5179.

(21)   Armiento, R.; Kozinsky, B.; Fornari, M.; Ceder, G. Screening for High-Performance Piezoelectrics Using High-Throughput Density Functional Theory. *Phys Rev B* **2011,** *84*, 014103.

(22)   Vogiatzis, K. D.; Polynski, M. V.; Kirkland, J. K.; Townsend, J.; Hashemi, A.; Liu, C.; Pidko, E. A. Computational Approach to Molecular Catalysis by 3d Transition Metals: Challenges and Opportunities. *Chem. Rev.* **2019,** *119*, 2453-2523.

(23)   Fey, N.; Lynam, J. M. Computational Mechanistic Study in Organometallic Catalysis: Why Prediction Is Still a Challenge. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2022,** *12*, e1590.

(24)   Barrozo, A.; Orio, M. Molecular Electrocatalysts for the Hydrogen Evolution Reaction: Input from Quantum Chemistry. *ChemSusChem* **2019,** *12*, 4905-4915.

34

(25)   Greeley, J.; Jaramillo, T. F.; Bonde, J.; Chorkendorff, I.; Nørskov, J. K. Computational High-Throughput Screening of Electrocatalytic Materials for Hydrogen Evolution. *Nat. Mater.* **2006,** *5*, 909-913.

(26)   Trincado, M.; Banerjee, D.; Grützmacher, H. Molecular Catalysts for Hydrogen Production from Alcohols. *Energy Environ. Sci.* **2014,** *7*, 2464-2503.

(27)   Masero, F.; Perrin, M. A.; Dey, S.; Mougel, V. Dinitrogen Fixation: Rationalizing Strategies Utilizing Molecular Complexes. *Chem. Eur. J.* **2021,** *27*, 3892-3928.

(28)   Chalkley, M. J.; Drover, M. W.; Peters, J. C. Catalytic N2-to-Nh3 (or -N2h4) Conversion by Well-Defined Molecular Coordination Complexes. *Chem. Rev.* **2020,** *120*, 5582-5636.

(29)   Roux, Y.; Duboc, C.; Gennari, M. Molecular Catalysts for N2 Reduction: State of the Art, Mechanism, and Challenges. *ChemPhysChem* **2017,** *18*, 2606-2617.

(30)   Rodriguez, M. M.; Bill, E.; Brennessel, W. W.; Holland, P. L. N2 Reduction and Hydrogenation to Ammonia by a Molecular Iron-Potassium Complex. *Science* **2011,** *334*, 780-783.

(31)   Claveau, E. E.; Sader, S.; Jackson, B. A.; Khan, S. N.; Miliordos, E. Transition Metal Oxide Complexes as Molecular Catalysts for Selective Methane to Methanol Transformation: Any Prospects or Time to Retire? *Phys. Chem. Chem. Phys.* **2023,** *25*, 5313-5326.

(32)   Ravi, M.; Ranocchiari, M.; van Bokhoven, J. A. The Direct Catalytic Oxidation of Methane to Methanol—a Critical Assessment. *Angew. Chem. Int. Ed.* **2017,** *56*, 16464-16483.

(33)   Latimer, A. A.; Kakekhani, A.; Kulkarni, A. R.; Nørskov, J. K. Direct Methane to Methanol: The Selectivity–Conversion Limit and Design Strategies. *ACS Catal.* **2018,** *8*, 6894-6907.

(34)   White, M. V.; Claveau, E. E.; Miliordos, E.; Vogiatzis, K. D. Electronic Structure and Ligand Effects on the Activation and Cleavage of N 2 on a Molybdenum Center. *J. Phys. Chem. A* **2024,** *128*, 2038-2048.

(35)   Jones, G. M.; Smith, B. A.; Kirkland, J. K.; Vogiatzis, K. D. Data-Driven Ligand Field Exploration of Fe (Iv)–Oxo Sites for C–H Activation. *Inorg. Chem. Front.* **2023,** *10*, 1062–1075.

(36)   Kirkland, J. K.; Khan, S. N.; Casale, B.; Miliordos, E.; Vogiatzis, K. D. Ligand Field Effects on the Ground and Excited States of Reactive Feo $^{2+}$ Species. *Phys. Chem. Chem. Phys.* **2018,** *20*, 28786-28795.

(37)   Christoforidis, K. C.; Pantazis, D. A.; Bonilla, L. L.; Bletsa, E.; Louloudi, M.; Deligiannakis, Y. Axial Ligand Effect on the Catalytic Activity of Biomimetic Fe-Porphyrin Catalyst: An Experimental and Dft Study. *J. Catal.* **2016,** *344*, 768-777.

(38)   Popoola, S. A.; Jaseer, E. A.; Al-Saadi, A. A.; Polo, V.; Casado, M. A.; Oro, L. A. Iridium Complexes as Catalysts in the Hydrogen Transfer of Isopropanol to Acetophenone: Ligand Effects and Dft Studies. *Inorg. Chim. Acta* **2015,** *436*, 146-151.

(39) Straub, B. F. Ligand Influence on Metathesis Activity of Ruthenium Carbene Catalysts: A Dft Study. *Adv. Synth. Catal.* **2007,** *349*, 204-214.

(40) Zhang, J.; Lin, J.; Li, Y.; Shao, Y.; Huang, X.; Zhao, C.; Ke, Z. The Effect of Auxiliary Ligand on the Mechanism and Reactivity: Dft Study on H 2 Activation by Lewis Acid–Transition Metal Complex (Tris(Phosphino)Borane)Fe(L). *Catal. Sci. Technol.* **2017,** *7*, 4866-4878.

(41) Jover, J.; Fey, N.; Harvey, J. N.; Lloyd-Jones, G. C.; Orpen, A. G.; Owen-Smith, G. J. J.; Murray, P.; Hose, D. R. J.; Osborne, R.; Purdie, M. Expansion of the Ligand Knowledge Base for Monodentate P-Donor Ligands (Lkb-P). *Organometallics* **2010,** *29*, 6245-6258.

(42) Fey, N.; Tsipis, A. C.; Harris, S. E.; Harvey, J. N.; Orpen, A. G.; Mansson, R. A. Development of a Ligand Knowledge Base, Part 1: Computational Descriptors for Phosphorus Donor Ligands. *Chem. Eur. J.* **2006,** *12*, 291-302.

(43) Gugler, S.; Paul Janet, J.; J. Kulik, H. Enumeration of De Novo Inorganic Complexes for Chemical Discovery and Machine Learning. *Mol Syst Des Eng* **2020,** *5*, 139-152.

(44) Francis, M. B.; Finney, N. S.; Jacobsen, E. N. Combinatorial Approach to the Discovery of Novel Coordination Complexes. *J. Am. Chem. Soc.* **1996,** *118*, 8983-8984.

(45) Francis, M. B.; Jamison, T. F.; Jacobsen, E. N. Combinatorial Libraries of Transition-Metal Complexes, Catalysts and Materials. *Curr. Opin. Chem. Biol.* **1998,** *2*, 422-428.

(46) Renom-Carrasco, M.; Lefort, L. Ligand Libraries for High ThroAput Screening of Homogeneous Catalysts. *Chem. Soc. Rev.* **2018,** *47*, 5038-5060.

(47) Chen, S.-S.; Meyer, Z.; Jensen, B.; Kraus, A.; Lambert, A.; Ess, D. H. Realigands: A Ligand Library Cultivated from Experiment and Intended for Molecular Computational Catalyst Design. *J. Chem. Inf. Model.* **2023,** *63*, 7412-7422.

(48) Reetz, M. T. Combinatorial Transition-Metal Catalysis: Mixing Monodentate Ligands to Control Enantio-, Diastereo-, and Regioselectivity. *Angew. Chem. Int. Ed.* **2008,** *47*, 2556-2588.

(49) Durand, D. J.; Fey, N. Building a Toolbox for the Analysis and Prediction of Ligand and Catalyst Effects in Organometallic Catalysis. *Acc. Chem. Res.* **2021,** *54*, 837-848.

(50) Foscato, M.; Jensen, V. R. Automated in Silico Design of Homogeneous Catalysts. *ACS Catal.* **2020,** *10*, 2354-2377.

(51) Foscato, M.; Occhipinti, G.; Venkatraman, V.; Alsberg, B. K.; Jensen, V. R. Automated Design of Realistic Organometallic Molecules from Fragments. *J. Chem. Inf. Model.* **2014,** *54*, 767-780.

(52) Foscato, M.; Venkatraman, V.; Occhipinti, G.; Alsberg, B. K.; Jensen, V. R. Automated Building of Organometallic Complexes from 3d Fragments. *J. Chem. Inf. Model.* **2014,** *54*, 1919-1931.

(53) Jin, H.; Merz, K. M. J. Liganddiff: De Novo Ligand Design for 3d Transition Metal Complexes with Diffusion Models. *J. Chem. Theory Comput.* **2024,** *20*, 4377-4384.

(54) Foscato, M.; Houghton, B. J.; Occhipinti, G.; Deeth, R. J.; Jensen, V. R. Ring Closure to Form Metal Chelates in 3d Fragment-Based De Novo Design. *J. Chem. Inf. Model.* **2015,** *55*, 1844-1856.

(55) Gensch, T.; dos Passos Gomes, G.; Friederich, P.; Peters, E.; Gaudin, T.; Pollice, R.; Jorner, K.; Nigam, A.; Lindner-D'Addario, M.; Sigman, M. S.; Aspuru-Guzik, A. A Comprehensive Discovery Platform for Organophosphorus Ligands for Catalysis. *J. Am. Chem. Soc.* **2022,** *144*, 1205-1217.

(56) Foscato, M.; Deeth, R. J.; Jensen, V. R. Integration of Ligand Field Molecular Mechanics in Tinker. *J. Chem. Inf. Model.* **2015,** *55*, 1282-1290.

(57) Kneiding, H.; Nova, A.; Balcells, D. Directional Multiobjective Optimization of Metal Complexes at the Billion-System Scale. *Nat. Comput. Sci.* **2024,** *4*, 263-273.

(58) Feldt, M.; Phung, Q. M.; Pierloot, K.; Mata, R. A.; Harvey, J. N. Limits of Coupled-Cluster Calculations for Non-Heme Iron Complexes. *J. Chem. Theory Comput.* **2019,** *15*, 922-937.

(59) Phung, Q. M.; Feldt, M.; Harvey, J. N.; Pierloot, K. Toward Highly Accurate Spin State Energetics in First-Row Transition Metal Complexes: A Combined Caspt2/Cc Approach. *J. Chem. Theory Comput.* **2018,** *14*, 2446-2455.

(60) Hait, D.; Tubman, N. M.; Levine, D. S.; Whaley, K. B.; Head-Gordon, M. What Levels of Coupled Cluster Theory Are Appropriate for Transition Metal Systems? A Study Using near-Exact Quantum Chemical Values for 3d Transition Metal Binary Compounds. *J. Chem. Theory Comput.* **2019,** *15*, 5370-5385.

(61) Drosou, M.; Mitsopoulou, C. A.; Pantazis, D. A. Reconciling Local Coupled Cluster with Multireference Approaches for Transition Metal Spin-State Energetics. *J. Chem. Theory Comput.* **2022,** *18*, 3538-3548.

(62) Riplinger, C.; Neese, F. An Efficient and near Linear Scaling Pair Natural Orbital Based Local Coupled Cluster Method. *J. Chem. Phys.* **2013,** *138*.

(63) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach. *J. Chem. Theory Comput.* **2015,** *11*, 2087-2096.

(64) Friesner, R. A.; Jerome, S. V. Localized Orbital Corrections for Density Functional Calculations on Transition Metal Containing Systems. *Coord. Chem. Rev.* **2017,** *344*, 205-213.

(65) Rinaldo, D.; Tian, L.; Harvey, J. N.; Friesner, R. A. Density Functional Localized Orbital Corrections for Transition Metals. *J. Chem. Phys.* **2008,** *129*, 164108.

(66) Janet, J. P.; Duan, C.; Yang, T.; Nandy, A.; Kulik, H. J. A Quantitative Uncertainty Metric Controls Error in Neural Network-Driven Chemical Discovery. *Chem. Sci.* **2019,** *10*, 7913-7922.

(67) Janet, J. P.; Liu, F.; Nandy, A.; Duan, C.; Yang, T.; Lin, S.; Kulik, H. J. Designing in the Face of Uncertainty: Exploiting Electronic Structure and Machine Learning Models for Discovery in Inorganic Chemistry. *Inorg. Chem.* **2019,** *58*, 10592-10606.

37

(68)   Benson, S. W.; Cruickshank, F. R.; Golden, D. M.; Haugen, G. R.; O'Neal, H. E.; Rodgers, A. S.; Shaw, R.; Walsh, R. Additivity Rules for the Estimation of Thermochemical Properties. *Chem. Rev.* **1969,** *69,* 279-324.

(69)   Benson, S. W.; Buss, J. H. Additivity Rules for the Estimation of Molecular Properties. Thermodynamic Properties. *J. Chem. Phys.* **1958,** *29,* 546-572.

(70)   Glerup, J.; Moensted, O.; Schaeffer, C. E. Nonadditive and Additive Ligand Fields and Spectrochemical Series Arising from Ligand Field Parameterization Schemes. Pyridine as a Nonlinearly Ligating .Pi.-Back-Bonding Ligand toward Chromium(Iii). *Inorg. Chem.* **1976,** *15,* 1399-1407.

(71)   Constable, E. C.; Cargill Thompson, A. M. W.; Tocher, D. A. In *Supramol. Chem.*; Balzani, V.;Cola, L., Eds.; Springer Netherlands: Dordrecht, 1992, DOI:10.1007/978-94-011-2492-8_14 10.1007/978-94-011-2492-8_14,219-233

(72)   Haga, M.-a.; Matsumura-lnoue, T.; Shimizu, K. Ligand Additivity in the Oxidation Potentials of Bidentate Mixed-Ligand Ruthenium(11) Complexes. *J. CHEM. SOC. DALTON TRANS.* **1989,** 3.

(73)   N. Tarakanova, E.; A. Tarakanov, P.; O. Simakov, A.; Furuyama, T.; Kobayashi, N.; V. Konev, D.; A. Goncharova, O.; A. Trashin, S.; Wael, K. D.; V. Sulimenkov, I.; V. Filatov, V.; I. Kozlovskiy, V.; G. Tomilova, L.; A. Stuzhin, P.; E. Pushkarev, V. Synthesis and Characterization of Heteroleptic Rare Earth Double-Decker Complexes Involving Tetradiazepinoporphyrazine and Phthalocyanine Macrocycles. *Dalton Trans.* **2021,** *50,* 6245-6255.

(74)   Zalis, S.; Krejcik, M.; Drchal, V.; Vlcek, A. A. The Sequence of Ligand Reductions in Heteroleptic Ruthenium-Diimine Complexes: Calculation of Redox Potentials as a Diagnostic Tool. *Inorg. Chem.* **1995,** *34,* 6008-6014.

(75)   Leandri, V.; Pizzichetti, A. R. P.; Xu, B.; Franchi, D.; Zhang, W.; Benesperi, I.; Freitag, M.; Sun, L.; Kloo, L.; Gardner, J. M. Exploring the Optical and Electrochemical Properties of Homoleptic Versus Heteroleptic Diimine Copper(I) Complexes. *Inorg. Chem.* **2019,** *58,* 12167-12177.

(76)   Punt, P. M.; Stratmann, L. M.; Sevim, S.; Knauer, L.; Strohmann, C.; Clever, G. H. Heteroleptic Coordination Environments in Metal-Mediated DNA G-Quadruplexes. *Front. Chem. (Lausanne, Switz.)* **2020,** *8,* 26.

(77)   Arunachalam, N.; Gugler, S.; Taylor, M. G.; Duan, C.; Nandy, A.; Janet, J. P.; Meyer, R.; Oldenstaedt, J.; Chu, D. B. K.; Kulik, H. J. Ligand Additivity Relationships Enable Efficient Exploration of Transition Metal Chemical Space. *J. Chem. Phys.* **2022,** *157,* 184112.

(78)   Cytter, Y.; Nandy, A.; Bajaj, A.; Kulik, H. J. Ligand Additivity and Divergent Trends in Two Types of Delocalization Errors from Approximate Density Functional Theory. *J. Phys. Chem. Lett.* **2022,** *13,* 4549-4555.

38

(79)   Bursten, B. E. Ligand Additivity: Applications to the Electrochemistry and Photoelectron Spectroscopy of D6 Octahedral Complexes. *J. Am. Chem. Soc.* **1982,** *104*, 1299-1304.

(80)   Deeth, R. J.; Foulis, D. L.; Williams-Hubbard, B. J. Molecular Mechanics for Multiple Spin States of Transition Metal Complexes. *Dalton Trans.* **2003**, DOI:10.1039/b305868a 10.1039/b305868a, 3949-3955.

(81)   Wood, J. S. Ligand Field Theory for Pentacoordinate Molecules. I. The Electronic Spectra of Species with D(Sub 3h) and C(Sub 3v) Symmetry. *Inorg. Chem.* **1968,** *7*, 852-859.

(82)   Vanquickenborne, L. G.; Ceulemans, A. Ligand-Field Models and the Photochemistry of Coordination Compounds. *Coord. Chem. Rev.* **1983,** *48*, 157-202.

(83)   Schäffer, C. E. The Non-Additive Ligand Field. *Theor. Chim. Acta* **1974,** *34*, 237-243.

(84)   Das, A. K.; Urban, L.; Leven, I.; Loipersberger, M.; Aldossary, A.; Head-Gordon, M.; Head-Gordon, T. Development of an Advanced Force Field for Water Using Variational Energy Decomposition Analysis. *J. Chem. Theory Comput.* **2019,** *15*, 5001-5013.

(85)   Heindel, J. P.; Xantheas, S. S. Molecular Dynamics Driven by the Many-Body Expansion (Mbe-Md). *J. Chem. Theory Comput.* **2021,** *17*, 7341-7352.

(86)   Babin, V.; Medders, G. R.; Paesani, F. Toward a Universal Water Model: First Principles Simulations from the Dimer to the Liquid Phase. *J. Phys. Chem. Lett.* **2012,** *3*, 3765-3769.

(87)   Straight, S. C.; Paesani, F. Exploring Electrostatic Effects on the Hydrogen Bond Network of Liquid Water through Many-Body Molecular Dynamics. *The Journal of Physical Chemistry B* **2016,** *120*, 8539-8546.

(88)   Gao, J.; Wang, Y. Communication: Variational Many-Body Expansion: Accounting for Exchange Repulsion, Charge Delocalization, and Dispersion in the Fragment-Based Explicit Polarization Method. *J. Chem. Phys.* **2012,** *136*, 071101.

(89)   Bull-Vulpe, E. F.; Riera, M.; Bore, S. L.; Paesani, F. Data-Driven Many-Body Potential Energy Functions for Generic Molecules: Linear Alkanes as a Proof-of-Concept Application. *J. Chem. Theory Comput.* **2022,** *19*, 4494-4509.

(90)   Richard, R. M.; Herbert, J. M. A Generalized Many-Body Expansion and a Unified View of Fragment-Based Methods in Electronic Structure Theory. *J. Chem. Phys.* **2012,** *137*, 064113.

(91)   Richard, R. M.; Lao, K. U.; Herbert, J. M. Aiming for Benchmark Accuracy with the Many-Body Expansion. *Acc. Chem. Res.* **2014,** *47*, 2828-2836.

(92)   Liu, K.-Y.; Herbert, J. M. Energy-Screened Many-Body Expansion: A Practical yet Accurate Fragmentation Method for Quantum Chemistry. *J. Chem. Theory Comput.* **2020,** *16*, 475-487.

(93)   Schmitt-Monreal, D.; Jacob, C. R. Density-Based Many-Body Expansion as an Efficient and Accurate Quantum-Chemical Fragmentation Method: Application to Water Clusters. *J. Chem. Theory Comput.* **2021,** *17*, 4144-4156.

(94) Dahlke, E. E.; Truhlar, D. G. Electrostatically Embedded Many-Body Expansion for Simulations. *J. Chem. Theory Comput.* **2008,** *4*, 1-6.

(95) Lao, K. U.; Herbert, J. M. Accurate and Efficient Quantum Chemistry Calculations for Noncovalent Interactions in Many-Body Systems: The Xsapt Family of Methods. *J. Phys. Chem. A* **2015,** *119*, 235-252.

(96) Janesko, B. G. Multiconfigurational Correlation at Dft + U Cost: On-Site Electron–Electron Interactions Yield a Block-Localized Configuration Interaction Hamiltonian. *J. Phys. Chem. A* **2024,** *128*, 5077-5087.

(97) Greiner, J.; Gianni, I.; Nottoli, T.; Lipparini, F.; Eriksen, J. J.; Gauss, J. Mbe-Casscf Approach for the Accurate Treatment of Large Active Spaces. *J. Chem. Theory Comput.* **2024,** *20*, 4663-4675.

(98) Abraham, V.; Mayhall, N. J. Cluster Many-Body Expansion: A Many-Body Expansion of the Electron Correlation Energy About a Cluster Mean Field Reference. *J. Chem. Phys.* **2021,** *155*, 054101.

(99) Ma, Q.; Werner, H.-J. Explicitly Correlated Local Coupled-Cluster Methods Using Pair Natural Orbitals. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018,** *8*, e1371.

(100) Eriksen, J. J.; Gauss, J. Generalized Many-Body Expanded Full Configuration Interaction Theory. *J. Phys. Chem. Lett.* **2019,** *10*, 7910-7915.

(101) Kurbanov, E. K.; Leverentz, H. R.; Truhlar, D. G.; Amin, E. A. Analysis of the Errors in the Electrostatically Embedded Many-Body Expansion of the Energy and the Correlation Energy for Zn and Cd Coordination Complexes with Five and Six Ligands and Use of the Analysis to Develop a Generally Successful Fragmentation Strategy. *J. Chem. Theory Comput.* **2013,** *9*, 2617-2628.

(102) Hua, D.; Leverentz, H. R.; Amin, E. A.; Truhlar, D. G. Assessment and Validation of the Electrostatically Embedded Many-Body Expansion for Metal–Ligand Bonding. *J. Chem. Theory Comput.* **2011,** *7*, 251-255.

(103) Wolfram Research, I.; 12.1 ed.; Wolfram Research, Inc.: Champaign, IL, 2020.

(104) Chu, D. B. K.; González-Narváez, D. A.; Meyer, R.; Nandy, A.; Kulik, H. J.; Zenodo, 2024, DOI:10.5281/zenodo.13338063 10.5281/zenodo.13338063.

(105) Nandy, A.; Adamji, H.; Kastner, D. W.; Vennelakanti, V.; Nazemi, A.; Liu, M.; Kulik, H. J. Using Computational Chemistry to Reveal Nature's Blueprints for Single-Site Catalysis of C–H Activation. *ACS Catal.* **2022,** *12*, 9281-9306.

(106) Seritan, S.; Bannwarth, C.; Fales, B. S.; Hohenstein, E. G.; Kokkila-Schumacher, S. I. L.; Luehr, N.; Snyder, J. W.; Song, C.; Titov, A. V.; Ufimtsev, I. S.; Martínez, T. J. Terachem: Accelerating Electronic Structure and *Ab Initio* Molecular Dynamics with Graphical Processing Units. *J. Chem. Phys.* **2020,** *152*, 224110.

(107) Seritan, S.; Bannwarth, C.; Fales, B. S.; Hohenstein, E. G.; Isborn, C. M.; Kokkila-Schumacher, S. I. L.; Li, X.; Liu, F.; Luehr, N.; Snyder Jr., J. W.; Song, C.; Titov, A. V.; Ufimtsev, I. S.; Wang, L.-P.; Martínez, T. J. Terachem: A Graphical Processing Unit-

Accelerated Electronic Structure Package for Large-Scale Ab Initio Molecular Dynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2021,** *11*, e1494.

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

(109) Becke, A. D. Density-Functional Thermochemistry. Iii. The Role of Exact Exchange. *J. Chem. Phys.* **1993,** *98*, 5648-5652.

(110) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *The Journal of Physical Chemistry* **1994,** *98*, 11623-11627.

(111) Hay, P. J.; Wadt, W. R. Ab Initio Effective Core Potentials for Molecular Calculations. Potentials for the Transition Metal Atoms Sc to Hg. *J. Chem. Phys.* **1985,** *82*, 270-283.

(112) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. Ix. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971,** *54*, 724-728.

(113) Saunders, V. R.; Hillier, I. H. A "Level–Shifting" Method for Converging Closed Shell Hartree–Fock Wave Functions. *Int. J. Quantum Chem* **1973,** *7*, 699-705.

(114) Wang, L.-P.; Song, C. Geometry Optimization Made Simple with Translation and Rotation Coordinates. *J. Chem. Phys.* **2016,** *144*, 214108.

(115) Neese, F. The Orca Program System. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012,** *2*, 73-78.

(116) Neese, F. Software Update: The Orca Program System—Version 5.0. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2022,** *12*, e1606.

(117) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The Orca Quantum Chemistry Program Package. *J. Chem. Phys.* **2020,** *152*, 224108.

(118) Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D. G.; Minenkov, Y.; Cavallo, L.; Neese, F. Communication: An Improved Linear Scaling Perturbative Triples Correction for the Domain Based Local Pair-Natural Orbital Based Singles and Doubles Coupled Cluster Method [Dlpno-Ccsd(T)]. *J. Chem. Phys.* **2018,** *148*, 011101.

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

(120) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005,** *7*, 3297-3305.

(121) Ioannidis, E. I.; Gani, T. Z. H.; Kulik, H. J. Molsimplify: A Toolkit for Automating Discovery in Inorganic Chemistry. *J. Comput. Chem.* **2016,** *37*, 2106–2117.

(122) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminf.* **2011,** *3*, 33.

(123) O'Boyle, N. M.; Morley, C.; Hutchison, G. R. Pybel: A Python Wrapper for the Openbabel Cheminformatics Toolkit. *Chem. Cent. J.* **2008**, *2*, 1-7.

(124) Nandy, A.; Duan, C.; Goffinet, C.; Kulik, H. J. New Strategies for Direct Methane-to-Methanol Conversion from Active Learning Exploration of 16 Million Catalysts. *JACS Au* **2022**, *2*, 1200-1213.

(125) Nandy, A.; Kulik, H. J. Why Conventional Design Rules for C–H Activation Fail for Open-Shell Transition-Metal Catalysts. *ACS Catal.* **2020**, *10*, 15033–15047.

(126) Duan, C.; Janet, J. P.; Liu, F.; Nandy, A.; Kulik, H. J. Learning from Failure: Predicting Electronic Structure Calculation Outcomes with Machine Learning Models. *J. Chem. Theory Comput.* **2019**, *15*, 2331–2345.

(127) Nandy, A.; Duan, C.; Janet, J. P.; Gugler, S.; Kulik, H. J. Strategies and Software for Machine Learning Accelerated Discovery in Transition Metal Chemistry. *Ind. Eng. Chem. Res.* **2018**, *57*, 13973–13986.

(128) Wang, Z.; Thompson, B. Is the Pearson R 2 Biased, and If So, What Is the Best Correction Formula? *The Journal of Experimental Education* **2007**, *75*, 109–125.

(129) Ezekiel, M. The Sampling Variability of Linear and Curvilinear Regressions: A First Approximation to the Reliability of the Results Secured by the Graphic "Successive Approximation" Method. *Ann. Math Stat.* **1930**, *1*, 275-333.

(130) Ezekiel, M. The Application of the Theory of Error to Multiple and Curvilinear Correlation. *J. Am. Stat. Assoc.* **1929**, *24*, 99-104.

(131) Fisher, R. A.; Russell, E. J. Iii. The Influence of Rainfall on the Yield of Wheat at Rothamsted. *Philos. Trans. R. Soc., B* **1923**, *213*, 89-142.

(132) Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. Array Programming with Numpy. *Nature* **2020**, *585*, 357-362.

(133) Haga, M.-a.; Matsumura-Inoue, T.; Shimizu, K.; Satô, G. P. Ligand Additivity in the Oxidation Potentials of Bidentate Mixed-Ligand Ruthenium(Ii) Complexes. *J. Chem. Soc., Dalton Trans.* **1989**, DOI:10.1039/DT9890000371 10.1039/DT9890000371, 371-373.

(134) Kramida, A.; Ralchenko, Y.; Reader, J.; Team, a. N. A. Nist Atomic Spectra Database (Version 5.10). 2022. https://physics.nist.gov/asd. (Accessed November 21, 2023).

(135) Ekberg, J. O. Wavelengths and Transition Probabilities of the 3d6-3d54p and 3d54s-3d54p Transition Arrays of Fe Iii. *Astron. Astrophys., Suppl. Ser.* **1993**, *101*, 1-36.

(136) DeYonker, N. J.; Peterson, K. A.; Steyl, G.; Wilson, A. K.; Cundari, T. R. Quantitative Computational Thermochemistry of Transition Metal Species. *J. Phys. Chem. A* **2007**, *111*, 11269-11277.

(137) Manivasagam, S.; Laury, M. L.; Wilson, A. K. Pseudopotential-Based Correlation Consistent Composite Approach (Rp-Ccca) for First- and Second-Row Transition Metal Thermochemistry. *J. Phys. Chem. A* **2015,** *119*, 6867-6874.

(138) Janet, J. P.; Kulik, H. J. Resolving Transition Metal Chemical Space: Feature Selection for Machine Learning and Structure–Property Relationships. *J. Phys. Chem. A* **2017,** *121*, 8939-8954.

(139) Vennelakanti, V.; Nandy, A.; Kulik, H. J. The Effect of Hartree-Fock Exchange on Scaling Relations and Reaction Energetics for C–H Activation Catalysts. *Top. Catal.* **2022,** *65*, 296-311.

(140) Gani, T. Z. H.; Kulik, H. J. Understanding and Breaking Scaling Relations in Single-Site Catalysis: Methane to Methanol Conversion by $Fe^{iv}=O$. *ACS Catal.* **2018,** *8*, 975-986.

(141) Harvey, J. N. In *Principles and Applications of Density Functional Theory in Inorganic Chemistry I*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2004; Vol. 112,151-184

(142) Mandal, D.; Shaik, S. Interplay of Tunneling, Two-State Reactivity, and Bell–Evans–Polanyi Effects in C–H Activation by Nonheme Fe(Iv)O Oxidants. *J. Am. Chem. Soc.* **2016,** *138*, 2094-2097.

(143) Massie, A. A.; Sinha, A.; Parham, J. D.; Nordlander, E.; Jackson, T. A. Relationship between Hydrogen-Atom Transfer Driving Force and Reaction Rates for an Oxomanganese(Iv) Adduct. *Inorg. Chem.* **2018,** *57*, 8253-8263.

43

For Table of Contents Use Only