**Electronic Supplementary Information for**

# *Detection of multi-reference character imbalances enables a transfer learning approach for chemical discovery with coupled cluster accuracy at DFT cost*

Chenru Duan[1,2], Daniel B. K. Chu[1], Aditya Nandy[1,2], and Heather J. Kulik[1,*]

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

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

**Contents**

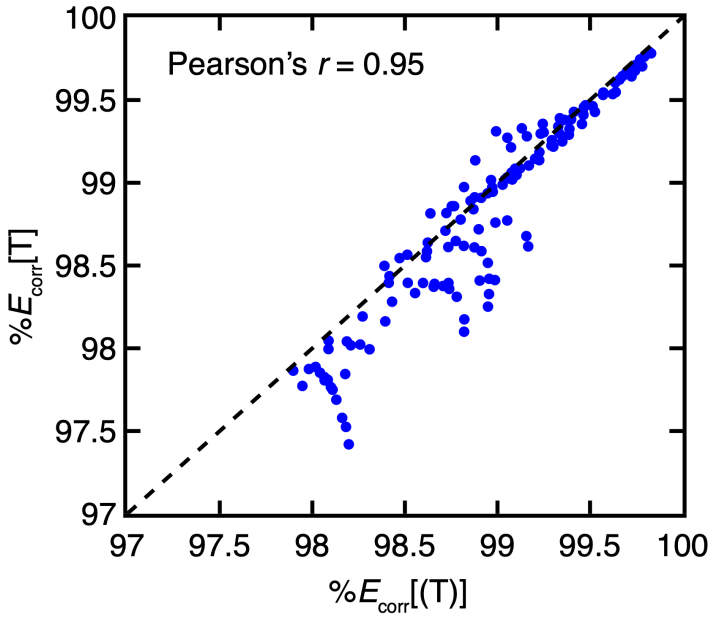


**Figure S1.** $\%E_{corr}[T]$ vs. $\%E_{corr}[(T)]$ for 132 complexes with six helium atoms as ligands at various metal–helium bond lengths. We find four clusters that do not lie on the parity line, which are LS Co(II)$He_6$, IS Fe(III)$He_6$, IS Mn(II)$He_6$, and IS Cr(II)$He_6$ at varied metal–He distances.

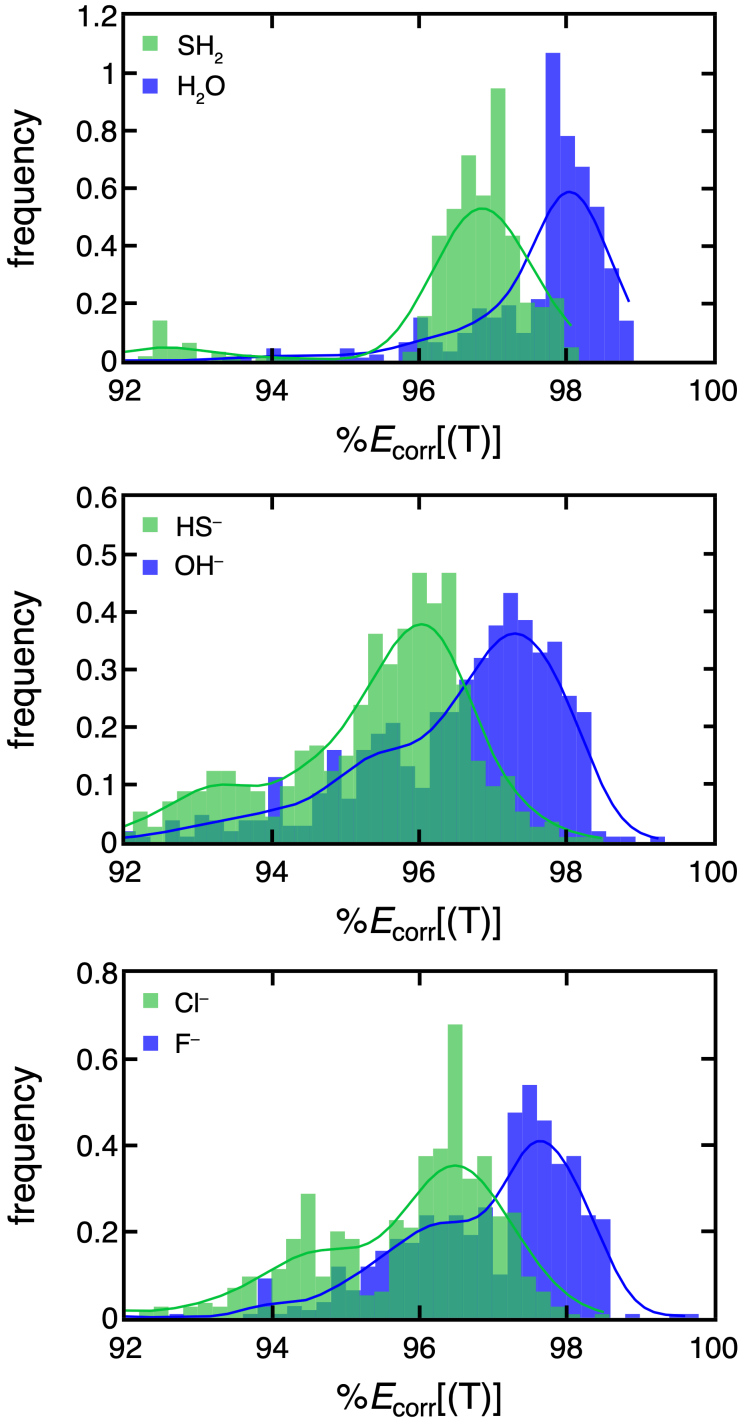


**Figure S2.** Distribution of $\%E_{corr}[(T)]$ for TMCs with isovalent 2*p*- and 3*p*-coordinating ligands. (top) $H_2O$ (blue) and $H_2S$ (green); (middle) $OH^-$ (blue) and $SH^-$ (green); (bottom) $F^-$ (blue) and $Cl^-$ (green). TMCs with both one and two non-He ligands are shown together in the distributions.

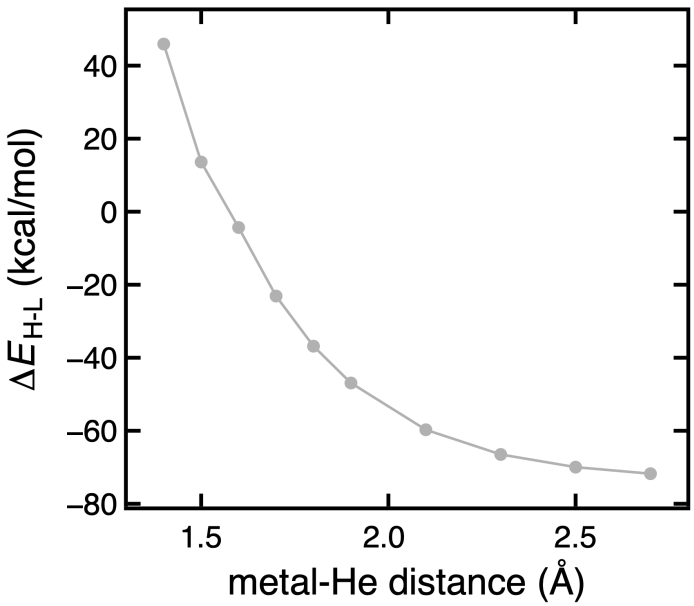


**Figure S3.** $\Delta E_{\text{H-L}}$ for Fe(II)(CO)$(\text{He})_5$ versus metal–helium distance in Å. The Fe-C distance of the carbonyl ligand is relaxed freely while the metal–helium distance is constrained. $\Delta E_{\text{H-L}}$ was computed with CCSD(T) at a mixed basis set with cc-pVTZ on Fe and cc-pVDZ on C, O, and He.

**Table S1.** Summary of MR diagnostics grouped by type and method used. Compared to our prior work[1], $D_1$ and $D_2$ diagnostics are not calculated due to their high linear correlations with some of the other diagnostics ($D_1$ and $T_1$, $D_2$ and $C_0^2$).

| Diagnostic | Method | Type | Extended description |
|---|---|---|---|
| $B_1$[2] | DFT | TAE | Differences in total atomization energy for BLYP and B1LYP (25% exchange) divided by number of pairs of bonded atoms |
| $A_{25}$[PBE][3] | DFT | TAE | 4x the difference in TAE[PBE] and TAE[PBE0] (25% exchange) divided by TAE[PBE] |
| $I_{ND}$[PBE][4-5] | DFT | occupations | Estimation of non-dynamical contribution from finite-temperature DFT with PBE functional (T = 5000 K) |
| $r_{ND}$[PBE][6] | DFT | occupations | ratio of FT-DFT $I_{ND}$ from PBE to the sum of $I_{ND}$ with the dynamical term, $I_D$ |
| $I_{ND}$[B3LYP][4-5] | DFT | occupations | Estimation of non-dynamical contribution from finite-temperature DFT with B3LYP functional (T = 9000 K) |
| $r_{ND}$[B3LYP][6] | DFT | occupations | ratio of FT-DFT $I_{ND}$ from B3LYP to the sum of $I_{ND}$ with the dynamical term, $I_D$ |
| $n_{HOMO}$[MP2][3, 7] | MP2 | occupations | MP2 highest occupied natural orbital occupation |
| $n_{LUMO}$[MP2][3, 7] | MP2 | occupations | MP2 lowest unoccupied natural orbital occupation |
| $T_1$[8] | CCSD | excitations | Frobenius norm of the single-excitation amplitude vector normalized by the square root of the number of electrons in CCSD |
| max($t_1$)[9] | CCSD | excitations | The largest eigenvalue of the matrix derived from the single-excitation amplitudes. |
| %TAE[(T)][10] | CCSD(T) | TAE | Percent difference in TAE from CCSD vs. CCSD(T) |
| $C_0^2$[14][8, 11-13] | CASSCF | occupations | CASSCF leading coefficient CSF at an active space of 14 orbitals |
| $n_{HOMO}$[14][3, 14] | CASSCF | occupations | CASSCF highest occupied natural orbital occupation at an active space of 14 orbitals |
| $n_{LUMO}$[14][3, 14] | CASSCF | occupations | CASSCF lowest unoccupied natural orbital occupation at an active space of 14 orbitals |

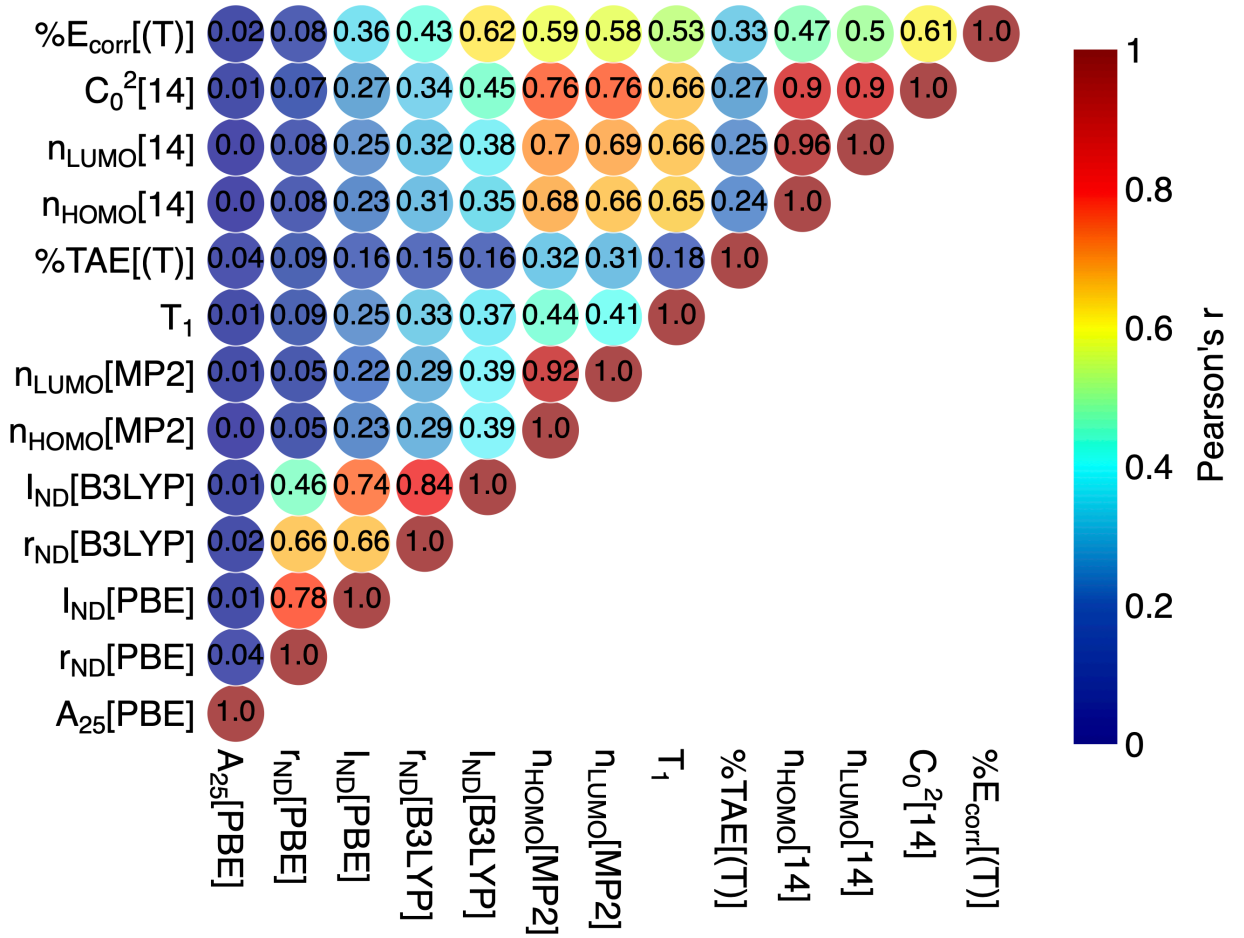


**Figure S4.** An upper triangular matrix of unsigned Pearson's *r* for pairs of MR diagnostics and $\%E_{corr}[(T)]$ on the set of more than 10,000 TMCs. For each pair, the circle is colored by the unsigned Pearson's *r* and the *r* value is explicitly shown. Fourteen active orbitals are used for the CASSCF calculations

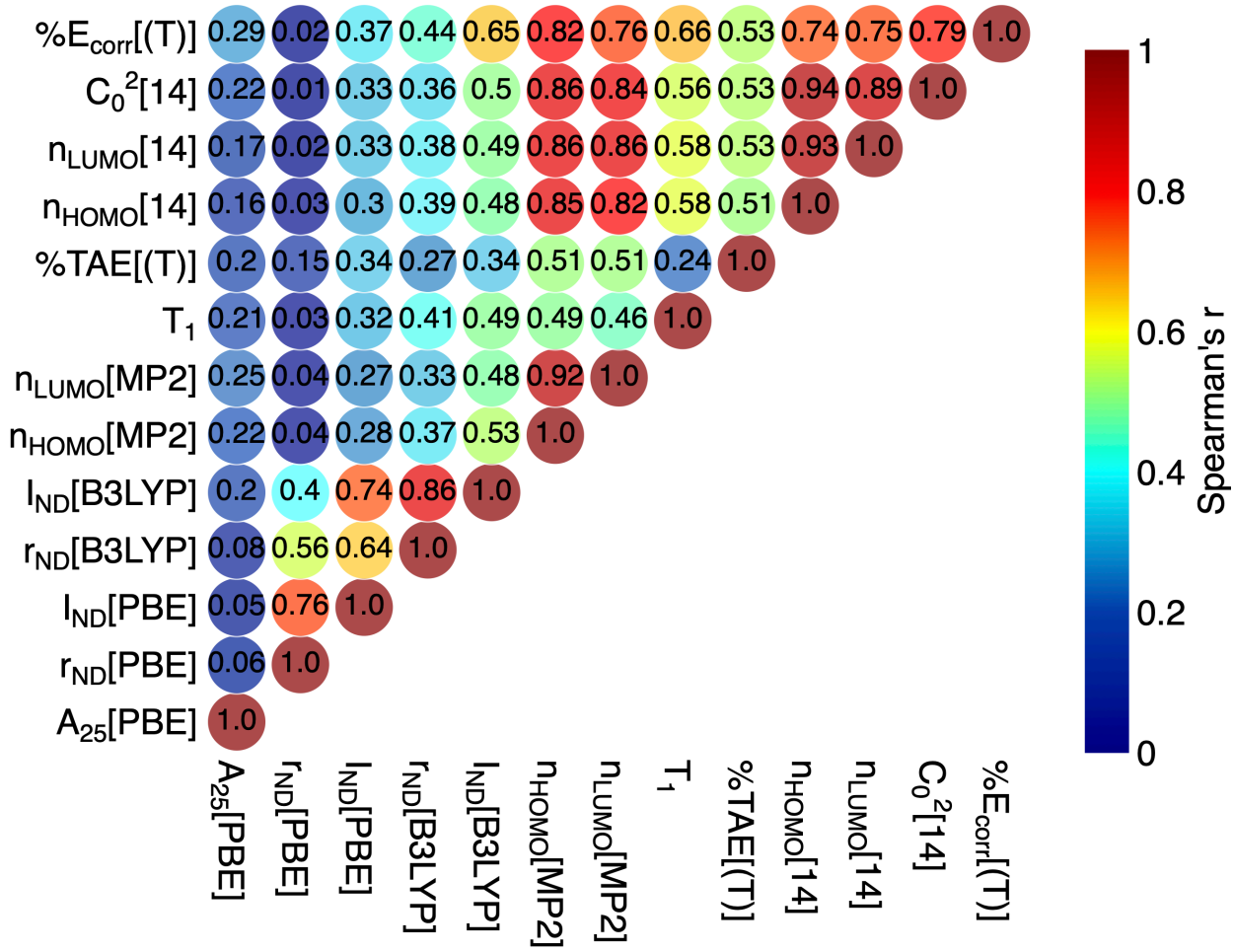


**Figure S5.** An upper triangular matrix of Spearman's *r* for pairs of MR diagnostics and %$E_{corr}$[(T)] on the set of more than 10,000 TMCs. For each pair, the circle is colored by the Spearman's *r* and the *r* value is explicitly shown. Fourteen active orbitals are used for the CASSCF calculations

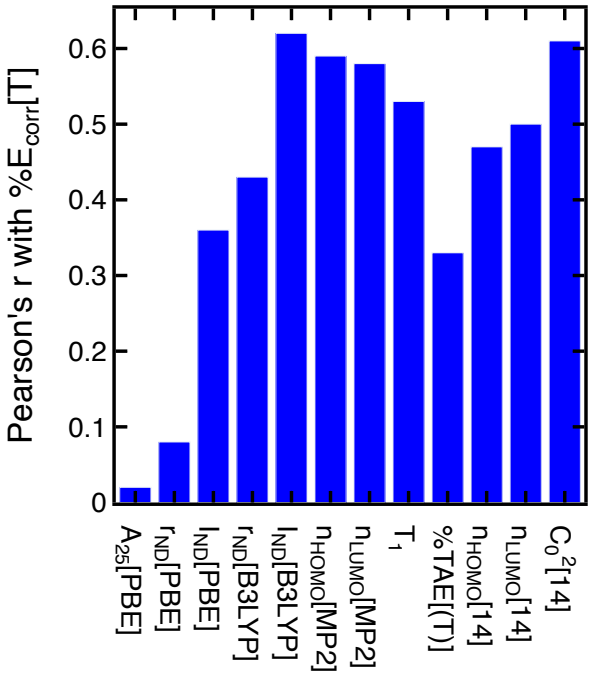


**Figure S6.** Bar plot of unsigned Pearson's *r* for %$E_{corr}$[(T)] with different MR diagnostics on more than 10,000 TMCs. Fourteen active orbitals are used for the CASSCF calculations

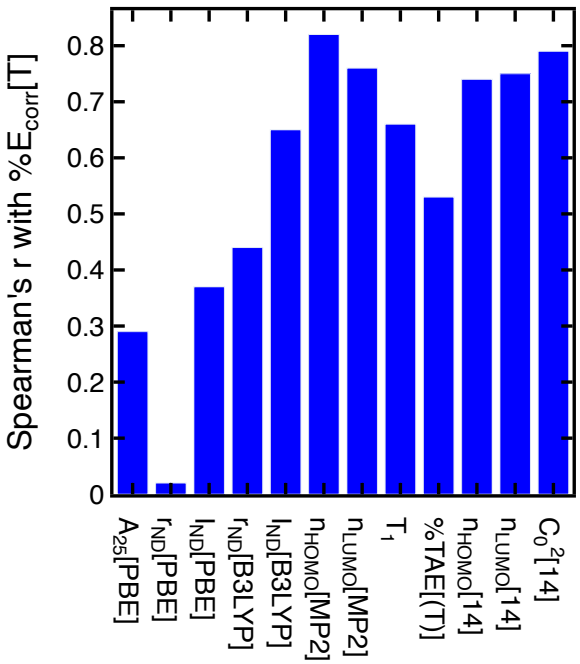


**Figure S7.** Bar plot of Spearman's *r* for %$E_{corr}$[(T)] with different MR diagnostics on more than 10,000 TMCs. Fourteen active orbitals are used for the CASSCF calculations

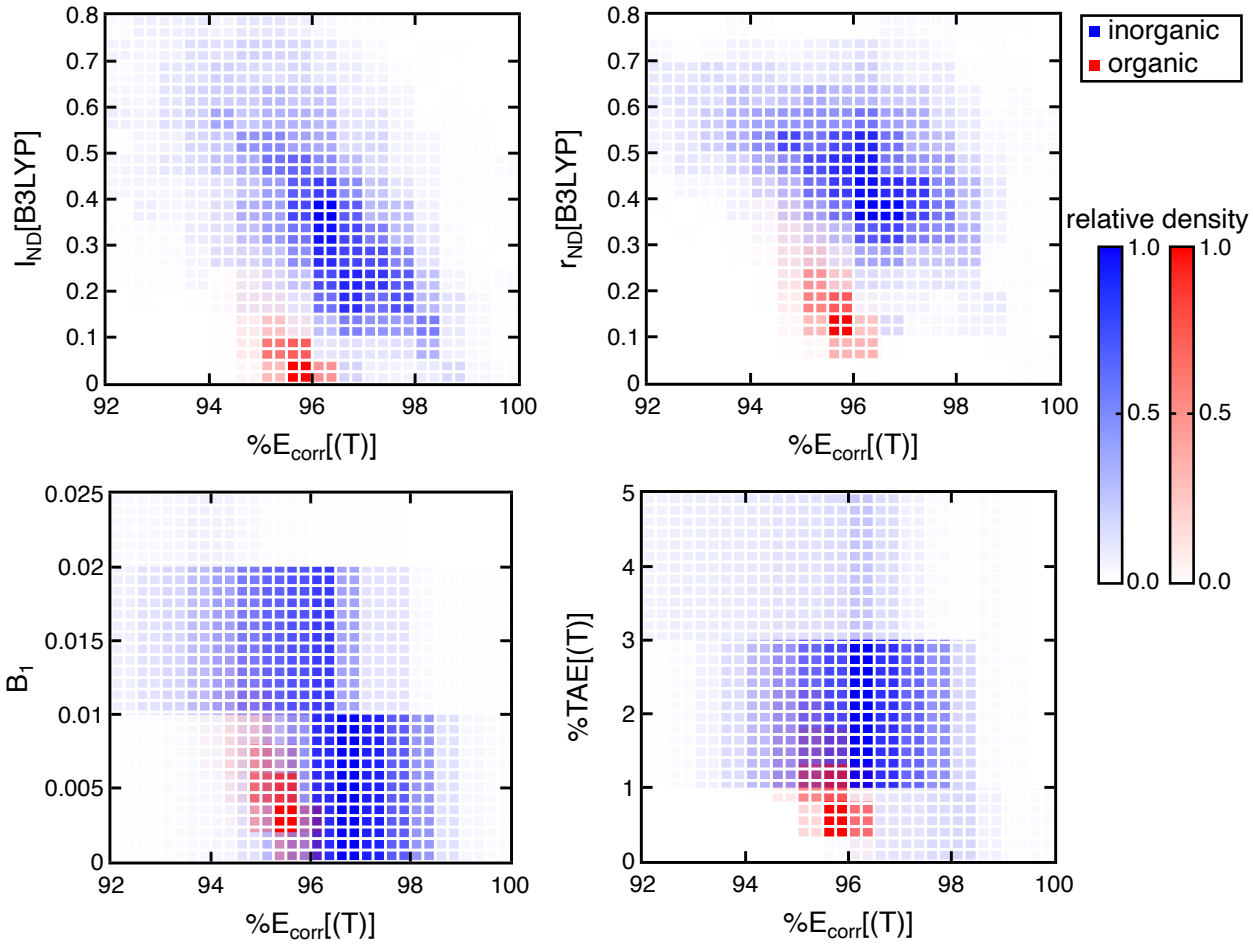


**Figure S8.** 2D histogram for %$E_{corr}$[(T)] vs. $I_{ND}$[B3LYP] (top left), $r_{ND}$[B3LYP] (top right), $B_1$ (bottom left), and %TAE[(T)] (bottom right) for more than 10,000 TMCs in this work (blue) and for the 12,500 equilibrium or distorted organic molecules in our prior work[1] (AD-3165, PS-401, and LG-8934, red). The relative density of systems lying at a specific bin is represented by the opacity of the coloring, as shown in the color bars at right.

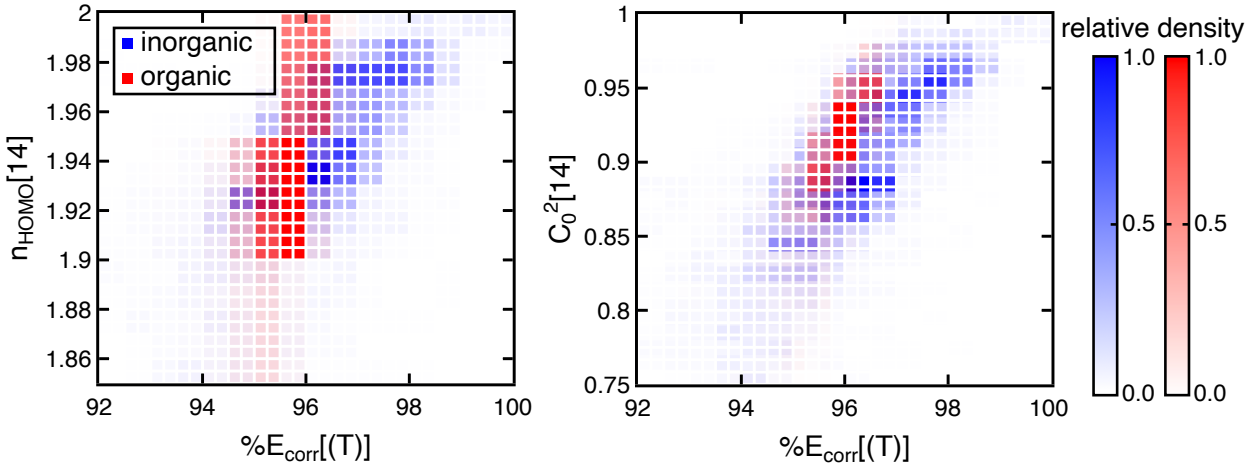


**Figure S9.** 2D histogram for %$E_{corr}$[(T)] vs. $n_{HOMO}$[14] (left) and $C_0^2$[14] (right) for the 10,000 TMCs in this work (blue) and for the 3566 equilibrium or distorted organic molecules in our prior work[1] (AD-3165 and PS-401, red). The relative density of systems lying at a specific bin is represented by the opacity of the coloring. Fourteen active orbitals are used for the CASSCF calculations. Note that LG-8934 is not included in the comparison since WFT-based diagnostics were not computed in that set.

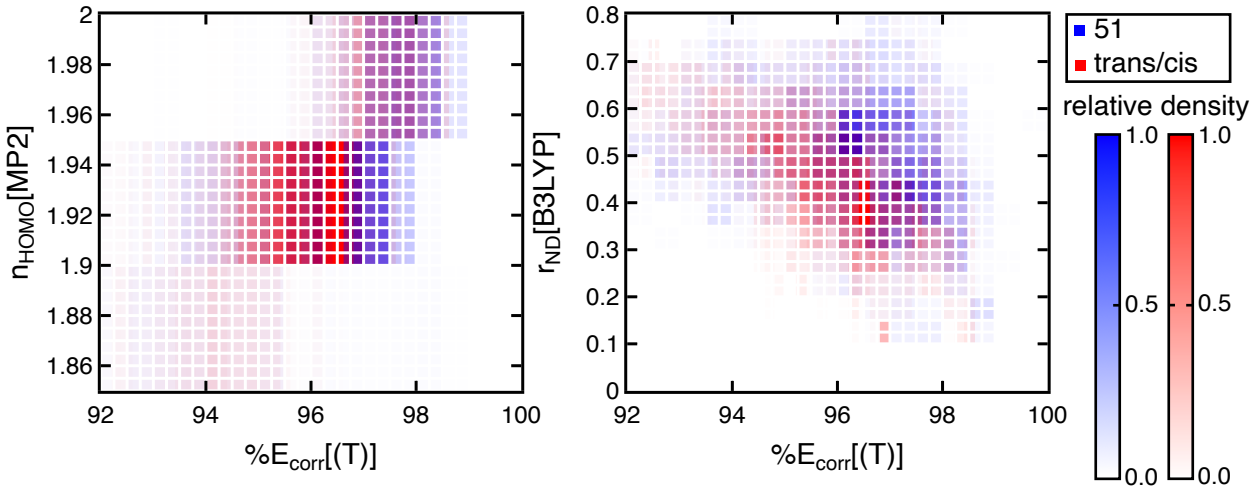


**Figure S10.** 2D histogram for %$E_{corr}$[(T)] vs. $n_{HOMO}$[MP2] (left) and $r_{ND}$[B3LYP] (right) for more than 10,000 TMCs in this work. The 51 complexes with only one non-He ligand are shown in blue, and the *trans* and *cis* complexes (i.e., with two non-He ligands) are shown in red. The 51 complexes have a smaller overall size than the *trans* and *cis* complexes. The relative density of systems lying at a specific bin is represented by the opacity of the coloring.

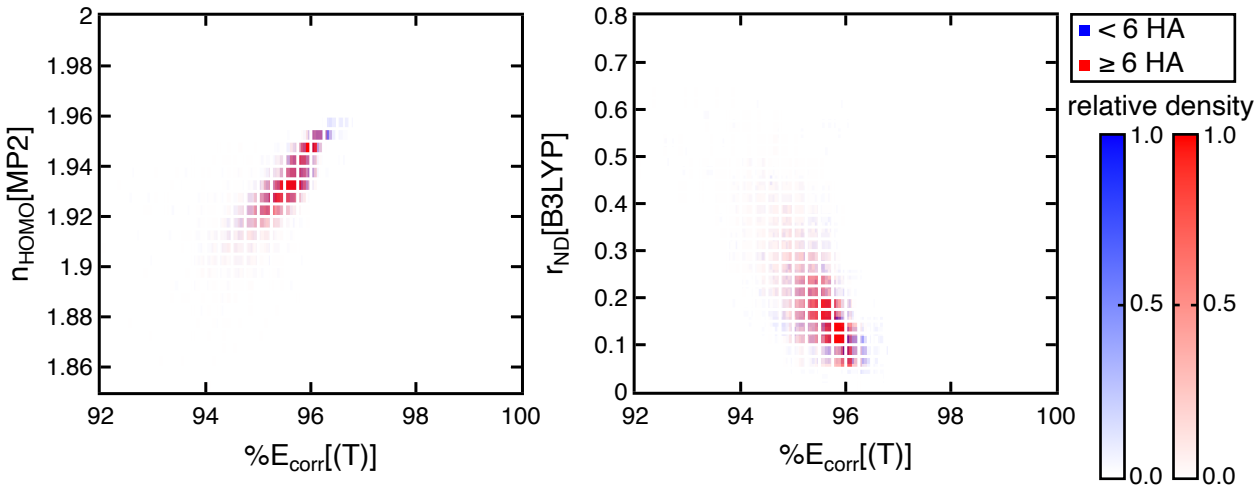


**Figure S11.** 2D histogram for $\%E_{corr}$[(T)] vs. $n_{HOMO}$[MP2] (left) and $r_{ND}$[B3LYP] (right) for the 12,500 equilibrium or distorted organic molecules in our prior work[1], with the 3566 smaller molecules with < 6 heavy atoms (HA) shown in blue and 8934 larger molecules with ≥ 6 HAs shown in red. The relative density of systems lying at a specific bin is represented by the opacity of the coloring.

**Table S2.** Summary of each ANN model's performance on predicting WFT-based MR diagnostics and $\%E_{corr}$[(T)] on a set-aside test set of 2,000 TMCs. Each line corresponds to an ANN trained to independently predict the quantity listed (i.e., we trained independent ANN model to predict each WFT-based diagnostic).

| | MAE | scaled MAE | Pearson's *r* |
|---|---|---|---|
| $n_{HOMO}$[MP2] | 0.009 | 0.013 | 0.88 |
| $n_{LUMO}$[MP2] | 0.009 | 0.010 | 0.91 |
| $n_{HOMO}$[14] | 0.027 | 0.027 | 0.85 |
| $n_{LUMO}$[14] | 0.033 | 0.024 | 0.81 |
| $C_0^2$[14] | 0.019 | 0.023 | 0.90 |
| %TAE[(T)] | 0.348 | 0.007 | 0.90 |
| $T_1$ | 0.006 | 0.020 | 0.89 |
| max($t_1$) | 0.048 | 0.025 | 0.80 |
| $\%E_{corr}$[(T)] | 0.211 | 0.016 | 0.94 |

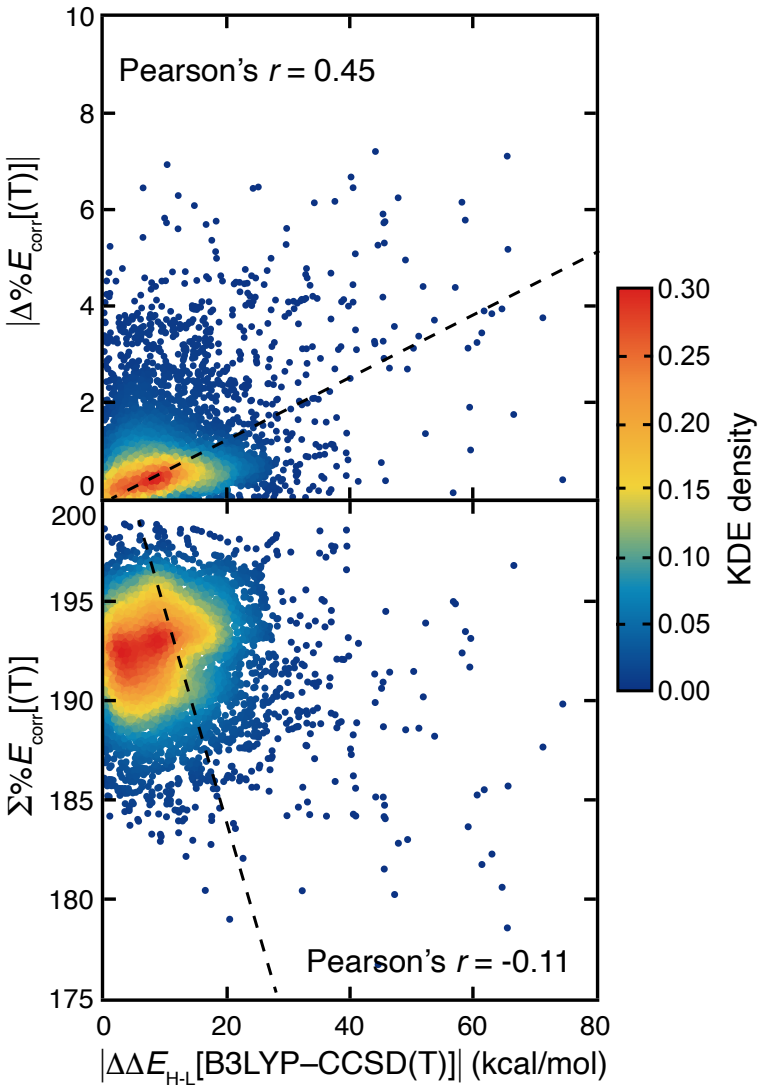


**Figure S12.** The absolute difference of adiabatic spin splitting between B3LYP and CCSD(T) (i.e., $|\Delta\Delta E_{\text{H-L}}[\text{B3LYP-CCSD(T)}]|$) vs. the absolute difference (top) and the sum (bottom) of $\%E_{\text{corr}}[(\text{T})]$ of the two spin states, colored by kernel density estimation (KDE) density values, as indicated by inset color bars. A black dashed linear regression line is shown in each case together with the Pearson correlation coefficients.

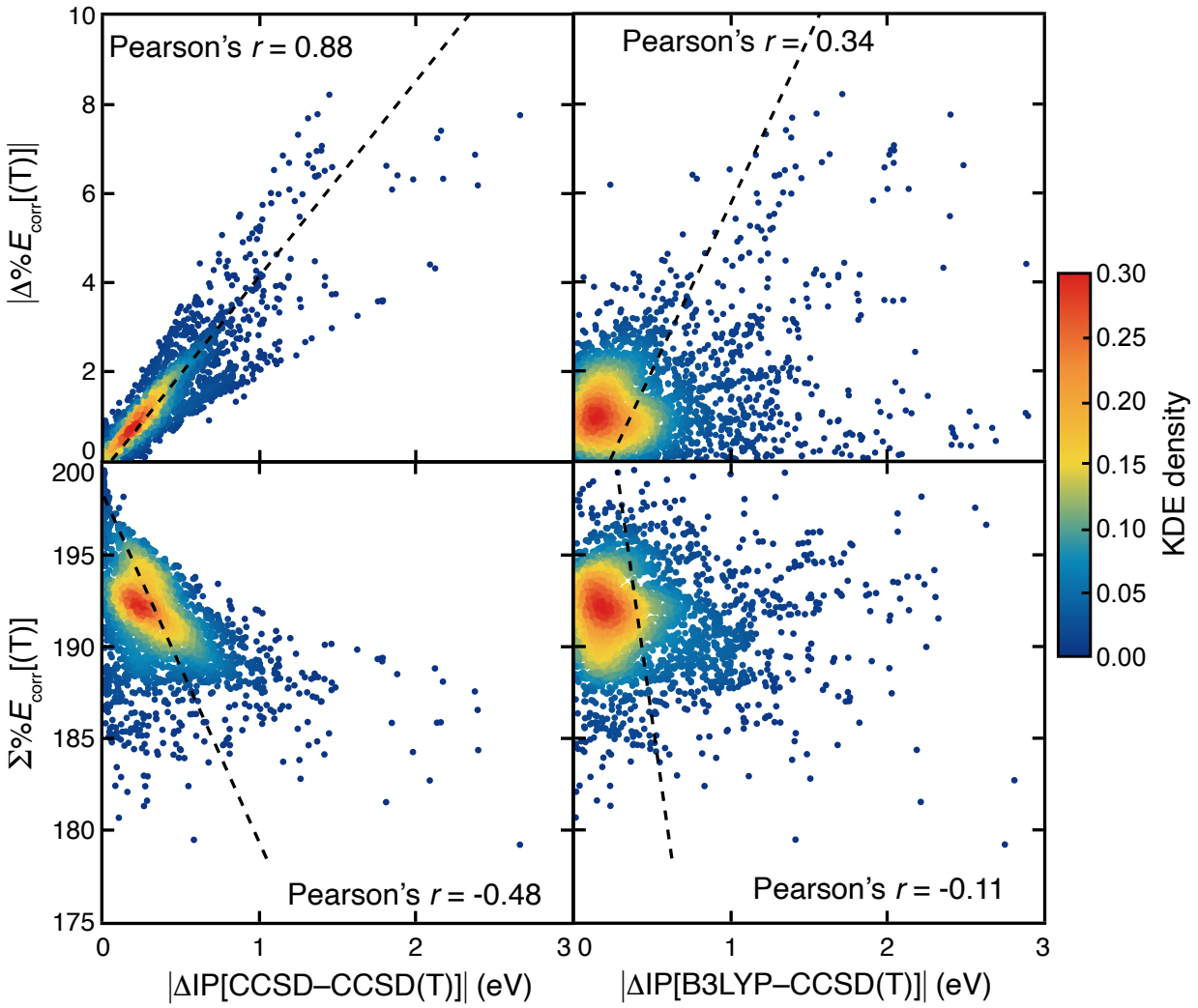


**Figure S13.** (left) The absolute difference of adiabatic IP between CCSD and CCSD(T) (i.e., |ΔIP[CCSD-CCSD(T)]|) vs. the absolute difference (top) and the sum (bottom) of %$E_{corr}$[(T)] of the two charge states. (right) The absolute difference of adiabatic IP between B3LYP and CCSD(T) (i.e., |ΔIP[B3LYP-CCSD(T)]|) vs. the absolute difference (top) and the sum (bottom) of %$E_{corr}$[(T)] of the two charge states. In both cases, points are colored by kernel density estimation (KDE) density values, as indicated by inset color bars. A black dashed linear regression line is also shown in each case together with the Pearson correlation coefficients.

**Table S3.** Features used for each prediction task.

| Target | Features |
|---|---|
| WFT-based diagnostics | CD-RACs, DFT-based diagnostics, oxidation state, spin state, and ligand charge |
| $\Delta\Delta E_{H\text{-}L}$ | CD-RACs from complexes in two spin states, oxidation state, and ligand charge, sums and differences of six DFT-based MR diagnostics, DFT evaluated $\Delta E_{H\text{-}L}$ with BLYP, B3LYP, PBE, and PBE0 |
| $\Delta$IP | CD-RACs from complexes in two oxidation states, spin state of the ox-II complex, and ligand charge, sums and differences of six DFT-based MR diagnostics, DFT evaluated $\Delta E_{H\text{-}L}$ with BLYP, B3LYP, PBE, and PBE0 |

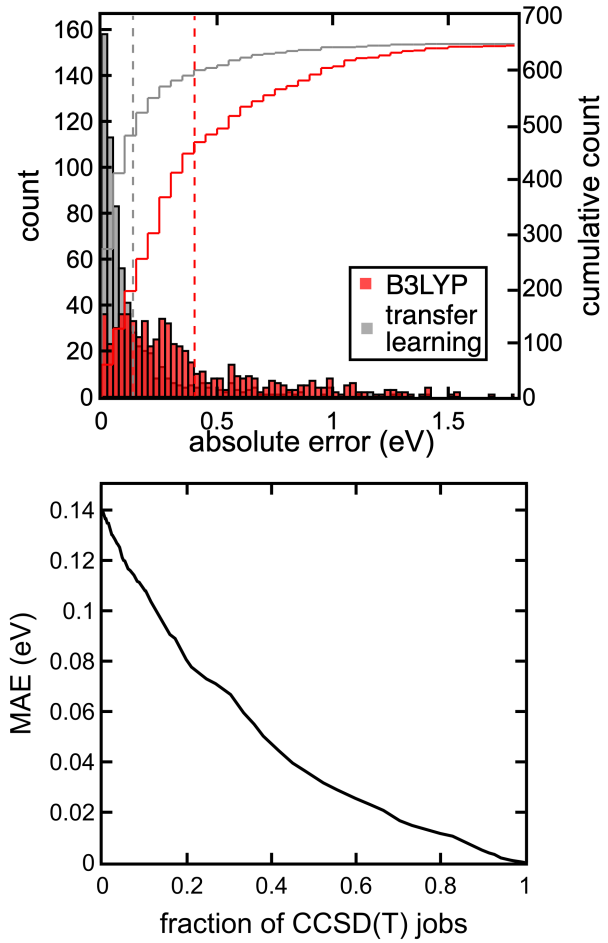


**Figure S14.** (top) Distributions of absolute errors for IP predicted with DFT using B3LYP (red) and our transfer learning models (gray) on the set-aside test data, with the cumulative count shown according to the axis on the right. The MAEs are shown as vertical bars at 0.40 eV for DFT and 0.14 eV for transfer learning. (bottom) MAE of our multi-pronged strategy of transfer learning, uncertainty quantification, and multi-level modeling vs. the percentage of CCSD(T) calculations required. In both cases, we treat CCSD(T) results as our reference.

**Table S4.** MAE of $\Delta E_{H\text{-}L}$ and IP for B3LYP, MP2, CCSD, and their corresponding transfer learning approaches on the set-aside test data (1355 points for $\Delta E_{H\text{-}L}$ and 657 points for IP). In both cases, CCSD(T) is treated as the reference. The transfer learning approach leads to 5- to 10-fold reductions in MAE and systematically improves with the level of the quantum chemistry method. The inputs of the models are described in Table S3 and all the models as well as hyperparameters are included in the zip file of the Supporting Information.

| | MAE of $\Delta E_{H\text{-}L}$ (kcal/mol) | MAE of IP (eV) |
|---|---|---|
| B3LYP | 10.2 | 0.40 |
| DFT-cost transfer learning | 2.8 | 0.14 |
| MP2 | 10.4 | 0.61 |
| MP2-cost transfer learning | 2.1 | 0.12 |
| CCSD | 5.6 | 0.35 |
| CCSD-cost transfer learning | 0.4 | 0.06 |

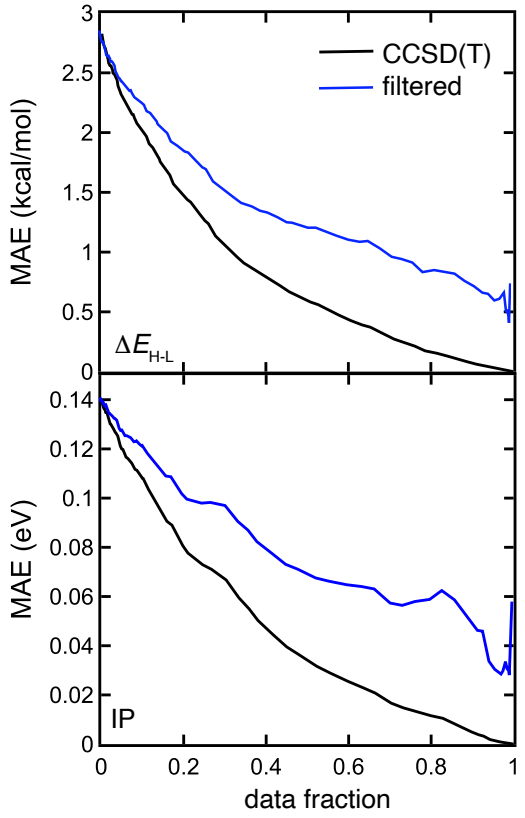


**Figure S15.** MAE of $\Delta E_{H\text{-}L}$ (top) and IP (bottom) for the multi-pronged strategy where the highest UQ points are explicitly computed by CCSD(T) (black) compared to those are directly filtered (i.e., eliminated) from the set-aside test data[15]. The x axis refers to the fraction of data considered to have high model uncertainty judged by a cutoff and are thus calculated or excluded. In the multi-pronged strategy, we perform CCSD(T) on this fraction of data and thus its error is counted as zero. In the direct filtering strategy, this fraction of data is removed from the test data. Note that if we randomly select points in the multi-pronged filtering strategy (instead of using UQ), the MAE would not decrease with respect to the data fraction (i.e., 2.8 kcal/mol $\Delta E_{H\text{-}L}$ and 1.4 eV for IP) on average (i.e., if we repeat this random process many times independently).

**Table S5.** Metals (M), oxidation states (ox), and spin states considered in this work. Cases where the high-spin state is not calculated are shown with "--".

| *d* electron configuration | M(ox) | High-spin state | Intermediate-spin state | Low-spin state |
|---|---|---|---|---|
| $d^3$ | Cr(III) | -- | quartet | doublet |
| $d^4$ | Mn(III)/Cr(II) | quintet | triplet | singlet |
| $d^5$ | Fe(III)/Mn(II) | sextet | quartet | doublet |
| $d^6$ | Co(III)/Fe(II) | quintet | triplet | singlet |
| $d^7$ | Co(II) | -- | quartet | doublet |

**Table S6.** Summary of ligands studied. The ligands are either from the spectrochemical series (spectro) or our previous OHLDB set[16]. For each ligand, the atoms with non-zero formal charges are also shown for simplicity.

| Chemical name | Formula | SMILES string | Source | Formal charge | Connecting atom |
|---|---|---|---|---|---|
| amine | $NH_2^-$ | [NH2-] | spectro | [("N": -1)] | N |
| ammonia | $NH_3$ | [NH3] | spectro | [] | N |
| phosphide | $PH_2^-$ | [PH2-] | spectro | [("P": -1)] | P |
| phosphine | $PH_3$ | [PH3] | spectro | [] | P |
| azide | $N_3^-$ | [N-]=[N+]=[N-] | spectro | [("N": -1), ("N": 1), ("N": -1)] | N |
| carbonyl | CO | [C-]#[O+] | spectro | [("C": -1), ("O": 1)] | C |
| chloride | $Cl^-$ | [Cl-] | spectro | [("Cl": -1)] | Cl |
| fluoride | $F^-$ | [F-] | spectro | [("F": -1)] | F |
| cyanide | $CN^-$ | [C-]#N | spectro | [("C": -1)] | S |
| hydrogen sulfide | $H_2S$ | [SH2] | spectro | [] | S |
| hydrosulfide | $HS^-$ | [HS-] | spectro | [("S": -1)] | O |
| water | $H_2O$ | [OH2] | spectro | [] | O |
| hydroxyl | $OH^-$ | [OH-] | spectro | [("O": -1)] | N |
| isothiocyanate | NCS- | [N-]=C=S | spectro | [("N": -1)] | S |
| thiocyanate | SCN- | [S-]-C#N | spectro | [("S": -1)] | N |
| nitrito | $NO_2^-$ | [O-]-N=O | spectro | [("O": -1)] | N |
| C2 | $C_2$ | C4C | OHLDB | [] | C |
| nitrogen | $N_2$ | N#N | OHLDB | [] | N |
| oxygen | $O_2$ | O#O | OHLDB | [] | O |
| sulfidocarbon | CS | [C-]#[S+] | OHLDB | [("C": -1), ("S": 1)] | C |
| hydrogen cyanide | HCN | [CH]#N | OHLDB | [] | N |
| cyanate | $NCO^-$ | N#C[O-] | OHLDB | [("O": -1)] | N |
| oxoazanide | $NO^-$ | [N-]=O | OHLDB | [("N": -1)] | N |
| sulfanylideneazanide | $NS^-$ | [N-]=S | OHLDB | [("N": -1)] | N |
| formaldehyde | $H_2CO$ | O=[CH2] | OHLDB | [] | O |

**Table S7.** Type of TMCs and their DFT geometry optimization success rate. Metal–He distances are studied in 0.1 Å intervals from 1.4 to 1.9 Å and 0.2 Å intervals from 2.1 to 2.7 Å

| Type of TMCs | all-He | 1-non-He | trans | cis |
|---|---|---|---|---|
| Theoretical size | 220 | 5500 | 5500 | 5500 |
| Successful cases | 220 | 4321 | 4730 | 3577 |
| Success rate | 100% | 79% | 86% | 65% |

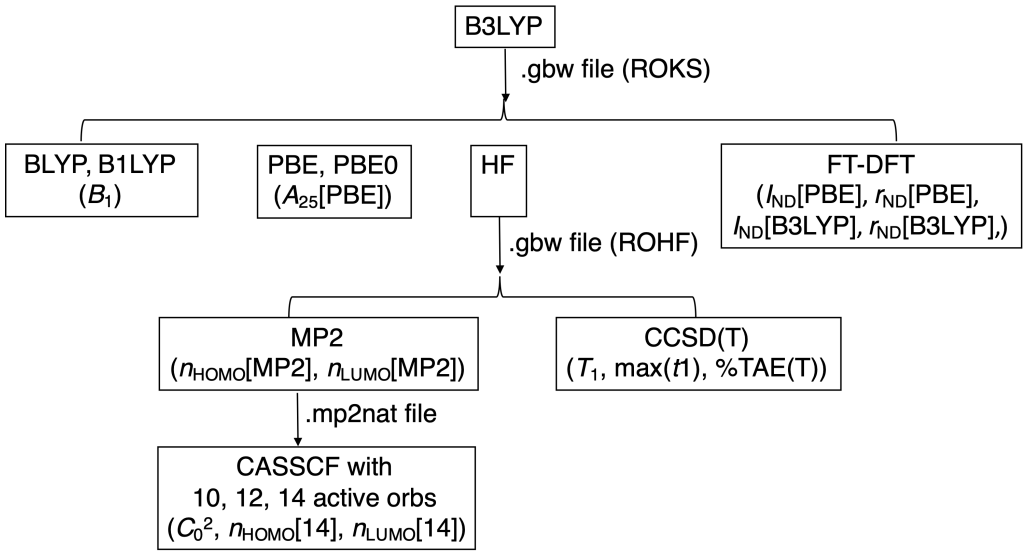


**Figure S16.** Workflow of computing 14 MR diagnostics. All calculations were performed with ORCA 4.0.2.1[17-18].

**Table S8.** Default convergence parameters used in self-consistent field calculations (i.e., HF, DFT, and CCSD(T)) for ORCA 4.0.2.1[17-18].

| Software | Energy convergence threshold (Ha) | DIIS error (Ha) | Maxiter |
|---|---|---|---|
| ORCA | 1e-6 | 1e-6 | 125 |

**Table S9.** MR diagnostics calculation attrition counts and reasons.

| Type | Count | Reason |
|---|---|---|
| Zero-temperature DFT (B3LYP, BLYP, B1LYP, PBE, and PBE0) | 983 | SCF convergence issue |
| Finite-temperature DFT (B3LYP and PBE) | 151 | SCF convergence issue |
| CASSCF at an active space of 14 orbitals | 424 | CASSCF convergence issue or exceeding the time limit of 48 hours. |
| CCSD(T) | 629 | MDCI convergence issue or exceeding the time limit of 48 hours. |

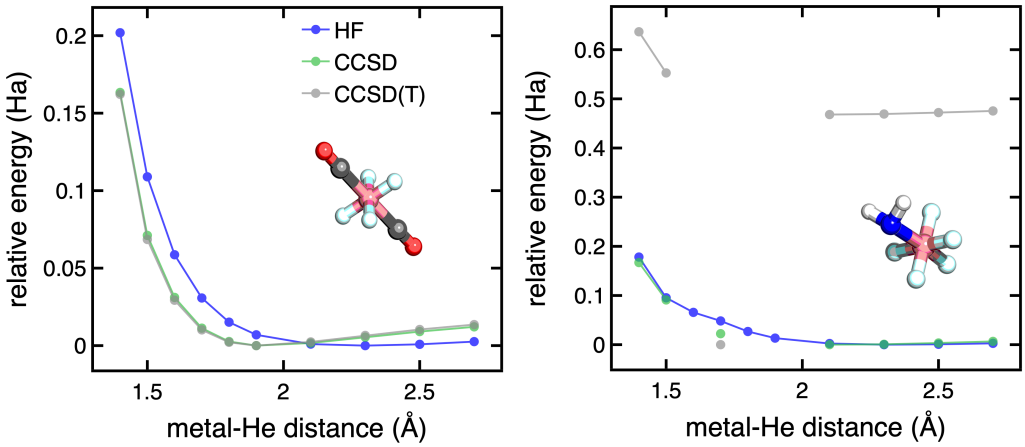


**Figure S17.** Potential energy curve for *trans* LS Co(II)$He_4(CO)_2$ (left) and LS Co(III)$He_5(NH_2^-)$ (right) for HF (blue), CCSD (green), and CCSD(T) (gray) in Ha with increasing metal–He distance. For each method, the energies are shifted such that the minimum is set to 0 Ha. A discontinuity can be seen in LS Co(III)$He_5(NH_2^-)$ at metal–He of 1.7 Å, resulting from the abnormally large perturbative correction at this metal–He distance.

**Table S10.** $T_1$ diagnostics for *trans* LS Co(II)$He_4(CO)_2$ and LS Co(III)$He_5(NH_2^-)$ with increasing metal–He distances. The $T_1$ diagnostic of *trans* LS Co(II)$He_4(CO)_2$ at a metal–He distance of 1.7 Å is bolded. This discontinuity in $T_1$ diagnostic suggests the CCSD wavefunction was not converged to the same electronic structure at this particular metal–He distance. In cases where the CCSD calculation did not converge, the $T_1$ diagnostic is shown as "--".

| Metal-He distance (Å) | 1.4 | 1.5 | 1.6 | 1.7 | 1.8 | 1.9 | 2.1 | 2.3 | 2.5 | 2.7 |
|---|---|---|---|---|---|---|---|---|---|---|
| *trans* LS Co(II)$He_4(CO)_2$ | 0.05 | 0.08 | -- | **0.28** | -- | -- | 0.09 | 0.09 | 0.09 | 0.09 |
| LS Co(III)$He_5(NH_2^-)$ | 0.02 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.02 | 0.02 |

**Table S11.** Cutoff values for the Grubbs test and Z-score test. A TMC was removed if it was marked as an outlier by both the Grubbs and Z-score tests.

| Grubbs | Z-score |
|---|---|
| 0.05 | 2.0 |

**Table S12.** Removed TMCs at each filtering step. The theoretical size of the dataset is 16,720.

| Reason | Number of points |
|---|---|
| Bad DFT optimized geometry | 3872 |
| Missing any of the 14 MR diagnostics | 2187 |
| Abnormal perturbative T correction | 274 |
| 14 orbitals not large enough as active space | 334 |

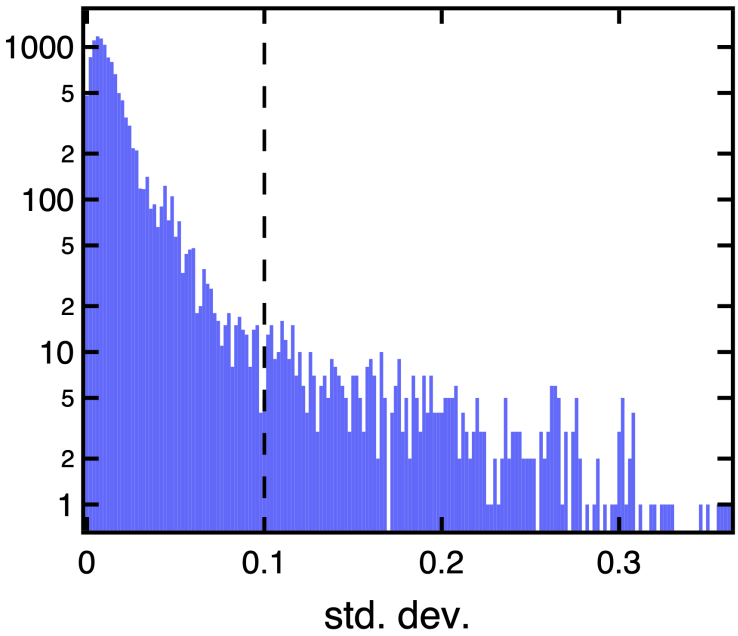


**Figure S18.** Distribution of the standard deviation (std. dev.) for the $C_0^2$ diagnostic obtained with three active spaces (i.e., 10 orbitals, 12 orbitals, and 14 orbitals). The cutoff value of 0.1 is shown by a vertical dashed line. The *y* axis is in log scale.

**Text S1.** We have introduced a systematic approach to featurize molecular inorganic complexes that blends metal-centric and whole-complex topological properties in a feature set referred to as revised autocorrelation functions (RACs).[19] These RACs, variants of graph autocorrelations (ACs),[20-23] are sums of products and differences of atomic properties, i.e., electronegativity ($\chi$), nuclear charge ($Z$), topology ($T$), covalent radius ($S$), and identity ($I$). Standard ACs are defined as

$$P_d = \sum_i \sum_j P_i P_j \delta(d_{ij}, d)$$

where $P_d$ is the AC for property $P$ at depth $d$, $\delta$ is the Dirac delta function, and $d_{ij}$ is the bond-wise path distance between atoms $i$ and $j$.

In our approach, we have five types of RACs:

- ${}^{\mathrm{f}}_{\mathrm{all}}P_d$: standard ACs start on the full molecule (*f*) and have all atoms in the scope (all).
- ${}^{\mathrm{f}}_{\mathrm{ax}}P_d$ and ${}^{\mathrm{f}}_{\mathrm{eq}}P_d$: restricted-*scope* ACs that start on the full molecule (*f*) and separately evaluate axial or equatorial ligand properties

$${}^{\mathrm{f}}_{\mathrm{ax/eq}}P_d = \frac{1}{|\mathrm{ax/eq\ ligands}|} \sum_i^{n_{\mathrm{ax/eq}}} \sum_i^{n_{\mathrm{ax/eq}}} P_i P_j \delta(d_{ij}, d)$$

where $n_{\mathrm{ax/eq}}$ is the number of atoms in the corresponding axial or equatorial ligand and properties are averaged within the ligand subtype.

- ${}^{\mathrm{mc}}_{\mathrm{all}}P_d$: restricted-scope, metal-centered (mc) descriptors that start on the metal center (mc) and have all atoms in the scope (all), in which one of the atoms, *i,* in the *i,j* pair is a metal center:

$$ {}^{\mathrm{mc}}_{\mathrm{all}}P_d = \sum_{i}^{\mathrm{mc}} \sum_{i}^{\mathrm{all}} P_i P_j \delta(d_{ij}, d) $$

- ${}^{\mathrm{lc}}_{\mathrm{ax}}P_d$:and ${}^{\mathrm{lc}}_{\mathrm{ax}}P_d$: restricted-scope, metal-proximal ACs that start on a ligand-centered (lc) and separately evaluate axial or equatorial ligand properties, in which one of the atoms, *i*, in the *i,j* pair is the metal-coordinating atom of the ligand:

$$ {}^{\mathrm{lc}}_{\mathrm{ax/eq}}P_d = \frac{1}{|\mathrm{ax/eq\ ligands}|} \frac{1}{|\mathrm{lc}|} \sum_{i}^{\mathrm{lc}} \sum_{i}^{n_{\mathrm{ax/eq}}} P_i P_j \delta(d_{ij}, d) $$

We also modify the AC definition, *P′*, to property differences rather than products for a minimum depth, d, of 1 (as the depth-0 differences are always zero):

$$ {}^{\mathrm{lc/mc}}_{\mathrm{ax/eq/all}}P'_d = \sum_{i}^{\mathrm{lc\ or\ mc}} \sum_{i}^{\mathrm{scope}} P_i P_j \delta(d_{ij}, d) $$

where scope can be axial, equatorial, or all ligands.

Although RACs have been demonstrated to be accurate for predicting equilibrium properties with ML models, including spin-splitting energies and redox potential[24-25], they encode no explicit 3D geometry information and cannot distinguish distorted geometries from equilibrium structures. Therefore, we combine RACs and a 3D geometry-based representation, the Coulomb matrix (CM), in a new representation we refer to as the Coulomb-decay revised autocorrelations (CD-RACs)[1].

$$ P_{d,\mathrm{CD}} = \frac{1}{n} \begin{cases} \sum_{i}^{n} \sum_{j}^{n} \frac{P_i P_j}{r_{ij}} \delta\left(d_{ij}, d\right), & d > 0 \\ \frac{1}{2} \sum_{i}^{n} P_i^{2.4}, & d = 0 \end{cases} $$

The form of CD-RACs is analogous to RACs, but is simply scaled by the pairwise atom distance when the bond depth is not zero. When the bond depth is zero, we use the power of 2.4 and introduce a pre-factor of 0.5 as in CM.

We calculate CD-RACs at all starting points (lc, mc, and f), all scopes (eq, ax, and all), and consider both the products and differences, which gives 180 CD-RACs in total. Among the CD-RACs, some are constant either due to their nature (e.g., full-complex depth-0 *I* CD-RAC is always 0.5) or small sizes of TMCs in the dataset (e.g., all full-complex depth-3 CD-RACs are 0). Eliminating those constant CD-RACs yields 134 CD-RACs in total.

**Table S13**. Range of hyperparameters sampled for ANN models trained from scratch with Hyperopt[26]. The lists in the architecture row can refer to two or three hidden layers (i.e., the number of items in the list), and the number of nodes in each layer are the elements in the list. The built-in Tree of Parzen Estimator algorithm in Hyperopt was used for the hyperparameter selection process.

| | |
|---|---|
| Architecture | ([128], [256], [512], [128, 128], [256, 256], [512, 512], [128, 128, 128], [256, 256, 256], [512, 512, 512]) |
| L2 regularization | [1e-6, 1] |
| Dropout rate | [0, 0.5] |
| Learning rate | [1e-5, 1e-3] |
| Beta1 | [0.75, 0.99] |
| Batch size | [16, 32, 64, 128, 256, 512] |

**References**


1. Duan, C.; Liu, F.; Nandy, A.; Kulik, H. J., Data-Driven Approaches Can Overcome the Cost– Accuracy Trade-off in Multireference Diagnostics. *J Chem Theory Comput* **2020,** https://dx.doi.org/10.1021/acs.jctc.0c00358 (16), 4373-4387.
2. Schultz, N. E.; Zhao, Y.; Truhlar, D. G., Density functionals for inorganometallic and organometallic chemistry. *Journal of Physical Chemistry A* **2005,** *109* (49), 11127-11143.
3. Fogueri, U. R.; Kozuch, S.; Karton, A.; Martin, J. M. L., A simple DFT-based diagnostic for nondynamical correlation. *Theoretical Chemistry Accounts* **2013,** *132* (1), 1291.
4. Ramos-Cordoba, E.; Salvador, P.; Matito, E., Separation of dynamic and nondynamic correlation. *Physical Chemistry Chemical Physics* **2016,** *18* (34), 24015-24023.
5. Ramos-Cordoba, E.; Matito, E., Local Descriptors of Dynamic and Nondynamic Correlation. *J Chem Theory Comput* **2017,** *13* (6), 2705-2711.
6. Kesharwani, M. K.; Sylvetsky, N.; Kohn, A.; Tew, D. P.; Martin, J. M. L., Do CCSD and approximate CCSD-F12 variants converge to the same basis set limits? The case of atomization energies. *J Chem Phys* **2018,** *149* (15), 154109.
7. Jensen, H. J. A.; Jorgensen, P.; Ågren, H.; Olsen, J., Second-order Moller–Plesset perturbation theory as a configuration and orbital generator in multiconfiguration self-consistent field calculations. *The Journal of Chemical Physics* **1988,** *88* (6), 3834-3839.
8. Lee, T. J.; Taylor, P. R., A Diagnostic for Determining the Quality of Single-Reference Electron Correlation Methods. *Int J Quantum Chem* **1989**, 199-207.
9. Janssen, C. L.; Nielsen, I. M. B., New diagnostics for coupled-cluster and Moller-Plesset perturbation theory. *Chemical Physics Letters* **1998,** *290* (4-6), 423-430.
10. Karton, A.; Daon, S.; Martin, J. M. L., W4-11: A high-confidence benchmark dataset for computational thermochemistry derived from first-principles W4 data. *Chemical Physics Letters* **2011,** *510* (4-6), 165-178.
11. Sears, J. S.; Sherrill, C. D., Assessing the performance of density functional theory for the electronic structure of metal-salens: The d(2)-metals. *Journal of Physical Chemistry A* **2008,** *112* (29), 6741-6752.
12. Sears, J. S.; Sherrill, C. D., Assessing the performance of density functional theory for the electronic structure of metal-salens: The 3d(0)-metals. *Journal of Physical Chemistry A* **2008,** *112* (15), 3466-3477.
13. Langhoff, S. R.; Davidson, E. R., Configuration interaction calculations on the nitrogen molecule. *Int J Quantum Chem* **1974,** *8* (1), 61-72.
14. Tishchenko, O.; Zheng, J. J.; Truhlar, D. G., Multireference model chemistries for thermochemical kinetics. *J Chem Theory Comput* **2008,** *4* (8), 1208-1219.
15. Janet, J. P.; Duan, C.; Yang, T.; Nandy, A.; Kulik, H. J., A quantitative uncertainty metric controls error in neural network-driven chemical discovery. *Chemical Science* **2019**.
16. Gugler, S.; Janet, J. P.; Kulik, H. J., Enumeration of de novo inorganic complexes for chemical discovery and machine learning. *Mol Syst Des Eng* **2020,** *5* (1), 139-152.
17. Neese, F., The ORCA program system. *Wires Comput Mol Sci* **2012,** *2* (1), 73-78.
18. Neese, F., Software update: the ORCA program system, version 4.0. *Wires Comput Mol Sci* **2018,** *8* (1), e1327.
19. Janet, J. P.; Kulik, H. J., Resolving transition metal chemical space: feature selection for machine learning and structure-property relationships. *Journal of Physical Chemistry A* **2017,** *121* (46), 8939-8954.

20. Devillers, J.; Domine, D.; Guillon, C.; Bintein, S.; Karcher, W., Prediction of partition coefficients (log p oct) using autocorrelation descriptors. *SAR QSAR Environ. Res.* **1997,** *7* (1-4), 151-172.
21. Broto, P.; Devillers, J., *Autocorrelation of properties distributed on molecular graphs*. Kluwer Academic Publishers: Dordrecht, The Netherlands: 1990.
22. Virshup, A. M.; Contreras-García, J.; Wipf, P.; Yang, W.; Beratan, D. N., Stochastic voyages into uncharted chemical space produce a representative library of all possible drug-like compounds. *Journal of the American Chemical Society* **2013,** *135* (19), 7296-7303.
23. Broto, P.; Moreau, G.; Vandycke, C., Molecular structures: perception, autocorrelation descriptor and sar studies: system of atomic contributions for the calculation of the n-octanol/water partition coefficients. *European journal of medicinal chemistry* **1984,** *19* (1), 71-78.
24. Janet, J. P.; Ramesh, S.; Duan, C.; Kulik, H. J., Accurate multi-objective design in a space of millions of transition metal complexes with neural-network-driven efficient global optimization. *ACS Central Science* **2020,** *6* (4), 513-524.
25. Nandy, A.; Duan, C.; Janet, J. P.; Gugler, S.; Kulik, H. J., Strategies and Software for Machine Learning Accelerated Discovery in Transition Metal Chemistry. *Industrial & Engineering Chemistry Research* **2018,** *57* (42), 13973-13986.
26. Bergstra, J.; Yamins, D.; Cox, D. D. In *Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms*, Proceedings of the 12th Python in science conference, 2013; pp 13-20.