

Autonomous First-Principles Design of Transition Metal Complexes

by

Naveen Arunachalam

B.S. Chemical Engineering,
California Institute of Technology, 2018

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN CHEMICAL ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2023

© 2023 Naveen Arunachalam. All rights reserved.

The author hereby grants to MIT a nonexclusive, worldwide, irrevocable, royalty-free license to exercise any and all rights under copyright, including to reproduce, preserve, distribute and publicly display copies of the thesis, or release the thesis under an open-access license.

Author
Department of Chemical Engineering
April 18, 2023

Certified by
Heather J. Kulik
Associate Professor of Chemical Engineering
Thesis Supervisor

Accepted by
Patrick S. Doyle
Robert T. Haslam (1911) Professor of Chemical Engineering
Chairman, Committee for Graduate Students

Autonomous First-Principles Design of Transition Metal Complexes

by

Naveen Arunachalam

Submitted to the Department of Chemical Engineering
On April 18, 2023, in partial fulfillment of the
Requirements for the degree of
Doctor of Philosophy in Chemical Engineering

Abstract

Designing novel transition metal complexes (TMCs) holds immense potential for advancing sustainability and chemical synthesis. While high-throughput virtual screening (HTVS) workflows and density functional theory (DFT) have emerged as powerful tools for discovering new TMCs, exhaustive design space exploration remains a formidable challenge due to the combinatorial growth of possible complexes with respect to ligands, metals, oxidation states, and ligand field symmetries. Machine learning models trained on databases of TMC structures and their computationally derived properties offer a promising avenue for rapidly and accurately evaluating the molecular properties of novel TMCs. However, the extrapolation error of these models in new regions of chemical space depends on the chemical space spanned by training data, making it crucial to carefully design training data to optimize the performance of HTVS workflows.

This thesis presents a comprehensive exploration of transition metal complex design, encompassing fundamental chemical insights and HTVS improvements derived from the systematic enhancement of database coverage, such as expanded coverage of ligand field symmetries, metal identities, and similarity to experimentally synthesized complexes. In addition, this work introduces interactive web-based tools for the design of TMCs and metal-organic frameworks, allowing users to explore and provide feedback on machine learning models. These tools facilitate collaboration between researchers and promote the iterative improvement of models by incorporating user feedback, leading to more effective and efficient exploration of the vast chemical space of TMCs.

Thesis Supervisor: Heather J. Kulik

Title: Associate Professor of Chemical Engineering

Contents

1	Introduction	24
1.1	Problem areas.....	25
1.2	Database population: symmetry classes and isovalents.....	27
1.3	Machine learning for property prediction.....	28
1.4	User interaction with machine learning models.....	28
1.5	Thesis overview	29
1.6	References	29
2	Ligand additivity relationships enable efficient exploration of transition metal chemical space	31
2.1	Introduction	31
2.2	Methods.....	33
2.2.1	Curation from the Cambridge Structural Database.....	33
2.2.2	Electronic structure calculations	34
2.3	Results and discussion.....	35
2.3.1	Symmetry classes and theoretical complex space.....	35
2.3.2	Ligand additivity for interpolating properties of transition metal complexes	42
2.3.3	Interpolation of chemical space from experimentally characterized complexes.....	52
2.4	Conclusions	57
2.5	References	59
3	Isovalent transition metal complex structures demonstrate additional trends in chemical space	65
3.1	Introduction	65
3.2	Methods.....	66
3.2.1	Data set construction – 3d/4d	66
3.2.2	Electronic structure calculations – 3d/4d	67
3.2.3	Data set construction and electronic structure calculations – 2p/3p	69

3.3	Results and discussion.....	70
3.3.1	3d vs. 4d transition metal complexes.....	70
3.3.2	Incorporating the effect of 2p vs. 3p ligands.....	72
3.4	Conclusion.....	73
3.5	References	74
4	Accelerating training data curation for autonomous workflows through parallel computing.....	77
4.1	Introduction	77
4.2	Methods.....	79
4.2.1	Bond-Stretching	81
4.2.2	Angular bending.....	82
4.2.3	Dihedral torsion	83
4.2.4	Inversion	83
4.2.5	van der Waals interaction.....	84
4.2.6	Electrostatic Interactions.....	84
4.3	Results and discussion.....	85
4.3.1	Performance analysis of parameter generation time vs. system size	86
4.3.2	Performance analysis of parameter generation time vs. degree of parallelization.....	87
4.3.3	Local optimization of butane using forward-mode automatic differentiation.....	87
4.3.4	Global optimization of butane using Optim.jl	89
4.4	Conclusion.....	90
4.4.1	Future work	91
4.5	References	91
5	Design tools for transition metal complexes and metal organic frameworks allow users to explore and provide feedback on machine learning models	93
5.1	Introduction	93
5.2	Methods.....	94
5.2.1	Frontend and backend components for molSimplify Lite.....	94
5.2.2	Form and database components for MOFSimplify	95
5.3	Results and discussion.....	97
5.3.1	Usage notes: molSimplify Lite.....	97
5.3.2	Usage notes: MOFSimplify	99
5.3.3	Impact.....	100

5.4	Conclusion.....	101
5.5	References	102
6	Ligand compatibility classification using stochastic negative addition: towards interactive investigation of new chemical structures	103
6.1	Introduction	103
6.1.1	Choice of design space.....	104
6.1.2	Classification via stochastic negative addition	106
6.2	Methods.....	107
6.2.1	Cambridge Structural Database search.....	107
6.2.2	Descriptors	108
6.2.3	Machine learning.....	109
6.3	Results and discussion.....	110
6.3.1	Indicators of ligand compatibility	110
6.3.2	Model evaluation.....	112
6.3.3	Usage notes	113
6.4	Conclusion.....	115
6.5	References	116
7	Recommendations for future work	117
7.1	Development and maintenance of fully autonomous workflows	117
7.2	Incorporation of web-based feedback into machine learning models	118
7.3	Expert-informed chemical space exploration	118
A.	Appendix: Ligand Additivity.....	120
	References	143
B.	Appendix: Isovalent TMCs.....	144
	References	170

List of Figures

Figure 1-1. Breakdown of factors affecting transition metal complex properties.....	26
Figure 1-2. Examples of high-symmetry classes using nomenclature from 2019.....	27
Figure 2-1. Stacked, unnormalized histogram of the number of complexes in the CSD grouped by the number of unique ligand types and by the highest denticity of ligands in the complex (monodentate in brown, bidentate in gray, tridentate in orange, tetradentate in blue, pentadentate in green, and hexadentate in red, as indicated in the inset). These counts are shown for all complexes (top), unique and computation-ready complexes (middle), and the Fe(II) unique, computation-ready subset (bottom).	36
Figure 2-2. Symmetry classes for transition metal complexes with up to two unique ligands, L ₁ and L ₂ , from left to right and top to bottom: homoleptic (HO) M(L ₁) ₆ , monoheteroleptic (5+1) M(L ₁) ₅ L ₂ , <i>trans</i> symmetric (TS) M(L ₁) ₄ (L ₂) ₂ , <i>cis</i> symmetric (CS) M(L ₁) ₄ (L ₂) ₂ , <i>fac</i> symmetric (FS) M(L ₁) ₃ (L ₂) ₃ , and <i>mer</i> symmetric (MS) M(L ₁) ₃ (L ₂) ₃ . For each pair of ligands, a total of two homoleptic and eight two-ligand isomers can be obtained because the 5+1, <i>trans</i> symmetric, and <i>cis</i> symmetric complexes are unique if the stoichiometry of L ₁ and L ₂ are swapped.....	38
Figure 2-3. Symmetry classes for transition metal complexes with three unique ligands, L ₁ , L ₂ , and L ₃ , from left to right and top to bottom: <i>cis</i> asymmetric (CA) M(L ₁) ₄ L ₂ L ₃ , double <i>cis</i> symmetric (DCS) M(L ₁) ₂ (L ₂) ₂ (L ₃) ₂ , <i>trans</i> asymmetric (TA) M(L ₁) ₄ L ₂ L ₃ , double <i>trans</i> symmetric (DTS) M(L ₁) ₂ (L ₂) ₂ (L ₃) ₂ , equatorial asymmetric (EA) M(L ₁) ₂ (L ₂) ₂ (L ₃) ₂ , <i>fac</i> asymmetric (FA) M(L ₁) ₂ (L ₂) ₃ L ₃ , <i>trans mer</i> asymmetric (MAT) M(L ₁) ₂ (L ₂) ₃ L ₃ , and <i>cis mer</i> asymmetric (MAC) M(L ₁) ₂ (L ₂) ₃ L ₃ . A total of 29 complexes can be obtained for any combination of three ligands due to additional isomers of the equatorial asymmetric type and those for which the stoichiometry of each ligand type is not equal.....	39
Figure 2-4. Percent of all unique mononuclear octahedral transition metal complexes in the CSD with user-defined charges (top, computation-ready) and the Fe(II) subset [middle, Fe(II)], grouped by a symmetry class for cases with two unique ligands (left: 5+1, <i>cis</i> symmetric, CS, <i>trans</i> symmetric, TS, <i>fac</i> , or <i>mer</i>) or three unique ligands (right: <i>cis</i> asymmetric, CA, double <i>cis</i> symmetric, DCS, <i>trans</i> asymmetric, TA, double <i>trans</i> symmetric, DTS, equatorial asymmetric, EA, <i>fac</i> asymmetric, FA, <i>cis mer</i>	

asymmetric, MAC, or *trans mer* asymmetric, MAT). The ratio of symmetry classes for the theoretical complexes from enumeration is shown at the bottom for comparison... 41

Figure 2-5. Calculated vs linearly interpolated ΔE_{H-L} (kcal/mol) for Fe(II) complexes with pairs of any of the three ligands: CH₃CN, H₂O, and CO. From left to right: interpolation between homoleptic complexes (HO only), interpolation using homoleptic complexes and CS and TS complex energies (CS+TS), or interpolation using homoleptic complexes and FS and MS complex energies (FS+MS). Points are colored according to the pair of ligands they correspond to: CH₃CN–H₂O (green circles), CH₃CN–CO (red squares), and CO–H₂O (blue triangles), as indicated in the inset. Key isomers are annotated. Points provided for the fit are translucent, whereas the remaining points are opaque. In all panels, a black dotted parity line is shown..... 44

Figure 2-6. Calculated vs linearly interpolated HOMO level (eV) for singlet Fe(II) complexes with pairs of any of the three ligands: CH₃CN, H₂O, and CO. From left to right: interpolation between homoleptic complexes (HO only), interpolation using homoleptic complexes and CS and TS complex energies (CS+TS), or interpolation using homoleptic complexes and FS and MS complex energies (FS+MS). Points are colored according to the pair of ligands they correspond to: CH₃CN–H₂O (green circles), CH₃CN–CO (red squares), and CO–H₂O (blue triangles), as indicated in the inset. Key isomers are annotated. Points provided for the fit are translucent, whereas the remaining points are opaque. In all panels, a black dotted parity line is shown. 44

Figure 2-7. Calculated vs linearly interpolated ΔE_{H-L} (kcal/mol) for Fe(II) complexes with at least one each of three ligands: CH₃CN, H₂O, and CO. From left to right: interpolation from homoleptic complexes (HO only), interpolation using homoleptic complexes and CS and TS complex energies derived from pairs of ligands (CS+TS), or interpolation using homoleptic complexes and FS and MS complex energies derived from pairs of ligands (FS+MS). Points are colored according to the ligand with the highest stoichiometric coefficient: H₂O (red circles), CO (gray squares), and CH₃CN (blue diamonds), or equal weight of all ligands (green triangles), as indicated in the inset. Key isomers are annotated. In all panels, a black dotted parity line is shown. 49

Figure 2-8. Calculated vs linearly interpolated HOMO level (eV) for singlet Fe(II) complexes with at least one each of three ligands: CH₃CN, H₂O, and CO. From left to right: interpolation from homoleptic complexes (HO only), interpolation using homoleptic complexes and CS and TS complex energies derived from pairs of ligands (CS+TS), or interpolation using homoleptic complexes and FS and MS complex energies derived from pairs of ligands (FS+MS). Points are colored according to the ligand with the highest stoichiometric coefficient: H₂O (red circles), CO (gray squares), and CH₃CN

(blue diamonds), or equal weight of all ligands (green triangles), as indicated in the inset. Key isomers are annotated. In all panels, a black dotted parity line is shown. 49

Figure 2-9. The $\Delta E_{\text{H-L}}$ (in kcal/mol) vs singlet HOMO level (in eV) for 56 homoleptic complexes (orange circles) and HO-only interpolation of all possible binary and ternary complexes colored by frequency from purple (low) to yellow (high). 1D histograms of each property are shown at top and right with bin widths of 2 kcal/mol and 0.25 eV, respectively, along with a kernel density estimate of the interpolated space shown as a dark blue line. A targeted zone of -4 to 4 kcal/mol for $\Delta E_{\text{H-L}}$ and -14.0 to -13.0 eV for the HOMO is annotated as a light blue square. 53

Figure 2-10. Left and middle: The $\Delta E_{\text{H-L}}$ (in kcal/mol) vs singlet HOMO level (in eV) for 12 homoleptic complexes (orange circles) and HO-only (left) or FS/MS-based (middle) interpolation of all possible binary and ternary complexes colored by frequency from purple (low) to yellow (high). The FS/MS complex energies are shown as pink circles. Right: The difference (i.e., HO-only minus FS/MS-based interpolation) of the two 2D histograms plotted from negative (blue, -81) to positive (red, 81). A targeted zone of -4 to 4 kcal/mol for $\Delta E_{\text{H-L}}$ and -14.0 to -13.0 eV for the HOMO is annotated as a light blue square. 56

Figure 2-11. Properties of eight binary (red or green) and six ternary (orange or blue) complexes in the validation set for HO-interpolation (circles) and FS/MS-augmented interpolation (squares) for $\Delta E_{\text{H-L}}$ (in kcal/mol, left) and the HOMO level (in eV, right). The targeted zone for each quantity is shown as a turquoise square, and three representative complexes are shown in the inset with their symmetry class, and the associated points are indicated with gray arrows. Structures are colored as follows: brown for Fe, gray for C, blue for N, white for H, green for Cl, red for O, and yellow for S. A dotted parity line is also shown. 56

Figure 3-1. (top) Qualitative diagrams of electron configurations in low-spin (LS), intermediate-spin (IS), and high-spin (HS) states for the mononuclear octahedral transition-metal complexes studied in this work (schematically shown at left). For both d^3/d^7 and d^4/d^6 M(II) or M(III) complexes, the additional electrons for the later transition metal are shown in red, and the electrons that apply to both states are shown in blue. The d^3 or d^7 complexes do not have a defined HS state. (bottom) The ten main monodentate ligands studied in this work ordered by their increasing ligand field strength, which tunes the octahedral field splitting (schematically shown at left). Atoms in the ball-and-stick representation are colored as follows: H in white, C in gray, N in blue, O in red, F in light blue, P in orange, S in yellow, and Cl in green. 66

Figure 3-2. Isovalent mutation example and isovalent counterparts 69

Figure 3-3. Construction of dataset for isovalent mutations	70
Figure 3-4. Comparisons of adiabatic spin splitting (ΔE , in kcal/mol) for HS-LS (green lines and shading), HS-IS (blue lines and shading) and IS-LS (red lines and shading) for pairs of homoleptic TMCs grouped first by the isovalent $3d$ and $4d$ metals (i.e., Fe(II)/Ru(II) vs Mn(III)/Tc(III)) and then by ligand (i.e., CO vs H ₂ O), as indicated on the x -axis. The solid lines correspond to values at $a_{\text{HF}} = 0.2$, the inner translucent shaded regions correspond to the $a_{\text{HF}} = 0.1$ – 0.3 range, and the outer translucent shaded regions correspond to $a_{\text{HF}} = 0.0$ – 0.4 . A zero axis is shown to indicate where ordering changes for any pair of states. The Ru(II)(H ₂ O) ₆ IS state was eliminated during filtering steps, and so its HS-IS or IS-LS data is unavailable.	71
Figure 3-5. Mean unsigned error (MUE, in kcal/mol) for the prediction of $4d$ TMC properties with only 20 $4d$ data points with eRAC-185 (red) or by including $3d$ data with 20 $4d$ data points and using either the RAC-155 (green) or eRAC-185 (blue) on the isovalent metal pairing data set. The colored bars represent the average from an ensemble of 25 feature-selected KRR models, and the error bars are the standard deviation of the ensemble.	72
Figure 3-6. Effect of simultaneous metal-ligand perturbation vs. sum of independent perturbations on normalized bond length. The illustrated TMCs contain $3d$ metals and $2p$ ligands.	73
Figure 4-1. Relationship between computation and experiment.....	78
Figure 4-2. Example of a transition metal complex, which contains a central transition metal surrounded by organic ligands.....	80
Figure 4-3. Calculated forms of the various potentials provided by UFF.	85
Figure 4-4. Plot illustrating how the time needed to generate a full set of parameters for N unique atom types scales with N	87
Figure 4-5. Performance of UFF.jl for generating full parameter sets in parallel and non-parallel environments.	88
Figure 4-6. Optimization trajectory for butane using forward-mode automatic differentiation to obtain gradients.....	89
Figure 4-7. Structures of butane before and after optimization.....	90
Figure 4-8. Example of a larger-scale simulation involving a protein in H ₂ O solvent. ...	90
Figure 5-1. Current frontend of molSimplify Lite.....	95
Figure 5-2. Hosting Structure of molSimplify Lite	95
Figure 5-3. Example feedback forms on MOFSimplify website.	96
Figure 5-4. Data upload form on MOFSimplify website.....	96

Figure 5-5. Database system diagram. iRun and Gibraltar refer to servers owned by the Kulik group. The mofsimply VM is hosted on the iRun server.....	97
Figure 5-6. a) A plot of the predicted ΔE_{H-L} for $\text{Fe}(\text{NH}_3)(\text{H}_2\text{O})_6$ with oxidation state 2 compared to other complexes in molSimplify’s training data. b) PCA showing the same complex in green among the training data in molSimplify in RAC space.	99
Figure 5-7. Example DFA recommendation for ΔE_{H-L} for $\text{Fe}(\text{NH}_3)(\text{H}_2\text{O})_6$. The website ranks DFAs by predicted error relative to DLPNO-CCSD(T).....	99
Figure 5-8. Sections of the MOFSimplify web interface. a) Interface for selecting a MOF for analysis and predicting properties of the selected MOF using ANNs trained on experimental data mined from the literature. The default MOF loaded upon selecting “Example MOF” is HKUST-1, a well-studied MOF. ⁸⁵ b) The feedback interface for evaluating model predictions. c) The interface listing similar (i.e., LSNN) MOFs to the selected MOF as determined by the ANNs. d) Visualization of the selected MOF’s components. e) Visualization of the selected MOF’s unit cell.	100
Figure 5-9. Locations of site accesses during first month of molSimplify Lite launch. Map was made with an online IP resolver and ArcGis.....	101
Figure 6-1. Paradigms for user interaction with machine learning models.....	104
Figure 6-2. Distribution of ligand denticities found in the Cambridge Structural Database. The ligands considered are unique ligands found across all octahedral TMCs (left) and tetrahedral TMCs (right). The orange regions represent the body of ligands that require a co-ligand to be studied, but cannot be studied as part of a homoleptic complex.	105
Figure 6-3. Molecular weight of true and false pairs of monodentate and tridentate ligands.	110
Figure 6-4. Comparison of property distributions (top left, number of atoms; top right, percent buried volume; bottom, ligand charge) for monodentate and tridentate ligands in true and false pairs.....	111
Figure 6-5. Spin splitting energy of true and fictional pairs of ligands.....	112
Figure 6-6. ROC curve for the trained ANN classifier (left) and the unsupervised UMAP representation of positive examples and negative examples in red and blue, respectively (right). The UMAP indicates a clear separation between the two groups, illustrating that stochastic negative addition creates negative examples that are distinguishable from positive examples in feature space. The high ROC-AUC of 0.87 illustrates that the ANN can learn the decision boundary between the positive and negative examples robustly and with high accuracy.	113

Figure 6-7. The Ligandify web interface. The user provides a SMILES string of a monodentate ligand including a placeholder metal atom, and Ligandify generates a 3D structure of the specified ligand. The “Find Candidates” button is used to find the top 5 tridentate co-ligand candidates for forming a tetrahedral complex with the provided monodentate ligand.115

Figure A-1. Calculated vs. linearly interpolated ΔE_{H-L} (kcal/mol) for Fe(II) complexes with pairs of any of the three ligands: NH₃, H₂O, and CO. From left to right: interpolation between homoleptic complexes (HO only), interpolation using homoleptic complexes as well as *cis* symmetric and *trans* symmetric complex energies (CS+TS), or interpolation using homoleptic complexes as well as *fac* and *mer* symmetric complex energies (FS+MS). Points are colored according to the pair of ligands they correspond to: NH₃-H₂O (green squares), NH₃-CO (gray circles), and CO-H₂O (blue triangles), as indicated in inset legend. Key isomers are annotated. Points provided for the fit are translucent, whereas the remaining points are opaque. In all panes, a black dotted parity line is shown.125

Figure A-2. Calculated vs. linearly interpolated HOMO level (eV) for singlet Fe(II) complexes with pairs of any of the three ligands: NH₃, H₂O, and CO. From left to right: interpolation between homoleptic complexes (HO only), interpolation using homoleptic complexes as well as *cis* symmetric and *trans* symmetric complex energies (CS+TS), or interpolation using homoleptic complexes as well as *fac* and *mer* symmetric complex energies (FS+MS). Points are colored according to the pair of ligands they correspond to: NH₃-H₂O (green squares), NH₃-CO (gray circles), and CO-H₂O (blue triangles), as indicated in inset legend. Key isomers are annotated. Points provided for the fit are translucent, whereas the remaining points are opaque. In all panes, a black dotted parity line is shown.128

Figure A-3. Calculated vs. linearly interpolated ΔE_{H-L} (kcal/mol) for Fe(II) complexes with at least one each of three ligands: NH₃, H₂O, and CO. From left to right: interpolation from homoleptic complexes (HO only), interpolation using homoleptic complexes as well as *cis* symmetric and *trans* symmetric complex energies derived from pairs of ligands (CS+TS), or interpolation using homoleptic complexes as well as *fac* and *mer* symmetric complex energies derived from pairs of ligands (FS+MS). Points are colored according to the ligand with the highest stoichiometric coefficient: H₂O (red circles), CO (gray squares), and NH₃ (blue down triangles), or equal weight of all ligands (green up triangles), as indicated in inset legend. Key isomers are annotated. In all panes, a black dotted parity line is shown.131

Figure A-4. Calculated vs. linearly interpolated HOMO level (eV) for singlet Fe(II) complexes with at least one each of three ligands: NH₃, H₂O, and CO. From left to right: interpolation from homoleptic complexes (HO only), interpolation using homoleptic complexes as well as *cis* symmetric and *trans* symmetric complex energies derived from pairs of ligands (CS+TS), or interpolation using homoleptic complexes as well as *fac* and *mer* symmetric complex energies derived from pairs of ligands (FS+MS). Points are colored according to the ligand with the highest stoichiometric coefficient: H₂O (red circles), CO (gray squares), and NH₃ (blue down triangles), or equal weight of all ligands (green up triangles), as indicated in inset legend. Key isomers are annotated. In all panes, a black dotted parity line is shown.133

Figure A-5. The ΔE_{H-L} (in kcal/mol) vs. singlet HOMO level (in eV) for 56 homoleptic Fe(II) complexes shown as a scatter plot (middle) with normalized marginal 1D histograms at top and right, respectively computed with B3LYP.139

Figure A-6. The ΔE_{H-L} (in kcal/mol) vs. singlet HOMO level (in eV) for 56 homoleptic Fe(II) complexes shown as a scatter plot (middle) with normalized marginal 1D histograms at top and right, respectively, computed with modified B3LYP with 10% HF exchange.139

Figure A-7. Homoleptic-only interpolation for Fe(II) complexes: the ΔE_{H-L} (in kcal/mol) vs. singlet HOMO level (in eV) for 12 homoleptic Fe(II) complexes shown as both a scatter plot (middle, orange) and with normalized marginal 1D histograms at top and right (orange bars), respectively. The interpolated values are shown as a 2D histogram colored from low (purple) to high (yellow) density, and the same data is shown as a kernel density estimate on the histogram panes.141

Figure A-8. Parity plot of calculated vs. HO-only interpolated ΔE_{H-L} (in kcal/mol, left) and HOMO level (in eV, right) for 66 *fac* (blue circles) and *mer* (red circles) derived from 12 homoleptic Fe(II) complexes. A black solid parity line is also shown.142

Figure B-1. $S(3d)$ vs $S(4d)$ for ΔE_{H-I} (in kcal/mol·HFX) of all TMCs with CO or H₂O ligands in both oxidation states, colored by element (Cr/Mo in gray, Mn/Tc in orange, or Fe/Ru in red) and with symbols corresponding to formal electron configuration (d^4 in triangles, d^5 in circles, and d^6 in squares). All hexa-aqua complexes are outlined in green, all hexa-carbonyl complexes are outlined in black, and the remaining symbols are outlined in dark gray. A dotted parity line is shown for reference.155

Figure B-2. $S(3d)$ vs $S(4d)$ for ΔE_{I-L} (in kcal/mol·HFX) of all TMCs with CO or H₂O ligands in both oxidation states, colored by element (Cr/Mo in gray, Mn/Tc in orange, Fe/Ru in red, and Co/Rh in blue) and with symbols corresponding to formal electron

configuration (d^3 in right-pointing triangles, d^4 in up-pointing triangles, d^5 in circles, d^6 in squares, and d^7 in diamonds). All hexa-aqua complexes are outlined in green, all hexa-carbonyl complexes are outlined in black, and the remaining symbols are outlined in dark gray. A dotted parity line is shown for reference.159

Figure B-3. Absolute LS M–L averaged distance (in Å) vs HS M–L averaged distance (in Å) for 155 pairs of equilibrium structures at $a_{\text{HF}} = 0.2$ in the 4-electron LS-to-HS spin state comparison for $3d$ (translucent green circles) and $4d$ (translucent blue squares) TMCs. A black dotted parity line is also shown. All $4d$ TMCs generally have longer bond lengths than $3d$ TMCs due to the larger metal covalent radius.161

Figure B-4. Normalized histograms for $3d$ TMCs (top) and $4d$ TMCs (bottom) with the same x-axis ranges and values for the d_{rel} (left) and difference in d_{rel} (right) plots. The left plots show the averaged (over all 6 metal–ligand bonds) d_{rel} values grouped in translucent histograms by LS (red), IS (gray), and HS (blue) states for 155 (or 247) pairs for which HS, IS, and LS states are all converged (247 is for IS-LS). The right plots show the difference in average d_{rel} values by state of the complexes in translucent normalized histograms: 2-electron IS-LS in magenta, 4-electron HS-LS in purple, and 2-electron HS-IS in cyan.163

Figure B-5. Relative distance averaged over two axial metal–ligand bonds vs. over four equatorial metal–ligand bonds for $3d$ TMCs: 155 pairs of LS (red translucent circles) or HS (blue translucent circles) as well as 247 pairs of IS (gray translucent circles). A parity line is shown in dotted black, and the plot area is square.165

Figure B-6. Relative distance averaged over two axial metal–ligand bonds vs. over four equatorial metal–ligand bonds for $4d$ TMCs: 155 pairs of LS (red translucent squares) or HS (blue translucent squares) as well as 247 pairs of IS (gray translucent squares). A parity line is shown in dotted black, and the plot area is square.165

Figure B-7. The difference in average relative metal–ligand distance, Δd_{rel} , for the 4-electron HS-LS state comparison vs. $\Delta E_{\text{H-L}}$ (in kcal/mol) for 155 pairs of HS and LS states at $a_{\text{HF}} = 0.2$ for $3d$ (translucent green circles) and $4d$ (translucent blue squares) TMCs. A zero axis shows when spin state ordering changes or the sign of the relative metal–ligand bond length difference changes.166

Figure B-8. Difference for $3d$ TMCs in average relative distance between relevant spin states, Δd_{rel} , vs the sensitivity of spin state splitting, $S(\Delta E)$, for the same spin states in kcal/mol-HFX grouped by 155 pairs of 4-electron HS-LS (H-L, purple translucent circles), 247 pairs of 2-electron IS-LS (I-L, magenta translucent circles), and 155 pairs of 2-electron HS-IS (H-I, cyan translucent circles). Best-fit lines have been obtained

through each spin state definition and shown as dotted solid lines in the same colors. A gray dotted line has been fit through all data.167

Figure B-9. Difference for $4d$ TMCs in average relative distance between relevant spin states, Δd_{rel} , vs the sensitivity of spin state splitting, $S(\Delta E)$, for the same spin states in kcal/mol·HFX grouped by 155 pairs of 4-electron HS-LS (H-L, purple translucent squares), 247 pairs of 2-electron IS-LS (I-L, magenta translucent squares), and 155 pairs of 2-electron HS-IS (H-I, cyan translucent squares). Best-fit lines have been obtained through each spin state definition and shown as dotted solid lines in the same colors. A gray dotted line has been fit through all data.168

List of Tables

Table 1-1. Advantages and Disadvantages of Computational vs. Experimental Approaches to Inorganic Discovery	25
Table 2-1 The number of theoretical complexes for each octahedral symmetry class considered in this work and the full octahedral space for an example single metal/oxidation/spin state ($m=1$) with an $N=12$ ligand pool. The configurations and isomers indicate the number of ways unique ligands can be arranged, and the cardinality indicates how many theoretical complexes can be enumerated.	37
Table 4-1. Atomic parameters provided by UFF.	80
Table 6-1. Percent changes in representation of octahedral complexes upon filtering for “catalysis” tag from CrossRef API	105
Table 6-2. Percent changes in representation of tetrahedral complexes upon filtering for “catalysis” tag from CrossRef API	106
Table A-1. Number of complexes grouped by number of unique ligands for all (all) mononuclear octahedral transition metal complexes, all unique complexes with user-defined charges and nondisordered structures ("computation-ready"), and the subset of unique complexes with Fe(II) centers (Fe(II)). The percent of each category is also shown.	121
Table A-2. Number of complexes grouped by symmetry class for all mononuclear octahedral transition metal complexes, all complexes with user-defined charges and nondisordered structures ("computation-ready"), and the subset of complexes with Fe(II) centers. The percent of each category is also shown.....	122
Table A-3. Number of unique ligands of each denticity grouped by symmetry class for all complexes with user-defined charges and non-disordered structures.	123
Table A-4. Number of unique ligands of each denticity grouped by symmetry class for all complexes with user-defined Fe(II) centers.	123
Table A-5. Number of monodentate ligands of each category from Fe(II) monodentate-only structures for the case where user defined charges and trustworthy structures were obtained ("All computation-ready") as well as the more stringent test where ligand charges were assigned as neutral. The bolded categories correspond to those used to study HO Fe(II) complexes as described in the main text.	124

Table A-6. Number of complexes from monodentate-only configurations for each symmetry from the total user defined charges and trustworthy structures ("computation-ready") or for Fe(II) only. For Fe(II) complexes, theoretical spaces of $N = 56$ and $N = 88$ ligands are compared for each symmetry class along with the % of that space present in the CSD. The relevant symmetry class and notation is also provided.....	124
Table A-7. Homoleptic Fe(II) transition metal complex properties, DE_{H-L} (in kcal/mol) and singlet HOMO level (in eV) for the ligands NH_3 , CH_3CN , CO , and H_2O obtained with B3LYP/LACVP* calculations.....	125
Table A-8. ΔE_{H-L} values for pairs of two ligand types in Fe(II) complexes (in kcal/mol) along with interpolated values using only HO complexes, HO+CS/TS complexes, or HO+FS/MS complexes.	127
Table A-9. Mean absolute errors (MAE) in kcal/mol for Fe(II) complex ΔE_{H-L} estimates from HO-only averaging, including CS/TS in the averaging, and including FS/MS averaging for pairs of ligands, L_1 and L_2 . The CS/TS and FS/MS errors are both averaged over only the points for which exact energies were not provided (interp.) as well as over all (all) points.	128
Table A-10. Singlet HOMO level values for pairs of two ligand types in Fe(II) complexes (in eV) along with interpolated values using only HO complexes, HO+CS/TS complexes, or HO+FS/MS complexes.	129
Table A-11. Mean absolute errors (MAE) in eV for singlet complex HOMO levels from HO-only averaging, including CS/TS in the averaging, and including FS/MS averaging for pairs of ligands, L_1 and L_2 . The CS/TS and FS/MS errors are both averaged over only the points for which exact energies were not provided (interp.) as well as over all (all) points.	130
Table A-12. ΔE_{H-L} values for Fe(II) complexes (in kcal/mol) with a combination of three ligand types, H_2O , CO , and CH_3CN , in Fe(II) complexes along with interpolated values using only HO complexes, HO+CS/TS complexes, or HO+FS/MS complexes. For the three EA isomers, L_3 is the ligand in the <i>trans</i> position as indicated in the main text..	131
Table A-13. ΔE_{H-L} values for Fe(II) complexes (in kcal/mol) with a combination of three ligand types, H_2O , CO , and NH_3 , in Fe(II) complexes along with interpolated values using only HO complexes, HO+CS/TS complexes, or HO+FS/MS complexes. For the three EA isomers, L_3 is the ligand in the <i>trans</i> position as indicated in the main text..	132
Table A-14. Singlet HOMO level values with a combination of three ligand types, H_2O , CO , and CH_3CN , in Fe(II) complexes (in eV) along with interpolated values using only	

HO complexes, HO+CS/TS complexes, or HO+FS/MS complexes. For the three EA isomers, L₃ is the ligand in the *trans* position as indicated in the main text.134

Table A-15. Singlet HOMO level values with a combination of three ligand types, H₂O, CO, and NH₃, in Fe(II) complexes (in eV) along with interpolated values using only HO complexes, HO+CS/TS complexes, or HO+FS/MS complexes. For the three EA isomers, L₃ is the ligand in the *trans* position as indicated in the main text.135

Table A-16. Mean absolute error (MAE) for singlet complex HOMO levels (in eV) and Fe(II) complex ΔE_{H-L} (in kcal/mol) for combinations of ligands, L₁, L₂, and L₃ evaluated with three averaging schemes: HO-only, CS/TS-derived averaging, and FS/MS-derived averaging.136

Table A-17. Properties of 56 homoleptic Fe(II) complexes derived from either homoleptic examples in the CSD (36 "HO") with neutral ligands or from neutral ligands only present in binary and ternary Fe(II) complexes (20 "B or T") from DFT. Both the ΔE_{H-L} (in kcal/mol) and the HOMO level of the singlet complex (in eV) are shown. Complexes are distinguished by stoichiometry.136

Table A-18. Twelve most common ligands in target zone (i.e., ΔE_{H-L} from -4 to +4 kcal/mol and singlet HOMO level from -14.0 to -13.0 eV) for Fe(II) complexes and the number of times the homoleptic-only averaging predicts that a complex in the zone will have that ligand in a binary complex or both binary and ternary complexes. Refcodes in italics refer to any complex that contains that ligand, whereas the remainder are the precise homoleptic Fe(II) complexes. The top 10 most frequent ligands were selected along with two high-frequency ligands (MeCN and ClPyz) for added diversity.....140

Table A-19. MAEs of homoleptic-only interpolation over the FS/MS 132 complex subset with respect to calculated values.....142

Table A-20. Binary complexes with their chemical formula and symmetry along with ΔE_{H-L} (in kcal/mol) and HOMO level (in eV) obtained through three approaches: homoleptic-only interpolation, FS/MS-augmented interpolation, and explicit calculation. For the cases that are FS or MS complexes, FS/MS-augmented interpolation refers to the explicitly calculated property. As a result, the reported MAEs are obtained over 5 complexes for FS/MS-interpolation but 8 complexes for HO-interpolation. The resulting calculated properties that are in the targeted zone are shown in bold.142

Table A-21. Ternary complexes with their chemical formula and symmetry along with ΔE_{H-L} (in kcal/mol) and HOMO level (in eV) obtained through three approaches: homoleptic-only interpolation, FS/MS-augmented interpolation, and explicit calculation. The resulting calculated properties that are in the targeted zone are shown in bold....143

Table B-1. Comparison of effects of equilibrium bond length differences from larger basis sets for B3LYP/LACVP* and B3LYP/def2-ZORA-TZVP (i.e., $a_{\text{HF}} = 0.2$) structures of Fe(II) or Ru(II)(CO)₆ in HS quintet and LS states.....144

Table B-2. Geometric and electronic criteria adapted from prior work¹ for excluding TMCs from data set with loose geometric cutoffs applied during resubmission indicated in parentheses.....144

Table B-3. Number of $3d$ or $4d$ TMCs eliminated at each sequential filtering step along with how many are retained, and the cumulative total is listed as overall. The filtering steps were: i) convergence (completion, passing loose thresholds upon 5 24 hour job resubmissions), ii) geometry metrics, iii) deviations of $\langle S^2 \rangle$ from the expected value by more than $1\mu_{\text{B}}$, and iv) deviations of metal spin of more than $1\mu_{\text{B}}$146

Table B-4. Summary of computed sensitivities: all $S(3d)$ or $S(4d)$ results that were obtained (“results”) along with those that were eliminated for having fewer than 4 points or due to the inability to eliminate slope changes between two points. All lines were retained regardless of R^2 values, but any points where LOOCV errors for a point were 5 kcal/mol or higher were eliminated.146

Table B-5. The ratio of S for $\Delta E_{\text{I-L}}$ to $\Delta E_{\text{H-L}}$ or $\Delta E_{\text{H-I}}$ to $\Delta E_{\text{H-L}}$ for $3d$ and $4d$ TMCs for all 154 TMCs for which all 3 quantities are available, and all sensitivities are negative. The average, standard deviation, minimum, maximum, and range for each ratio is reported.147

Table B-6. Summary of sensitivities (in kcal/mol-HFX) for pairs of $3d$ and $4d$ TMCs grouped by metal and oxidation state. The total number of $3d/4d$ pairs is indicated along with the average, standard deviation, minimum, maximum, and range for the $3d$ or $4d$ TMC subsets. A best-fit line for the relationship between the sensitivities and the associated R^2 value are also shown.....148

Table B-7. Outliers where the $S(4d)$ value exceeds the $S(3d)$ value, both shown in kcal/mol-HFX. The R^2 value for each sensitivity determination and the difference between the first-row and second-row sensitivity are shown grouped by the type of spin splitting.149

Table B-8. Outliers where the $S(4d)$ or $S(3d)$ value is positive (both shown in kcal/mol-HFX) for 4-electron $\Delta E_{\text{H-L}}$ and 2-electron $\Delta E_{\text{H-I}}$. The R^2 value for each sensitivity determination and the difference between the first-row and second-row exchange sensitivity are shown grouped by the type of spin splitting.....149

Table B-9. Outliers where the $S(4d)$ or $S(3d)$ value is positive (both shown in kcal/mol-HFX) for 2-electron $\Delta E_{\text{I-L}}$. The R^2 value for each sensitivity determination and

the sensitivity difference are shown. Only the bottom three cases are $S(3d) > 0$, the rest are $S(4d) > 0$150

Table B-10. Adiabatic spin-splitting energies, ΔE_{H-L} , for the 4-electron difference HS-to-LS energies in kcal/mol for $3d$ TMCs containing CO and H₂O ligands with varied a_{HF} values as indicated in the table alongside the computed sensitivity and the R^2 value from a linear fit. Values in italics have been averaged between neighboring points or extrapolated using the linear fit if not between neighboring points. The symmetry and composition of the complex is indicated in the table.....151

Table B-11. Adiabatic spin-splitting energies, ΔE_{H-L} , for the 4-electron difference HS-to-LS energies in kcal/mol for $4d$ TMCs containing CO and H₂O ligands with varied a_{HF} values as indicated in the table alongside the computed sensitivity and the R^2 value from a linear fit. Values in italics have been averaged between neighboring points or extrapolated using the linear fit if not between neighboring points. The symmetry and composition of the complex is indicated in the table.....152

Table B-12. Sensitivities, S , in kcal/mol-HFX, for ΔE_{H-L} 4-electron difference adiabatic spin-splitting energies computed among pairs of $3d$ and $4d$ TMCs containing CO and H₂O ligands. The symmetry and composition of the complex is indicated in the table.153

Table B-13. Adiabatic spin-splitting energies, ΔE_{H-I} , for the 2-electron difference HS-to-IS energies in kcal/mol for $3d$ TMCs containing CO and H₂O ligands with varied a_{HF} values as indicated in the table alongside the computed sensitivity and the R^2 value from a linear fit. Values in italics have been averaged between neighboring points or extrapolated using the linear fit if not between neighboring points. The symmetry and composition of the complex is indicated in the table.....153

Table B-14. Adiabatic spin-splitting energies, ΔE_{H-I} , for the 2-electron difference HS-to-IS energies in kcal/mol for $4d$ TMCs containing CO and H₂O ligands with varied a_{HF} values as indicated in the table alongside the computed sensitivity and the R^2 value from a linear fit. Values in italics have been averaged between neighboring points or extrapolated using the linear fit if not between neighboring points. The symmetry and composition of the complex is indicated in the table.....154

Table B-15. Sensitivities, S , in kcal/mol-HFX for ΔE_{I-L} and ΔE_{H-I} 2-electron difference adiabatic spin-splitting energies computed among pairs of $3d$ and $4d$ TMCs containing CO and H₂O ligands. The symmetry and composition of the complex is indicated in the table.....155

Table B-16. Adiabatic spin-splitting energies, ΔE_{I-L} , for the 2-electron difference IS-to-LS energies in kcal/mol for $3d$ TMCs containing CO and H₂O ligands with varied a_{HF}

values as indicated in the table alongside the computed sensitivity, S , and the R^2 value from a linear fit. Values in italics have been averaged or extrapolated using the linear fit if not between neighboring points. The symmetry and composition of the complex is indicated in the table.....156

Table B-17. Adiabatic spin-splitting energies, ΔE_{I-L} , for the 2-electron difference IS-to-LS energies in kcal/mol for 4*d* TMCs containing CO and H₂O ligands with varied a_{HF} values as indicated in the table alongside the computed sensitivity and the R^2 value from a linear fit. Values in italics have been averaged between neighboring points or extrapolated using the linear fit if not between neighboring points. The symmetry and composition of the complex is indicated in the table.....158

Table B-18. Dependence of HS/LS preference based on sign of ΔE_{H-L} for all 239 pairs of 3*d* and 4*d* complexes for which ΔE_{H-L} was evaluated. This analysis is completed for different a_{HF} fractions labeled at the top of the table. The $a_{HF} = 0.4$ value is extrapolated from the sensitivity. Each preference is grouped by metal and oxidation state and then summarized at the bottom as overall.159

Table B-19. Dependence of GS preference (HS, LS, or IS) for the 247 pairs for which all compatible spin states have been converged. This corresponds to the 155 pairs of 3*d* and 4*d* complexes for which HS/IS/LS states are all converged in addition to the 92 cases of d^3 Cr(III)/Mo(III) or d^7 Co(II)/Rh(II) pairs for which IS/LS states were converged. The variation is shown with a_{HF} fractions, as labeled at the top of the table. The $a_{HF} = 0.4$ value is extrapolated from the sensitivity. The overall LS/IS/HS count is shown at the bottom of the table along with the LS/IS count for the d^3/d^7 cases.160

Table B-20. Dependence of HS/LS preference based on sign of ΔE_{H-L} for the 155 pairs of 3*d* and 4*d* complexes for which HS/IS/LS states are all converged with a_{HF} fractions labeled at the top of the table. The $a_{HF} = 0.4$ value is extrapolated from the sensitivity. This analysis excludes the existence of the IS state and only compares the HS and LS energies. Each preference is grouped by metal and oxidation state and then summarized at the bottom as overall.161

Table B-21. Covalent radii used for 3*d* and 4*d* metals (left) and ligand coordinating elements (right) to obtain relative distances in this work. Recommended covalent radii were obtained from Ref. 2. Where LS and HS covalent radii were provided (e.g., Fe), they were averaged, and where multiple radii based on hybridization (i.e., C) were provided, they were also averaged.....162

Table B-22. Average, minimum, maximum, and standard deviation (std.) of relative distances for all HS, IS, and LS states as well as the statistics for the HS-LS, IS-LS, and HS-IS differences compared for 155 pairs of 3*d* and 4*d* TMCs for which the LS, IS, and

HS states are both defined and successfully converged. The relative distances were computed with respect to the covalent radii of the metal and coordinating atoms and they were then averaged over all equatorial and axial bond lengths.....163

Table B-23. Average, minimum, maximum, and standard deviation (std.) of relative distances for IS and LS states as well as the statistics for the IS-LS differences compared for 92 pairs of $3d$ and $4d$ TMCs that have d^3 and d^7 electron configurations and therefore only LS and IS states are defined. The relative distances were computed with respect to the covalent radii of the metal and coordinating atoms and they were then averaged over all equatorial and axial bond lengths. The trends in the averages of these LS and IS sets are roughly comparable to those obtained for LS or IS states in the cases where HS states were also valid.....164

Table B-24. Best-fit lines ($y = mx + b$) of the difference in spin state relative bond lengths, Δd_{rel} , vs. sensitivity, $S(\Delta E)$ (in kcal/mol·HFX), for $3d$ TMCs and $4d$ TMCs grouped by relative spin states: 155 cases of 4-electron H-L, 155 cases of 2-electron H-I, and 247 cases of 2-electron I-L. The best-fit line for the overall $3d$ or $4d$ TMC set containing 557 points is also shown. The correlation (R^2) for each line is reported. We also report at the bottom of the table the trendlines obtained for the $3d$ or $4d$ $S(\Delta E_{\text{H-L}})$ vs. the LS d_{rel} as well as the line obtained by correlating both data sets together.....166

Table B-25. Properties of Fe(II)(He)₆ and Ru(II)(He)₆ LS and HS complexes: the M–He bond length (in Å) for the LS singlet and HS quintet states, and the relative bond lengths with respect to covalent radii.....168

Table B-26. Properties of Fe(II)(CO)₆ and Ru(II)(CO)₆ LS and HS complexes: the M–He bond length (in Å) for the LS singlet and HS quintet states, and the relative bond lengths with respect to covalent radii.....169

Table B-27. Characteristics of total energy sensitivities for the HS and LS states, $S(E_{\text{HS}})$ and $S(E_{\text{LS}})$ in kcal/mol·HFX of $3d$ and $4d$ TMCs: minimum, maximum, average, and the number that are positive.169

1 Introduction

Inorganic materials are integral to many sectors of the modern economy due to their industrial, pharmaceutical, and chemical applications. Part of the reason for their usefulness is their tunability, as their properties can be manipulated by delicate tuning of the ligand field. As the demand for inorganic materials grows, there will be rising demand for “designer” materials with targeted properties. In order to meet this demand at scale, it is important to have a systematic approach for discovering new materials.

Inorganic discovery is currently a field ripe for inquiry. The question of how to discover new complexes with targeted properties is one that brings together theoretical understanding, creativity, and the ability to make predictions about a wide range of materials. The main difficulty with inorganic discovery is combinatorial intractability: due to a multitude of possible choices of ligands, a wider range of materials is possible compared to organic chemistry. Additionally, testing new inorganic materials in the lab is lengthy and expensive, and the combinatorial explosion in the number of possible complexes makes experimental approaches almost futile for scanning large regions of chemical space.

Therefore, in order to tackle the question of inorganic discovery, we should attempt to improve tractability by limiting the scope of research and turning to alternative research paradigms. Because there are so many possible geometries with wildly different behavior, it could be useful to limit the scope of geometries we look at. Thus, we focus in one region of design space: octahedral inorganic complexes, which consist of a metal center and (usually organic) ligands, with 6 total metal-atom connections arranged octahedrally. However, even within this class of compounds, there is a huge number of possible complexes. To improve tractability over experimental approaches, we adopt a computational approach using first-principles methods such as DFT and data-based methods such as machine learning to accelerate discovery time. Some advantages and disadvantages are summarized in Table 1-1. The relationship between the two approaches (i.e. that computation is fast and possibly inaccurate vs. experiments, which are slower and correspond better to real-world conditions) suggests that we focus on discovering leads computationally, then feed those leads to experimental groups to test.

Table 1-1. Advantages and Disadvantages of Computational vs. Experimental Approaches to Inorganic Discovery

	Experiment	Computation
Advantages	<ul style="list-style-type: none"> • Results can be more readily ported to industrial and real-world applications • Results are more reliable re. expected behavior during process scale-up 	<ul style="list-style-type: none"> • Fast – no need for synthesis • Accurate (in a sense) – No noise from experimental conditions • Cheap – lots of data can be generated in an automated fashion at relatively low cost (e.g. on a GPU workstation) • Can investigate many different properties at once (e.g. SSE, redox, etc.) based on electronic structure
Disadvantages	<ul style="list-style-type: none"> • Expensive, time-consuming • “Experimental noise” can confound results • Limited characterization techniques 	<ul style="list-style-type: none"> • Inherent errors that don’t show up in experiment (e.g. due to limited basis set size, numerical error)¹ • Can do “experiments” on molecules that aren’t even synthetically accessible

1.1 Problem areas

The question of how to discover new materials in such a vast design space is naturally ill-defined due to the abundance of factors that affect their properties. To lend more structure to the question, we use a systematic view to highlight areas where intervention can lead to improved quality and quantity of discovery. A system breakdown is shown in Figure 1-1; by developing better control or understanding of what happens in the boxes or arrows, we can better approach discovery.

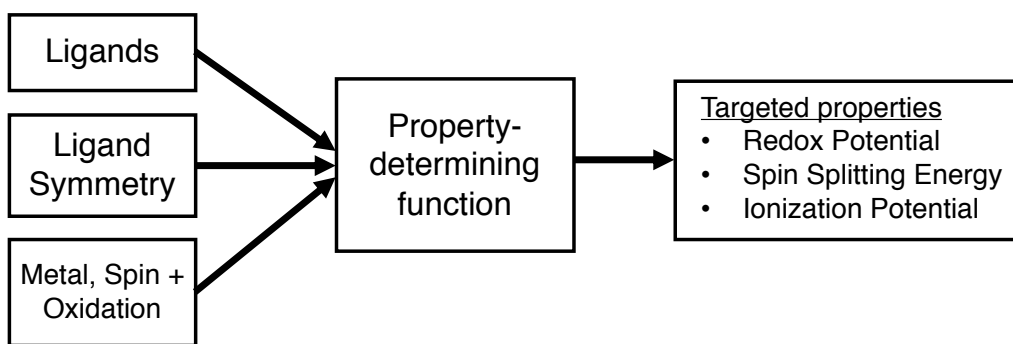


Figure 1-1. Breakdown of factors affecting transition metal complex properties.

The inputs to the model in Figure 1-1 are ligands, ligand symmetry, and metal properties. Possible variables for the “ligand” input are the identities of the ligands (i.e. the topology, denticity, charge distribution, and atoms of the ligands); changing these leads to changes in various aspects of electronic structure, such as spin localization on the metal and spin-splitting energy. The next input is ligand symmetry, which can occasionally be a significant factor affecting the properties of materials²; a summary of high-symmetry compounds is provided in Figure 1-2. The third input is the metal center, whose identity, oxidation state, and spin state can be varied. Although the design space of metals is much smaller than that of the ligands and their symmetry, it is important to note that some choices, such as the choice of 1st row metals versus 2nd row metals, can significantly change the type of insight gained from manipulating other variables. The central box, the property-determining function, represents nature; to better understand or predict its behavior, we can perform electronic structure calculations, develop theoretical understanding, or use methods such as machine learning to get “faster-than-fast” predictions at the cost of accuracy.

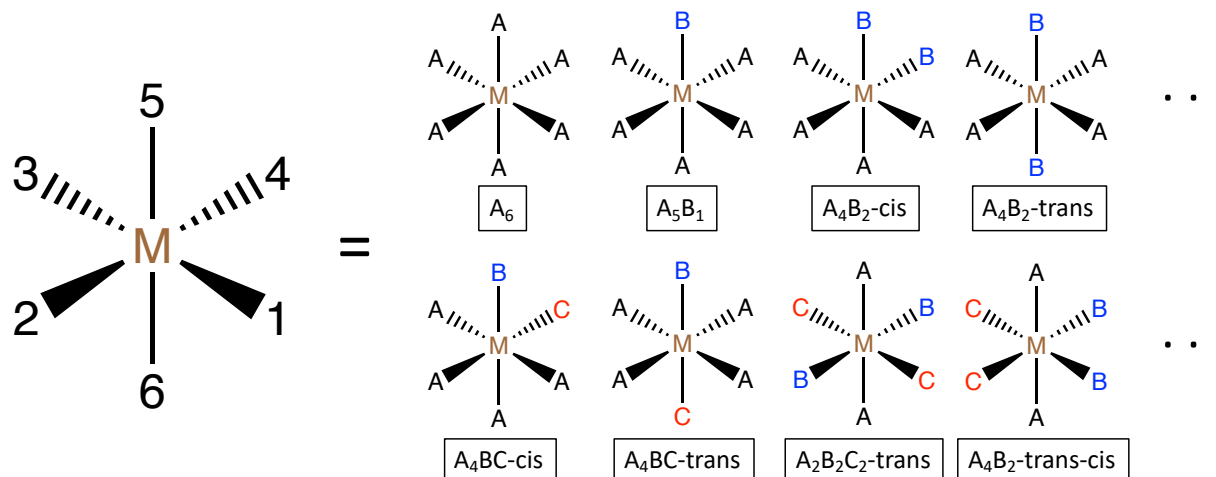


Figure 1-2. Examples of high-symmetry classes using nomenclature from 2019

Prior to Spring 2019, our group has engaged in efforts across all areas of Figure 1-1. With respect to the choice of ligands, we have used genetic algorithms³ and direct enumeration⁴ to generate candidates for study. Regarding ligand symmetry, our studies at the time had looked mostly at compounds with full symmetry (A_6) and high symmetry (A_5B_1 , A_4B_2 -trans). For the choice of metal, our group had typically looked at the first-row ($3d$) elements Cr, Mn, Fe, and Co. To accelerate predictions for and understand the property-determining function, we had developed tools such as mAD to run DFT calculations on GPUs using MolSimplify, used RACs with KRR or neural networks to predict properties such as redox potential and spin-splitting energy,⁵ made efforts to quantify uncertainty with machine learning methods, and developed various theoretical insights and design principles.⁶ Additionally, for the arrow flowing out of the property-determining function, we had developed static and dynamic classifiers to truncate DFT jobs that are not likely to converge⁷.

Although our efforts provided great coverage of factors that contribute to faster and broader discovery, more still had to be done in order to make property predictions in new regions of chemical space. Thus, in this thesis, I have focused on expanding database coverage by studying $4d$ metals, new symmetry classes, and chemically-synthesizable compounds. Additionally, after gaining insights from this data and developing machine learning models in these new chemical spaces, I was interested in translating these models into a form usable by experts, leading to the web tools molSimplify, MOFSimplify, and Ligandify. In the following sections, I provide a motivation and broad overview of the three main aims of my thesis.

1.2 Database population: symmetry classes and isovalents

Symmetry classes and isovalents are of interest because they can improve tunability of the ligand field as well as provide access to interesting and new chemistries. Previously, our group only looked at a limited set of $3d$ complexes with high-symmetry classes such as A_6 , A_5B_1 , and A_4B_2 -trans, which are shown in Figure 1-2.

Symmetry classes are hard to study experimentally because of synthetic restrictions (i.e. some arrangements of ligands are difficult to create in the lab). They are also hard to study computationally because of combinatorial explosion; according to Polya’s Enumeration theorem, a total of 29,260 complexes can be made from a pool of

10 ligands.⁴ When the effect of the metal center (i.e., $3d$ vs. $4d$) or ligand connecting atom (i.e., $2p$ vs. $3p$) is taken into account, this number increases drastically. By populating a database containing diverse TMCs spanning metal identities and ligand field symmetries, we can improve the robustness of our machine learning models in broader areas of chemical space.

1.3 Machine learning for property prediction

Machine learning for property prediction has emerged as a promising way to accelerate inorganic chemical design, particularly in the context of transition metal complexes (TMCs) and metal-organic frameworks (MOFs). Traditional computational approaches, such as density functional theory, can be resource-intensive and time-consuming, making it infeasible to explore this immense chemical space exhaustively. Machine learning models offer a powerful and efficient alternative, enabling rapid and accurate property prediction based on the patterns and relationships present in the training data. By employing machine learning techniques, we can effectively navigate the expansive inorganic chemical design space, accelerate the discovery of promising TMCs and MOFs, and ultimately contribute to advancements in sustainability and chemical synthesis. In this work, the data generated for isovalents and structures derived from the Cambridge Structural Database (CSD) was used to create machine learning models to predict $4d$ TMC properties and classify the synthesizability of novel complexes.

1.4 User interaction with machine learning models

As the methods for inorganic chemical design advance, the development of effective and user-friendly tools becomes increasingly essential for researchers to efficiently explore the vast chemical space of transition metal complexes (TMCs) and metal-organic frameworks (MOFs). User interactions with these models can be recorded to determine areas of current interest to practitioners, and the feedback from these interactions can be incorporated into machine learning models so that they can adapt and improve based on user feedback, ensuring that they remain relevant and accurate in an ever-evolving research landscape. In this work, MOFSimplify, molSimplify Lite, and Ligandify were developed to help bridge the gap between machine learning models and end users.

1.5 Thesis overview

My thesis has made several contributions towards a comprehensive exploration of inorganic chemical design space, leveraging computational chemistry, machine learning, and web-based tools to address the challenges associated with high-throughput virtual screening workflows. This thesis is organized into seven chapters, each addressing a specific aspect of inorganic chemical discovery.

The first three chapters set the foundation for the thesis by delving into the fundamental aspects of TMCs and their design. Chapter 1 introduces the problem areas associated with TMC design. Chapters 2 and 3 focus on the exploration of transition metal chemical space, with an emphasis on ligand additivity relationships and isovalent transition metal complex structures. These chapters provide a basis for understanding under-explored regions of TMC design space and illustrate ways in which chemical space exploration can be accelerated.

Chapters 4 through 6 build upon the foundational knowledge by focusing on tools that directly enhance the HTVS process from both a computational and user-based perspective. Chapter 4 discusses how parallel computing accelerates training data curation, which can improve the efficiency of computational approaches. Chapter 5 introduces interactive design tools for TMCs and metal-organic frameworks (molSimplify Lite and MOFSimplify, respectively), allowing users to explore and provide feedback on machine learning models. Chapter 6 presents a novel approach to ligand compatibility classification using stochastic negative addition, enabling interactive investigation of new chemical structures, and enhancing the ability to explore new TMC designs.

Finally, Chapter 7 offers recommendations for future work in the field, with a focus on fully autonomous workflows and the incorporation of web-based feedback into machine learning models. This chapter outlines the potential directions for advancing the state of the art in TMC design and further improving the efficiency and effectiveness of HTVS workflows.

1.6 References

- ¹ W. Jiang, N.J. DeYonker, J.J. Determan, and A.K. Wilson, *The Journal of Physical Chemistry A* **116**, 870 (2011).
- ² H. Bi, D. Chen, D. Li, Y. Yuan, D. Xia, Z. Zhang, H. Zhang, and Y. Wang, *Chemical Communications* **47**, 4135 (2011).
- ³ J.P. Janet, L. Chan, and H.J. Kulik, *The Journal of Physical Chemistry Letters* **9**, 1064 (2018).
- ⁴ S. Gugler, J. P. Janet, and H. J. Kulik, *Mol. Syst. Des. Eng.* **5**, 139–152 (2020).
- ⁵ A. Nandy, C. Duan, J.P. Janet, S. Gugler, and H.J. Kulik, *Industrial & Engineering Chemistry Research* **57**, 13973 (2018).
- ⁶ J.P. Janet, F. Liu, A. Nandy, C. Duan, T. Yang, S. Lin, and H.J. Kulik, *Inorganic Chemistry* **58**, 10592 (2019).
- ⁷ C. Duan, J.P. Janet, F. Liu, A. Nandy, and H.J. Kulik, *Journal of Chemical Theory and Computation* **15**, 2331 (2019).

2 Ligand additivity relationships enable efficient exploration of transition metal chemical space

This work has previously appeared as N. Arunachalam, S. Gugler, M.G. Taylor, C. Duan, A. Nandy, J.P. Janet, R. Meyer, J. Oldenstaedt, D.B. Chu, and H.J. Kulik, Ligand additivity relationships enable efficient exploration of transition metal chemical space, *The Journal of Chemical Physics* **157**, 184112 (2022).

2.1 Introduction

In recent years, virtual high-throughput screening (VHTS)^{1,2,3,4,5,6,7,8} with first-principles density functional theory (DFT) and machine learning (ML) models^{9,10,11,12,13,14,15,16,17} has greatly accelerated the discovery of new molecules and materials.^{18,19,20,21,22,23} Nevertheless, the theoretical space of all possible compounds or materials is so large as to challenge even the most accelerated methods, with 10^{30} – 10^{60} theoretical drug-like molecules being enumerated^{24,25,26,27,28,29} from a relatively small number of elements and atoms.^{30,31} Within the space of theoretical transition metal complexes, additional variables emerge, such as the metal identity, spin, and oxidation state, as well as denticity of the ligands.^{4,9} Indeed, significant analysis has been carried out by Fey and co-workers in understanding the role of privileged (e.g., phosphine) ligands in determining transition metal complex properties.^{32,33,34} Jensen and co-workers devised elegant strategies to explore the space of favored complexes, e.g., by adjusting the denticity during complex optimization carried out with efficient semi-empirical- or force-field-based scoring.^{35,36,37,38} One key challenge for high-throughput screening with DFT of transition metal complex space is that the smallest non-trivial mononuclear octahedral complex consists of at least seven heavy atoms and nearly 100 electrons, also challenging the speed of conventional simulation techniques in comparison to readily computed datasets of closed-shell transition metal³⁹ and organic molecules.^{40,41} Further compounding the challenges of exploring transition metal chemical space are potential issues with convergence success or the presence of multireference character.^{42,43,44} Thus,

it is attractive to identify the minimal set of explicit first-principles calculations that can be used to build a model of the properties of the full dataset.^{45,46,47}

Recently, toward the goal of exploring a more diverse transition metal chemical space in comparison to complexes comprising of frequently studied ligands, we devised a strategy for enumerating hypothetical, small (i.e., 1–2 heavy atoms per coordination site) ligands for mononuclear octahedral transition metal complexes.⁴⁸ These ligands sampled a diverse combination of coordinating atoms and their bonding environments,⁴⁸ and only a small fraction were represented in prior databases of organic molecules.^{25,49,50} We showed how incorporating these molecules could improve the fidelity of artificial neural network (ANN) models^{48,51} when applied to larger, realistic complexes present in the Cambridge Structural Database (CSD).⁵² That study was limited to the properties of homoleptic combinations of those ligands (i.e., all ligands are the same) and therefore did not capture effects of mixing ligands that give rise to the compelling properties of many heteroleptic complexes and catalysts. Enumerating combinations of these ligands, however, would give rise to a combinatorial explosion, motivating strategies to understand which combinations are likely to be valuable or informative for a specific application.

Despite the challenge of combinatorial explosion, there are some established precedents of ligand additivity⁵³ that suggest that the properties of heteroleptic complexes can be inferred from combinations of homoleptic complexes.⁵⁴ For example, ligand additivity has been demonstrated in force field and DFT energetics⁵⁴ and DFT errors.⁵⁵ It has also been used in correction schemes, such as the DBLOC method.^{56, 57, 58, 59} We also recently exploited additivity to learn the degree of multireference character in a complex from the multireference character in its constituent ligands.⁶⁰ Additivity is also exploited heavily in fragmentation methods^{61,62} and locally correlated methods.^{63,64} In the present work, we carry out a survey of the symmetry classes and ligand diversity present in the CSD to confirm that the theoretical chemical space is orders of magnitude larger than the number that have been characterized. Motivated by the need to devise efficient but accurate methods for the exploration of chemical space, we introduce improved interpolation schemes for heteroleptic compounds to incorporate *cis* and *trans* effects. Finally, we demonstrate how these approaches can be used for efficient but accurate discovery of transition metal complexes with targeted properties.

2.2 Methods

2.2.1 Curation from the Cambridge Structural Database

A set of 85 575 mononuclear octahedral transition metal complexes were curated from the Cambridge Structural Database⁵² (CSD) version 5.41 (November 2019). This procedure employed both the Conquest graphical interface to the CSD and the Python application programming interface, in all cases applied to the v5.41 dataset with complexes from the November 2019 dataset with both the March 2020 and May 2020 data updates (Appendix A, Text A-1). For the complexes identified as octahedral, equatorial planes and axial positions were assigned based on prior reported rules.⁶⁵ To identify the symmetry of the ligands, unique ligands were identified by removing the metal atom to create independent molecular graphs for each ligand. Each ligand was identified as chemically unique within a given octahedral complex if it differed from all other ligands in the complex by (1) heavy atom chemical symbols, (2) metal-connecting-atom element, or (3) more than three hydrogens (Appendix A, Text A-2). The symmetry of the complex was identified by distinguishing ligand denticity overall and in the equatorial plane along with the total number of unique ligands and whether ligands that were *trans* to each other were identical (Appendix A, Text A-2). This led to a nomenclature for 66 ligand symmetry classes (Appendix A, Text A-2).

To identify the set of unique ligands in each complex, a dummy atom with identical connectivity to the metal with an atomic number of 0 was introduced to preserve the connectivity of the ligands to the metal without preserving the metal identity. For this ligand and dummy atom combination, the atomic-number and bond-order weighted connectivity matrix determinant was calculated as described in Ref. 65. We also computed the determinant of the atomic-number and bond-order weighted connectivity matrix where the off-diagonal elements, $Z_i Z_j$ ($i \neq j$), were set to the CSD-assigned bond order for each ligand. Ligands with both distinct atomic-number-weighted connectivity matrix determinants and bond-order-weighted connectivity matrix determinants were identified as distinct ligands across monometallic transition metal complexes in the CSD. A second search was carried out by requiring that oxidation states and charges be assigned by the uploader along with no disorder (i.e., as judged by the CSD flags) or missing hydrogen atoms in the structure (i.e., none were added by the CSD algorithm), leading to 17 085 unique “computation-ready” complexes. Finally, we curated a subset of 1202 Fe(II)-containing “computation-ready” complexes based on the oxidation state reported by the uploader. From the ligands identified in this Fe(II)

complex set, heteroleptic calculations from CSD ligands were carried out using a previously developed procedure⁶⁰ that enabled the assignment of the per-ligand charge. The goal of this curation is for subsequent analysis outlined in Sec. 3. We note that complementary datasets have been developed, including the tmQM dataset of 86k closed-shell transition metal complexes,⁶⁶ the cell2mol set of 31k inferred transition metal complex charges from the crystal unit cell along with 13k ligand charges,⁶⁶ and our set of over 5k transition metal complex ligands with confident charge assignment.⁶⁰ Our rationale for curating another set of CSD transition metal complexes is threefold: (i) to focus on open-shell transition metal chemistry absent from some of the aforementioned sets, (ii) to leverage our recently developed ligand-derived charge scheme,⁶⁰ and (iii) to primarily analyze the theoretical vs observed propensities of ligand types, densities, and complex symmetries in the CSD.

2.2.2 Electronic structure calculations

DFT geometry optimizations were performed using a development version of TeraChem v1.9.^{67,68} The B3LYP^{69,70,71} global hybrid functional was employed with the LANL2DZ⁷² effective core potential for transition metals and the 6-31G* basis⁷³ for all other atoms. All transition metal complexes were studied with Fe(II) centers in low-spin singlet and high-spin quintet multiplicities. Singlet calculations were carried out in a spin-restricted formalism, while quintet calculations were unrestricted. Level shifting⁷⁴ was employed to aid self-consistent field convergence with the majority-spin and minority-spin virtual orbitals each shifted by 0.25 Ha. Geometry optimizations were carried out in the gas phase in translation rotation internal coordinates⁷⁵ using the BFGS algorithm. Default tolerances of 4.5×10^{-4} hartree/bohr and 10^{-6} hartree were applied in the convergence criteria for the maximum gradient and energy difference between steps, respectively.

For the representative model complexes of CH₃CN, H₂O, CO, and NH₃, initial structures were generated with molSimplify,^{51,76,77} which uses OpenBabel^{78,79} as a backend. The same protocol was applied to generate homoleptic complexes of the 20 neutral ligands derived from monodentate-only, non-homoleptic Fe(II) complexes obtained from the CSD⁵² after discarding one bulky ligand, OP(Ph)₃, that could not form a stable homoleptic structure for steric reasons. For the 36 homoleptic Fe(II) complexes in the CSD, the structures were directly extracted for subsequent geometry optimization. Heteroleptic complexes of the 12 representative ligands were also generated with molSimplify and optimized following the same procedure. The curated

CSD database and structures, as well as the scripts to generate these databases and structures, along with the properties of all transition metal complexes studied with DFT are provided on Zenodo with DOI 10.5281/zenodo.7224793.⁸⁰

2.3 Results and discussion

2.3.1 Symmetry classes and theoretical complex space

The diversity present in the chemical space of transition metal complexes is derived from the variability in the metal, its oxidation and spin state, and the chemistry of the coordinating ligands. Our experimental knowledge of this chemical space is unevenly distributed. We thus first examined the structures deposited in the CSD to uncover trends in the arrangement of ligands in previously characterized complexes (see Sec. 2). From 85 575 mononuclear octahedral transition metal complexes of which 17 085 are identified as unique and computation-ready (e.g., have user-defined charges), the vast majority (95%–98%) contain no more than three unique ligands (Figure 2-1 and Appendix A, Table A-1). In fact, 28% of all unique computation-ready complexes are homoleptic, and a majority (76%) contain no more than two ligand types (Figure 2-1 and Appendix A, Table A-1). Because the metal identity and ligand diversity are expected to be coupled, we also evaluated statistics on a subset of 1202 unique Fe(II) complexes and confirm that the preference for complexes with no more than two ligands is preserved and even strengthened (Figure 2-1 and Appendix A, Table A-1). We, nevertheless, further distinguish the denticity of these ligand types because monodentate ligands are theoretically compatible with a wider range of symmetries than higher denticity ligands (Figure 2-1). In subsequent analysis in this paper, we will only focus on monodentate ligands for this reason and note, therefore, that while two is the most common number of unique ligands, many of those ligands are bidentate and their study is motivated in the future.

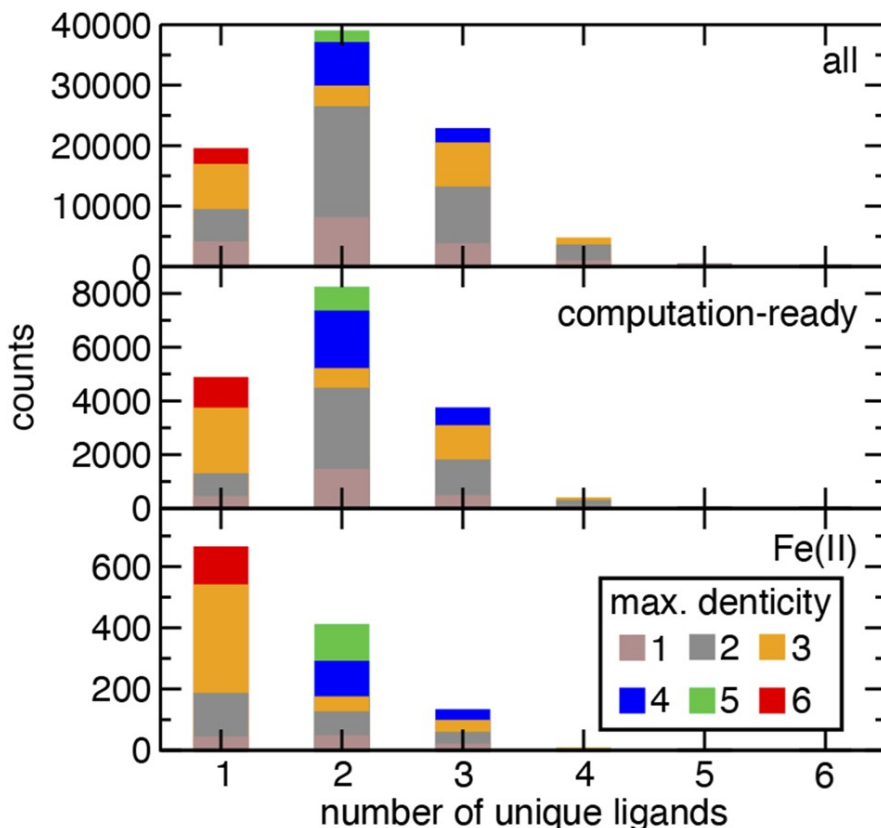


Figure 2-1. Stacked, unnormalized histogram of the number of complexes in the CSD grouped by the number of unique ligand types and by the highest denticity of ligands in the complex (monodentate in brown, bidentate in gray, tridentate in orange, tetradentate in blue, pentadentate in green, and hexadentate in red, as indicated in the inset). These counts are shown for all complexes (top), unique and computation-ready complexes (middle), and the Fe(II) unique, computation-ready subset (bottom).

To identify how the structures sampled in the CSD compare to the theoretical space of all hypothetical complexes, we enumerate the overall pool of theoretical complexes and the number in each symmetry class to compare to the most frequently characterized symmetry classes in the CSD. We applied Pólya's enumeration theorem to octahedral coordination geometries to obtain all possible symmetry classes from the cycle index (i.e., theoretical sum) of a symmetry group (Appendix A, Text A-3).^{81,82,83} The total number of complexes depends on both the number of ways the same stoichiometry can be arranged and the number of ways the ligands can be combined (Appendix A, Text A-3). For example, in $(L_1)_4(L_2)_2$, there is one occurrence of a four-substitution site and one occurrence of a two-substitution site, whereas in $(L_1)_2(L_2)_2(L_3)_2$, there are three occurrences of a two-substitution site. In total, the

theoretical number of ways distinct ligands can be enumerated together leads to a large number of hypothetical complexes even for a small pool of ligands (Table I).

For the trivial case of homoleptic (HO) complexes, N ligands produce N complexes (i.e., 12 for $N = 12$, Table 2-1). For up to two unique ligand types, the five unique symmetry classes consist of monoheteroleptic (5+1) $M(L_1)_5(L_2)_1$ complexes, *trans* symmetric (TS) or *cis* symmetric (CS) $M(L_1)_4(L_2)_2$ complexes, and *fac* symmetric (FS) or *mer* symmetric (MS) $M(L_1)_3(L_2)_3$ complexes (Figure 1-1 and Table 2-1). Both 5+1 and CS/TS complexes each form $N(N-1)$ complexes for N ligands (i.e., 132 each for $N = 12$, Table 2-1). The degeneracy of the stoichiometry in FS and MS complexes gives rise to $N(N-1)/2!$ complexes (i.e., 66 each for $N = 12$, Table I). Thus, from $N = 12$ ligands, a total of 540 complexes may be formed with up to two unique ligand types.

Table 2-1 The number of theoretical complexes for each octahedral symmetry class considered in this work and the full octahedral space for an example single metal/oxidation/spin state ($m = 1$) with an $N = 12$ ligand pool. The configurations and isomers indicate the number of ways unique ligands can be arranged, and the cardinality indicates how many theoretical complexes can be enumerated.

Name	Configuration	Isomers	Cardinality	Complexes
HO	x_6	1	N	12
5+1	x_1x_5	1	$N(N-1)$	132
TS/CS	x_4x_2	2	$N(N-1)$	264
FS/MS	x_3x_3	2	$N(N-1)/2!$	132
CA/TA	$x_4x_4x_1$	2	$N(N-1)(N-2)/2!$	1 320
FA/MAC/MAT	$x_3x_2x_1$	6	$N(N-1)(N-2)/2!$	3 960
EA/DCS/DTS	$x_2x_2x_2$	5	$N(N-1)(N-2)/3!$	1 100
Up to two ligands				540
Up to three ligands				6 920
Full			$1/48 (N^6 + 3N^5 + 9N^4 + 13N^3 + 14N^2 + 8N)$	82 160

Expanding to up to three unique ligand types introduces eight additional symmetry classes (Figure 2-3). These include $M(L_1)_4L_2L_3$ *cis* asymmetric (CA) and *trans* asymmetric (TA) complexes; the three types of $M(L_1)_2(L_2)_2(L_3)_2$ configurations in equatorial asymmetric (EA), double *cis* symmetric (DCS), or double *trans* symmetric

(DTS) symmetries; and three $M(L_1)_2(L_2)_3L_3$ in *fac* asymmetric (FA), *mer* asymmetric *trans* (MAT), or *mer* asymmetric *cis* (MAC) symmetries (Fig. 3). For the EA complexes, there are three occurrences of a two-substitution site that fulfill the EA definition. This combines with the DCS and DTS isomers to form a total of five isomers with $N(N-1)(N-2)/3!$ possible combinations of ligands (i.e., 1100 for $N = 12$, Table 2-1). The less degenerate CA/TA complexes form a total of $N(N-1)(N-2)/2!$ complexes for each of the two isomers (i.e., 1320 for $N = 12$, Table 2-1). Similarly, FA/MAC/MAT complexes can each form $N(N-1)(N-2)/2!$ complexes in two isomers each (i.e., 3960 for $N = 12$, Table 2-1).

In total, 6920 complexes can be formed from $N = 12$ ligands for the three symmetry classes considered here. For the same ligand pool, there is a much larger set of 82 160 theoretical complexes that could be created from a greater number of unique ligands. This analysis does not consider cases where the ligand chemistry prevents the formation of a complex, e.g., only monodentate and pentadentate ligands are likely to form the 5+1 symmetry class, whereas only monodentate, bidentate, tridentate, or hexadentate ligands can form HO complexes.

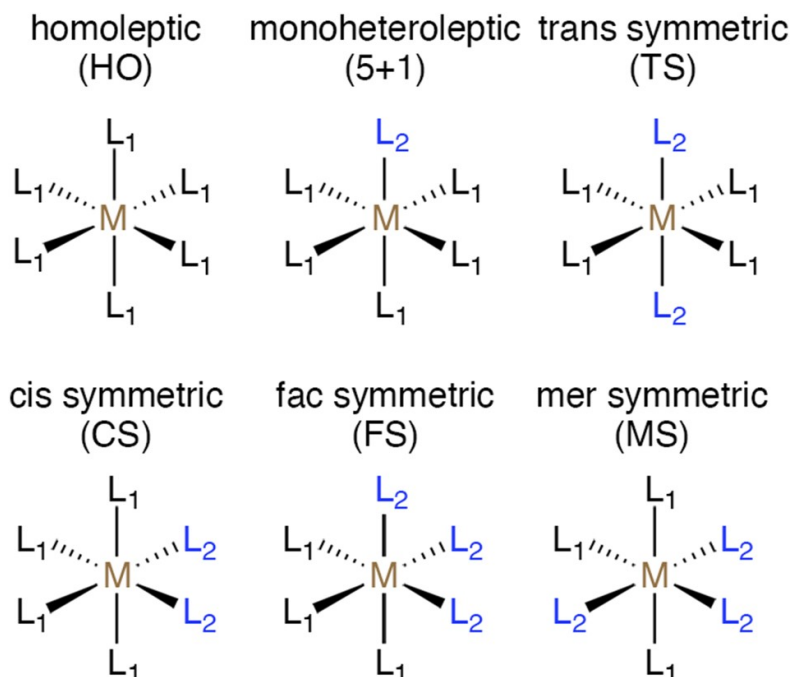


Figure 2-2. Symmetry classes for transition metal complexes with up to two unique ligands, L_1 and L_2 , from left to right and top to bottom: homoleptic (HO) $M(L_1)_6$, monoheteroleptic (5+1) $M(L_1)_5L_2$, *trans* symmetric (TS) $M(L_1)_4(L_2)_2$, *cis* symmetric (CS) $M(L_1)_4(L_2)_2$, *fac* symmetric (FS) $M(L_1)_3(L_2)_3$, and *mer* symmetric (MS) $M(L_1)_3(L_2)_3$. For each pair of ligands, a total of two homoleptic and eight two-ligand

isomers can be obtained because the 5+1, *trans* symmetric, and *cis* symmetric complexes are unique if the stoichiometry of L₁ and L₂ are swapped.

Returning to the diversity observed in complexes deposited in the CSD, we qualitatively observe that relatively little of the theoretical space has been sampled. As we have shown for a representative example, for any set of unique ligands, a much larger theoretical number of binary and ternary complexes can be formed in comparison to homoleptic complexes. Nevertheless, there are far fewer binary and especially ternary complexes in the CSD (Appendix A, Table A-1). Within the binary complexes, 5+1 and TS complexes are overrepresented in comparison to those for FS, MS, or CS based on our theoretical enumeration (Figure 2-4 and Appendix A, Table A-2). All ternary complexes are underrepresented, but those with equal stoichiometry (i.e., EA, DCS, or DTS) have very few examples in the CSD (Table 2-1 and Appendix A, Table A-2). If we simplify our analysis by focusing only on Fe(II) complexes, the same trends hold although Fe(II) complexes have an even greater relative number of 5+1 and TS complexes and a lower number of CS complexes (Figure 2-4 and Appendix A, Table A-2).

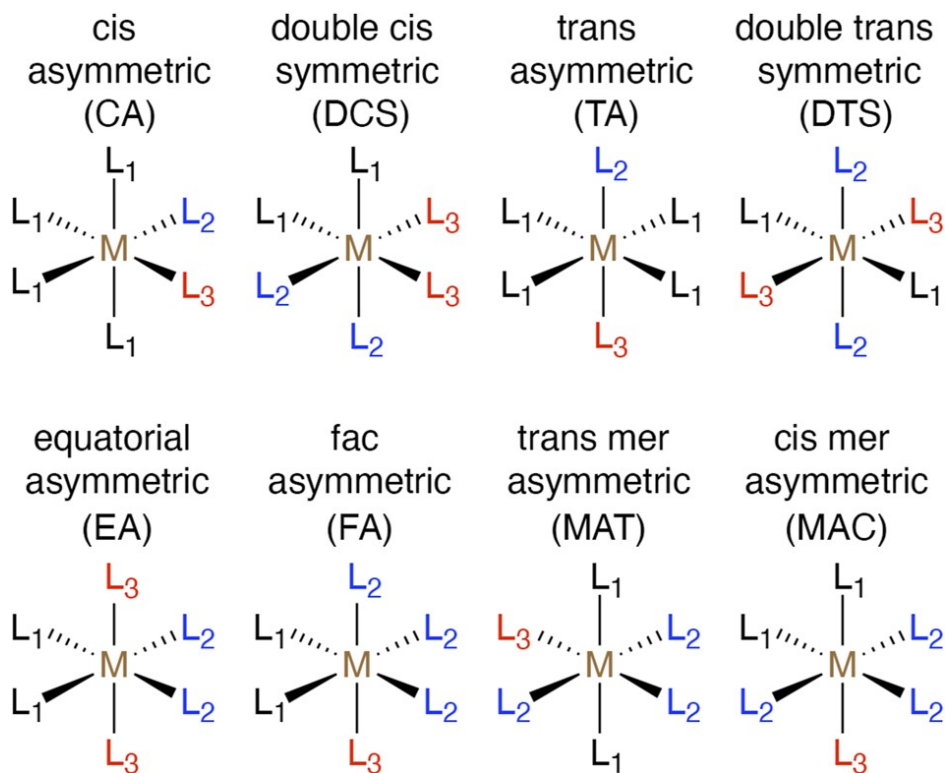


Figure 2-3. Symmetry classes for transition metal complexes with three unique ligands, L₁, L₂, and L₃, from left to right and top to bottom: *cis* asymmetric (CA) M(L₁)₄L₂L₃,

double *cis* symmetric (DCS) $M(L_1)_2(L_2)_2(L_3)_2$, *trans* asymmetric (TA) $M(L_1)_4L_2L_3$, double *trans* symmetric (DTS) $M(L_1)_2(L_2)_2(L_3)_2$, equatorial asymmetric (EA) $M(L_1)_2(L_2)_2(L_3)_2$, *fac* asymmetric (FA) $M(L_1)_2(L_2)_3L_3$, *trans mer* asymmetric (MAT) $M(L_1)_2(L_2)_3L_3$, and *cis mer* asymmetric (MAC) $M(L_1)_2(L_2)_3L_3$. A total of 29 complexes can be obtained for any combination of three ligands due to additional isomers of the equatorial asymmetric type and those for which the stoichiometry of each ligand type is not equal.

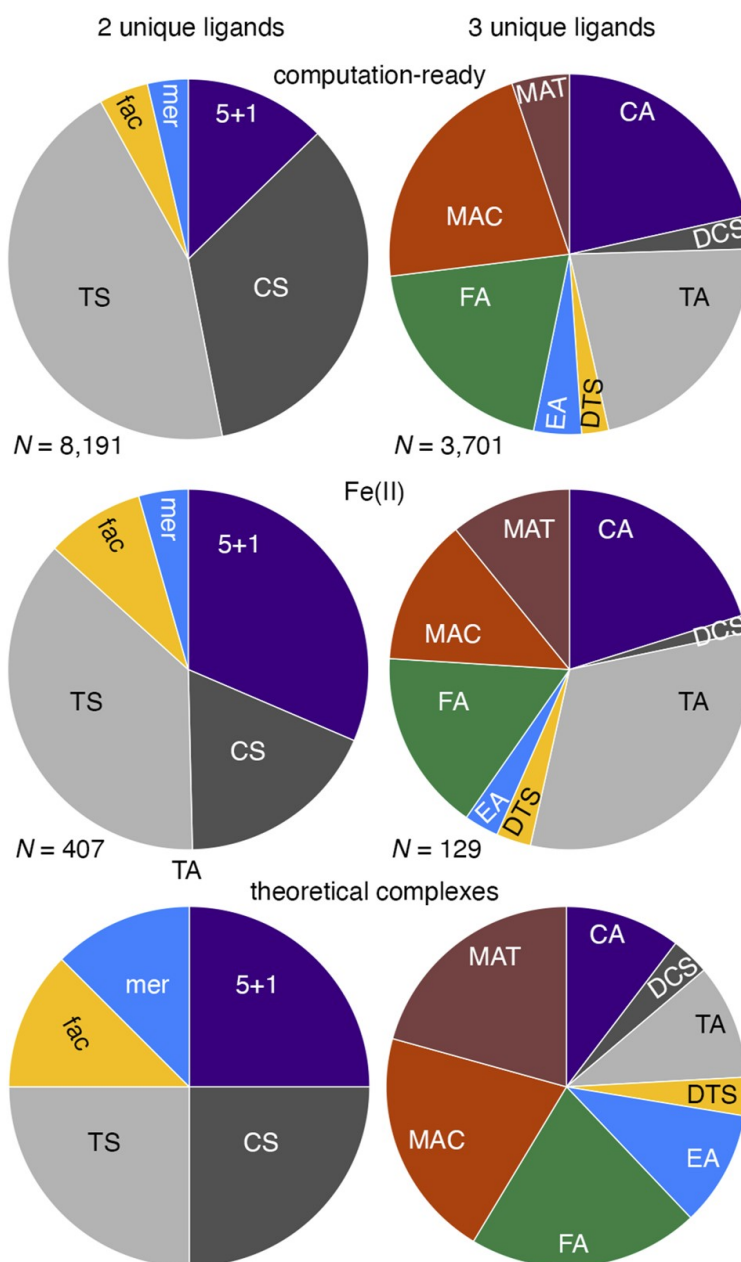


Figure 2-4. Percent of all unique mononuclear octahedral transition metal complexes in the CSD with user-defined charges (top, computation-ready) and the Fe(II) subset [middle, Fe(II)], grouped by a symmetry class for cases with two unique ligands (left: 5_+1 , *cis* symmetric, CS, *trans* symmetric, TS, *fac*, or *mer*) or three unique ligands (right: *cis* asymmetric, CA, double *cis* symmetric, DCS, *trans* asymmetric, TA, double *trans* symmetric, DTS, equatorial asymmetric, EA, *fac* asymmetric, FA, *cis mer* asymmetric, MAC, or *trans mer* asymmetric, MAT). The ratio of symmetry classes for the theoretical complexes from enumeration is shown at the bottom for comparison.

An additional factor in analysis of symmetry classes is the extent to which ligand denticity plays a role. While enumeration is straightforward for monodentate ligands, higher-denticity ligands (e.g., pentadentates) may only be compatible with some of the symmetry classes. Indeed, over both the computation-ready CSD set and the Fe(II) subset, a significant number of complexes consist of higher-denticity ligands that would be incompatible with full enumeration (Figure 2-1 and Appendix A, Table A-3 and Table A-4). Some multidentate ligands are also restricted in which symmetry classes they can form due to rigidity, while other more flexible ligands are less restricted. Nevertheless, we can still conclude that there is vastly higher sampling of homoleptic structures. For example, tridentate ligands can be present in a homoleptic complex and any binary FS/MS or ternary TA/FA/MAC/MAT complex. Nearly an order-of-magnitude more unique tridentate ligands have been characterized in homoleptic Fe(II) complexes in comparison to either the binary or ternary cases (Appendix A, Table A-4). Similar trends hold for all unique computation-ready complexes (Appendix A, Table A-3). Overall, the number of unique ligands for binary or ternary complexes is still lower than the number of unique complexes, but the gap is smaller than could be expected from enumeration alone (Appendix A, Table A-3 and Table A-4).

To simplify a quantitative comparison between the theoretical space and the enumerated space, we focus on Fe(II) complexes with only monodentate ligands in any of the 14 symmetry classes considered. For this set, we identify 40 unique monodentate ligands present in 40 HO complexes (Appendix A, Table A-5). There are additional 48 ligands present in binary or ternary complexes not observed in homoleptics for which we can confidently assign a charge to the ligand (Appendix A, Table A-5). While the majority of HO complex monodentate ligands were neutral, over half of the additional binary or ternary ligands have a non-zero charge (Appendix A, Table A-6). Thus, although we identify 88 unique monodentate ligands in previously synthesized Fe(II)

complexes, ligands not sampled in HO complexes may be incompatible with the HO symmetry class if they give rise to high overall complex charges.

Taking the set of 88 ligands from any HO, binary, or ternary Fe(II) complex as the theoretical space for which compatibility across symmetry classes should be maximal, we then quantified the theoretical vs actual coverage of Fe(II) complex chemical space. A significant number (39%) of all HO complexes have been characterized. In comparison, the binary symmetry classes have been much less well explored, with TS complexes being the highest at 0.3% (i.e., 26 of 7656 theoretical complexes and Appendix A, Table A-6). The number of theoretical ternary complexes grows rapidly with this ligand pool, ranging from 109 736 theoretical DTS/DCS complexes to 329 208 FA/MAC/MAT/CA/TA complexes (Appendix A, Table A-6). In total, the 1202 Fe(II) complexes represent a tiny fraction of the theoretical 3 213 056 homoleptic, binary, or ternary complexes that could form from 88 experimentally synthesized ligands. Of the 88 ligands, we exclude one [i.e., OP(Ph)₃] from further analysis due to its large bulk that prevents building a homoleptic complex that remains intact after geometry optimization. Even if we restrict ourselves to the 56 ligands that are neutral, closed-shell singlets, and amenable to homoleptic complex construction, the theoretical space of homoleptic, binary, or ternary complexes is still large (i.e., 816 256 complexes). Thus, we conclude that efficient strategies to infer heteroleptic properties from homoleptic properties are necessary to “fill in” the remainder of this unexplored space.

2.3.2 Ligand additivity for interpolating properties of transition metal complexes

We next aimed at determining the extent to which the properties of lower-symmetry heteroleptic transition metal complexes could be inferred from those of higher-symmetry complexes. We constructed complexes with two to three unique ligand types from small sets of ligands that spanned a large range of ligand field strengths: weak-field water, strong-field carbonyl, and either strongfield methyl isocyanide or weak-field ammonia. For the calculation of the adiabatic high-spin [e.g., quintet Fe(II)] to low-spin [e.g., singlet Fe(II)] splitting, $\Delta E_{\text{H-L}}$, it can be expected that the weakfield ligands will lead to homoleptic complexes that favor high-spin states, whereas strong-field ligands will make homoleptic complexes that favor low-spin states. Thus, heteroleptic combinations of these ligands are expected to reside between the two limits.

One simple way to obtain estimates of the spin splitting of the heteroleptic complexes (e.g., with up to three unique ligand types) is to take a weighted average of the spin splitting of the parent homoleptic complexes,

$$E(M(L_1)_x(L_2)_y(L_3)_{6-x-y}) = \frac{x}{6}E(M(L_1)_6) + \frac{y}{6}E(M(L_2)_6) + \frac{6-x-y}{6}E(M(L_3)_6)$$

(Equation 2-1)

Indeed, we observe that heteroleptics reside between the homoleptic limits for spin splitting, and the linear averaging roughly holds for complexes with methyl isocyanide (CH₃CN), H₂O, and CO ligands (Figure 2-5 and Appendix A, Table A-7). Similar observations can be made on combinations with H₂O, CO, and NH₃ (Appendix A, Figure A-1). Nevertheless, there are significant outliers in the interpolated vs actual $\Delta E_{\text{H-L}}$, which are particularly evident when comparing heteroleptic complexes with the same stoichiometry and, therefore, the same prediction from a simple linear model, but distinct ligand symmetry (Figure 2-5). For example, the CS complex Fe(II)(H₂O)₄(CO)₂ $\Delta E_{\text{H-L}}$ is predicted accurately (predicted: -7.9 kcal/mol vs calculated: -9.9 kcal/mol) from homoleptic interpolation (Figure 2-5 and Appendix A, Table A-8). The same prediction significantly overestimates the TS complex with the same stoichiometry (predicted: -7.9 kcal/mol vs calculated: -19.3 kcal/mol), which instead behaves much more similarly to the 5+1 complex, as observed in prior work⁸⁴ (Figure 2-5). Overall, for this combination of ligands, CS or FS complexes that have one minority ligand in the axial position and another in the equatorial plane are much better predicted than the equivalent MS or TS complexes (Figure 2-5). For the case of CO, H₂O, and NH₃, mixing between weak-field NH₃ and H₂O is relatively accurately predicted from homoleptic averaging, whereas for NH₃ and CO, it is the MS and TS complexes that are more accurately predicted than the FS or CS counterparts (Appendix A, Figure A-2, Table A-8, and Table A-9).

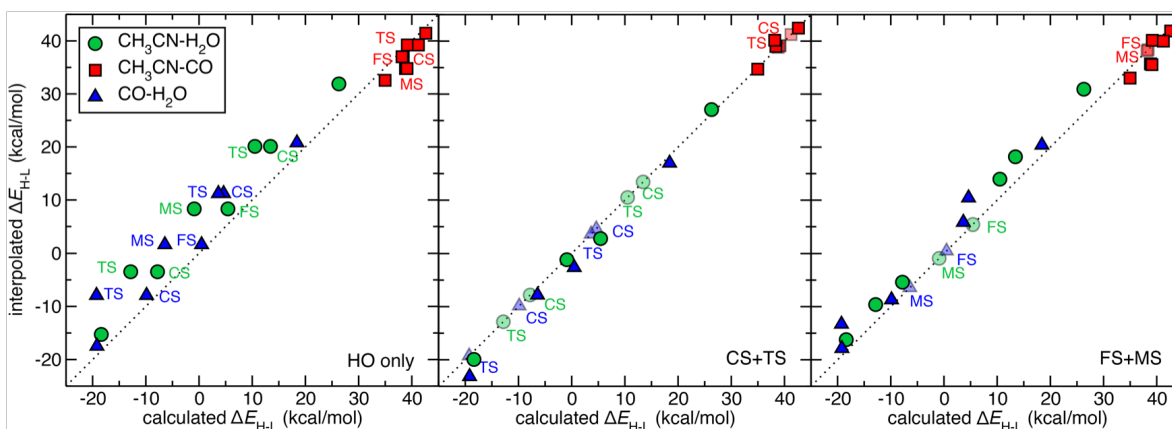


Figure 2-5. Calculated vs linearly interpolated ΔE_{H-L} (kcal/mol) for Fe(II) complexes with pairs of any of the three ligands: CH_3CN , H_2O , and CO . From left to right: interpolation between homoleptic complexes (HO only), interpolation using homoleptic complexes and CS and TS complex energies (CS+TS), or interpolation using homoleptic complexes and FS and MS complex energies (FS+MS). Points are colored according to the pair of ligands they correspond to: $\text{CH}_3\text{CN-H}_2\text{O}$ (green circles), $\text{CH}_3\text{CN-CO}$ (red squares), and $\text{CO-H}_2\text{O}$ (blue triangles), as indicated in the inset. Key isomers are annotated. Points provided for the fit are translucent, whereas the remaining points are opaque. In all panels, a black dotted parity line is shown.

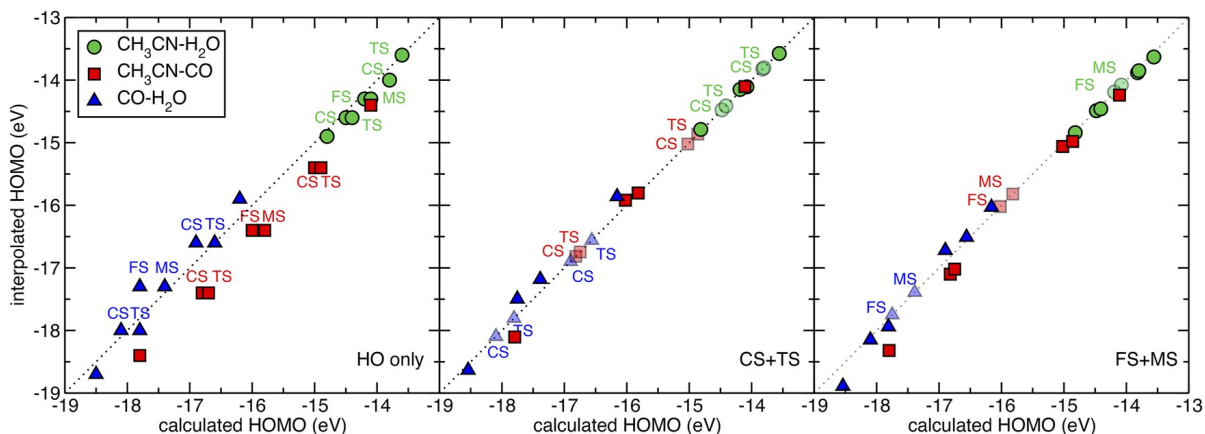


Figure 2-6. Calculated vs linearly interpolated HOMO level (eV) for singlet Fe(II) complexes with pairs of any of the three ligands: CH_3CN , H_2O , and CO . From left to right: interpolation between homoleptic complexes (HO only), interpolation using homoleptic complexes and CS and TS complex energies (CS+TS), or interpolation using homoleptic complexes and FS and MS complex energies (FS+MS). Points are colored according to the pair of ligands they correspond to: $\text{CH}_3\text{CN-H}_2\text{O}$ (green circles), $\text{CH}_3\text{CN-CO}$ (red squares), and $\text{CO-H}_2\text{O}$ (blue triangles), as indicated in the inset. Key

isomers are annotated. Points provided for the fit are translucent, whereas the remaining points are opaque. In all panels, a black dotted parity line is shown.

Because adiabatic spin splitting involves geometry optimizations in two distinct spin states, we also evaluated HOMO levels of each singlet complex as a property that only depends on a single geometry. Overall, interpolation of HOMO energies of heteroleptic complexes of CH₃CN, CO, and H₂O from homoleptic complexes reproduces trends between the homoleptic limits (Figure 2-6 and Appendix A, Table A-7, Table A-10, and Table A-11). As with spin more varied results for the HOMO level, with some cases occurring where the TS and FS complexes are less accurately predicted (Figure 2-6 and Appendix A, Table A-10). Nevertheless, over the full splitting, there are key differences for complexes with identical stoichiometry that cannot be captured by interpolation from homoleptic complexes alone (Fig. 6). Interestingly, for the same complexes for which the CS complex had a higher (i.e., more off-parity) $\Delta E_{\text{H-L}}$ and the TS complex was more like the equivalent 5+1 complex, we find range of data, the outliers in the HOMO level prediction are more modest than what was observed for $\Delta E_{\text{H-L}}$. These trends also hold for the complexes with NH₃, H₂O, and CO (Appendix A, Figure A-2, Table A-10, and Table A-11).

Given the differences between CS and TS complexes despite having identical stoichiometry, we next identified strategies for improving the interpolation. First, we employed CS and TS complexes along with homoleptic complexes to predict the spin-splitting energetics of the other binary heteroleptic complexes as follows:

$$E(5 + 1) = \frac{1}{2}(E(HO) + E(TS))$$

(Equation 2-2)

where $E(5+1)$ is the interpolated energy of the $M(L_1)_5L_2$ complex from the $M(L_1)_6$ HO energy and the $M(L_1)_4(L_2)_2$ TS energy, and we select the weights of the averaging here and throughout to reflect the stoichiometry of the final complex (i.e., here, one L_2 ligand and five L_1 ligands).

Similarly, we estimate the FS and MS energies as

$$E(FS) = \frac{1}{2}(E(CS M(L_1)_4(L_2)_2) + E(CS M(L_2)_4(L_1)_2))$$

(Equation 2-3)

$$E(MS) = \frac{1}{2}(E(TS M(L_1)_4(L_2)_2) + E(TS M(L_2)_4(L_1)_2))$$

(Equation 2-4)

where the weights are again chosen stoichiometrically, but we assume that *fac* complexes, which contain more *cis* interactions than *mer* complexes, are a better predictor of *cis* symmetric interactions and vice versa for the *trans* case. In practice, this corresponds to computing six energies and interpolating four remaining energies for each pair of ligands for a computational saving of 40%. Indeed, we observe reduced mean absolute error (MAE) over the remaining points that are interpolated (2.4 kcal/mol vs 5.1 kcal/mol) in comparison to the HO-only interpolation (Figure 2-5 and Appendix A, Table A-8 and Table A-9). In particular, FS complexes of CO and H₂O are now correctly predicted to be much more low-spin-directing than the MS complex of the same stoichiometry (Figure 2-5). The HOMO levels for these and other complexes are also improved (Appendix A, Table A-10 and Table A-11). For example, the modified interpolation is able to capture the fact that MS Fe(II)(CO)₄(H₂O)₂ has a shallower HOMO level than its FS counterpart (Figure 2-6). Nevertheless, not all points are uniformly improved by this interpolation. For the case of NH₃, CO, and H₂O where the interpolation already performed well, some points such as ΔE_{H-L} for 5+1 Fe(II)(CO)₅(NH₃) are slightly worsened in the modified interpolation (Appendix A, Figure A-1). For the same complex, the HOMO level is equivalently predicted by both interpolation schemes, and most HOMO level estimates are improved (Appendix A, Figure A-2). Overall, errors are on average significantly lower for all properties and sets of ligands considered when the modified interpolation expressions are employed.

Although the CS/TS-derived interpolation schemes greatly reduce errors in estimating the energetics of heteroleptic complexes, they still require significant computational overhead. Thus, we next aimed at identifying if FS and MS complexes, which contain three ligands *cis* to each other or two ligands *cis* and two sets of ligands *trans*, respectively, could be used instead in the interpolation (see Figure 2-2). If the FS and MS complexes impart sufficient information, using them in an interpolation scheme along with homoleptic complex properties corresponds to evaluating four properties (e.g., energies) to predict six properties for a computational savings of 60%. In this interpolation scheme, we estimated the complex properties from FS and MS complexes as follows:

$$E(5+1) = \frac{2}{3}E(HO) + \frac{1}{3}E(FS)$$

(Equation 2-5)

where $E(5+1)$ is the interpolated energy of the $M(L_1)_5L_1$ complex from the $M(L_1)_6$ HO energy and the $M(L_1)_3(L_2)_3$ FS energy, and the weights are selected based on stoichiometry. Similarly, we obtain expressions for the CS and TS complexes as

$$E(CS) = \frac{1}{3}E(HO) + \frac{2}{3}E(FS M(L_1)_3(L_2)_3)$$

(Equation 2-6)

$$E(TS) = \frac{1}{3}E(HO) + \frac{2}{3}E(MS M(L_1)_3(L_2)_3)$$

(Equation 2-7)

where the weights are chosen to match the stoichiometry of the final complexes and the *fac* complex is again chosen to better mimic the *cis* complex. Indeed, using this approach, we achieve errors only slightly larger than that for the CS/TS averaging scheme, with the added benefit of requiring fewer energies to obtain the same fidelity of the interpolation (Figure 2-5). For example, the higher spinsplitting energy of CS complexes relative to TS complexes is captured here because it can be directly derived from the strength of the high-spin directing character of the FS complex relative to that of the MS complex (Figure 2-5). Mixing of the weak-field NH_3 and H_2O ligands, which had slightly worsened with CS/TS interpolation, is also significantly improved in this scheme, and the other sets of complexes are of comparable accuracy (Appendix A, Figure A-1, Table A-8, and Table A-9). While, CS/TS interpolation tended to underestimate the HOMO level, FS/MS interpolation slightly overestimates HOMO levels, but errors are much smaller than for the HO-only interpolation (Figure 2-6 and Appendix A, Table A-10 Table A-11). For the set of ligands including ammonia, almost all points are predicted to comparable or slightly improved values (Appendix A, Figure A-2). Thus, HO-only interpolation provides a highly efficient scheme for predicting heteroleptic transition metal complexes, but the best trade-off in accuracy and computational cost for interpolation is likely achieved through estimating properties using information from FS and MS complexes.

We next investigated whether we could generalize our observations to heteroleptics with three unique ligand types (Figure 2-3). This extension is motivated by the fact that 96% of all mononuclear transition metal complexes in the CSD contain no more than three ligand types, and over 99% of Fe(II) complexes contain three or fewer

ligand types (Figure 2-1 and Appendix A, Table A-1). Expansion to three ligand types introduces 29 complex energies that need computation for any set of three ligands. HO-only averaging performs poorly here for complexes that are mixtures of CH₃CN, CO, and H₂O, with a large difference between CA and TA complexes of H₂O and CO being treated completely equivalently in this scheme (Figure 2-7). Similar differences in TA and CA HOMO levels are also missed in this averaging scheme (Figure 2-8). Generally, the CA complex spin-splitting energies are better predicted by the homoleptic averaging than the TA are for both ternary complexes with CH₃CN and with NH₃ (Figure 2-7 and Appendix A, Figure A-3, Table A-12, and Table A-13). For the HOMO level, results are more varied, with the HOMO energies of the CH₃CN ternary complexes being underestimated, while those of NH₃ ternary complexes are overestimated (Figure 2-8 and Appendix A, Figure A-4, Table A-14, and Table A-15).

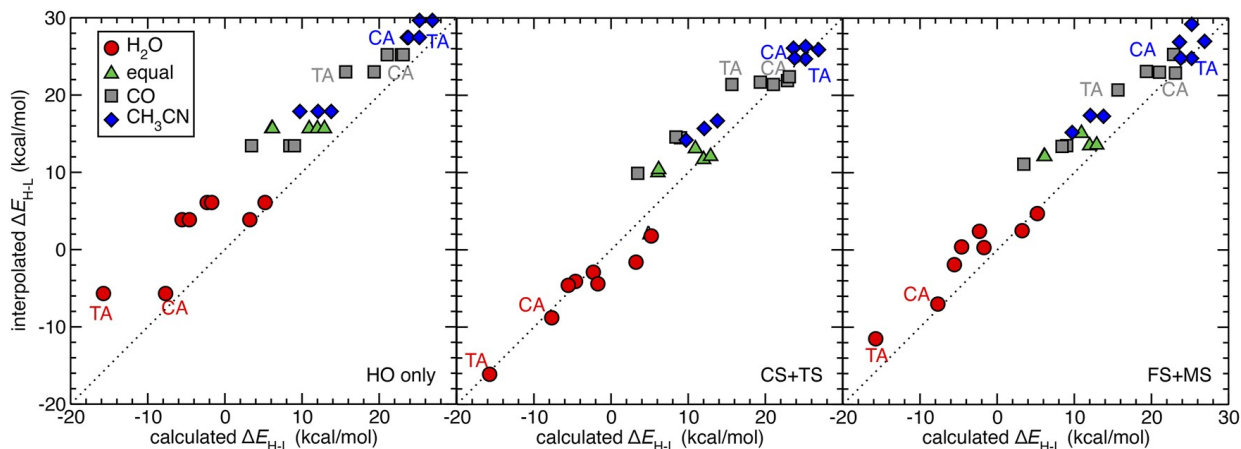


Figure 2-7. Calculated vs linearly interpolated ΔE_{H-L} (kcal/mol) for Fe(II) complexes with at least one each of three ligands: CH₃CN, H₂O, and CO. From left to right: interpolation from homoleptic complexes (HO only), interpolation using homoleptic complexes and CS and TS complex energies derived from pairs of ligands (CS+TS), or interpolation using homoleptic complexes and FS and MS complex energies derived from pairs of ligands (FS+MS). Points are colored according to the ligand with the highest stoichiometric coefficient: H₂O (red circles), CO (gray squares), and CH₃CN (blue diamonds), or equal weight of all ligands (green triangles), as indicated in the inset. Key isomers are annotated. In all panels, a black dotted parity line is shown.

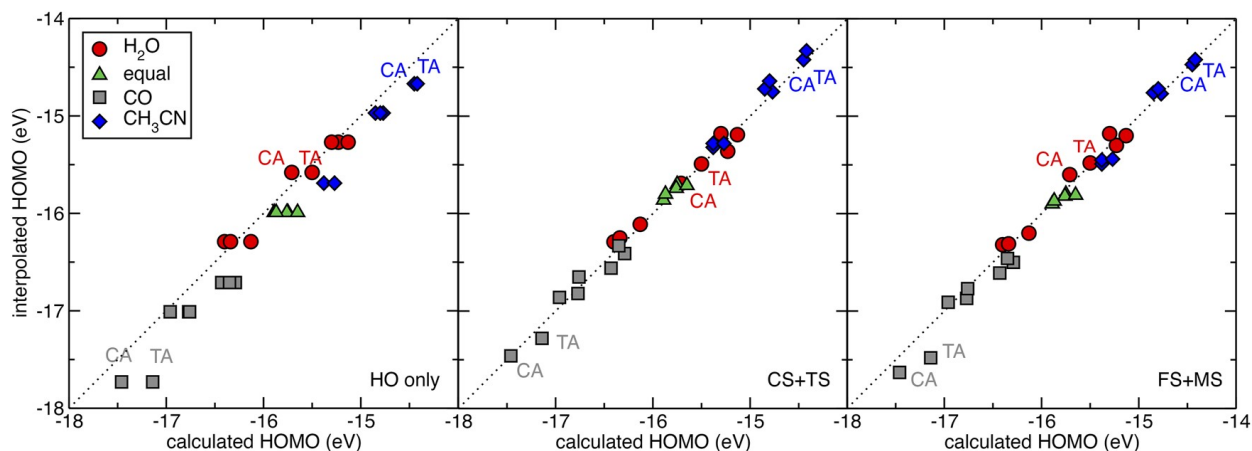


Figure 2-8. Calculated vs linearly interpolated HOMO level (eV) for singlet Fe(II) complexes with at least one each of three ligands: CH₃CN, H₂O, and CO. From left to right: interpolation from homoleptic complexes (HO only), interpolation using homoleptic complexes and CS and TS complex energies derived from pairs of ligands (CS+TS), or interpolation using homoleptic complexes and FS and MS complex energies derived from pairs of ligands (FS+MS). Points are colored according to the ligand with the highest stoichiometric coefficient: H₂O (red circles), CO (gray squares), and CH₃CN (blue diamonds), or equal weight of all ligands (green triangles), as indicated in the inset. Key isomers are annotated. In all panels, a black dotted parity line is shown.

Thus, we next investigated interpolation schemes that reincorporated CS/TS or FS/MS complex energies of the binary heteroleptics. Interpolation of these 29 energies from 12 CS/TS and 3 HO energies or 6 FS/MS and 3 HO energies would still represent significant computational savings. The binary complex energies can also be reused for ternary complexes that share a pair of ligand types, as is the case in the two worked examples presented here. Next, we obtained expressions for all eight symmetry classes of heteroleptics with three ligand types from energies derived from only either FS/MS or

CS/TS complexes. Specifically, the FS/MS-interpolated expressions for these heteroleptics are as follows:

$$E(CA M(L_1)_4(L_2)_1(L_3)_1) = \frac{1}{3}(E(HO M(L_1)_6) + E(FS M(L_1)_3(L_2)_3) + E(FS M(L_1)_3(L_3)_3))$$

(Equation 2-8)

$$E(TA M(L_1)_4(L_2)_1(L_3)_1) = \frac{1}{3}(E(HO M(L_1)_6) + E(MS M(L_1)_3(L_2)_3) + E(MS M(L_1)_3(L_3)_3))$$

(Equation 2-9)

$$E(FA M(L_1)_3(L_2)_2(L_3)_1) = \frac{1}{3}\left(\frac{1}{2}E(HO M(L_1)_6) + \frac{1}{2}E(HO M(L_2)_6) + E(FS M(L_1)_3(L_2)_3) + E(FS M(L_1)_3(L_3)_3)\right)$$

(Equation 2-10)

$$E(MAC M(L_1)_3(L_2)_2(L_3)_1) = \frac{1}{3}\left(\frac{1}{2}E(HO M(L_1)_6) + \frac{1}{2}E(HO M(L_2)_6) + E(FS M(L_1)_3(L_2)_3) + E(MS M(L_1)_3(L_3)_3)\right)$$

(Equation 2-11)

$$E(MAT M(L_1)_3(L_2)_2(L_3)_1) = \frac{1}{3}\left(\frac{1}{2}E(HO M(L_1)_6) + \frac{1}{2}E(HO M(L_2)_6) + E(MS M(L_1)_3(L_2)_3) + E(MS M(L_1)_3(L_3)_3)\right)$$

(Equation 2-12)

$$E(DCS M(L_1)_2(L_2)_2(L_3)_2) = \frac{1}{3}\left(\frac{1}{3}E(HO M(L_1)_6) + \frac{1}{3}E(HO M(L_2)_6) + \frac{1}{3}E(HO M(L_3)_6) + \frac{2}{3}E(FS M(L_1)_3(L_2)_3) + \frac{2}{3}E(FS M(L_1)_3(L_3)_3) + \frac{2}{3}E(FS M(L_2)_3(L_3)_3)\right)$$

(Equation 2-13)

$$\begin{aligned} & E(DCS\ M(L_1)_2(L_2)_2(L_3)_2) \\ &= \frac{1}{3} \left(\frac{1}{3} E(HO\ M(L_1)_6) + \frac{1}{3} E(HO\ M(L_2)_6) + \frac{1}{3} E(HO\ M(L_3)_6) \right. \\ & \quad \left. + \frac{2}{3} E(MS\ M(L_1)_3(L_2)_3) + \frac{2}{3} E(MS\ M(L_1)_3(L_3)_3) + \frac{2}{3} E(MS\ M(L_2)_3(L_3)_3) \right) \end{aligned}$$

(Equation 2-14)

$$\begin{aligned} & E(EA\ M(L_1)_2(L_2)_2(L_3)_2) \\ &= \frac{1}{3} \left(\frac{1}{3} E(HO\ M(L_1)_6) + \frac{1}{3} E(HO\ M(L_2)_6) + \frac{1}{3} E(HO\ M(L_3)_6) \right. \\ & \quad \left. + \frac{2}{3} E(FS\ M(L_1)_3(L_2)_3) + \frac{2}{3} E(MS\ M(L_1)_3(L_3)_3) + \frac{2}{3} E(MS\ M(L_2)_3(L_3)_3) \right) \end{aligned}$$

(Equation 2-15)

where the third ligand in the EA complex is the one that is *trans* to itself, so energies involving that ligand are derived from the binary MS complexes, whereas the remaining components are derived from FS complexes. Analogous expressions were also obtained for the CS/TS energetics (Appendix A, Text A-3). We again obtained these expressions by identifying the most representative interactions (i.e., *cis* or *trans* and *fac* or *mer*) in the final ternary complex and matching the stoichiometry of the resulting complex. Overall, both interpolation schemes significantly improve the estimation of both spin splitting and HOMO level energies for both sets of ternary complexes in comparison to HO-only interpolation (Figure 2-5, Figure 2-6 and Appendix A, Figures A-3 to A-4 and Tables A-14 to A-16). Differences in CA and TA complex properties are much better predicted by both interpolation schemes, with the CS/TS scheme performing best for the combination that includes CH₃CN (Figure 2-5 and Figure 2-6). While this may be expected as a generalization of observations on the binary heteroleptics, it is noteworthy that trends in the fully equivalent stoichiometries (i.e., all three EA isomers, DCS, or DTS complexes) are reasonably well predicted by the improved schemes, whereas they were indistinguishable with simple linear interpolation because they differ solely by *cis* vs *trans* positioning effects (Figure 2-5 and Figure 2-6). Overall, errors for the interpolation scheme using FS/MS ligands are sufficiently low to warrant its use in chemical space exploration (Appendix A, Table A-16). We have thus demonstrated how over a set of three ligands, nine explicit calculations can be used to

interpolate the properties of 18 binary and 29 ternary complexes to around 2 kcal/mol accuracy in spin-splitting energies or 0.1–0.2 eV accuracy in orbital energy levels.

2.3.3 Interpolation of chemical space from experimentally characterized complexes

Given the strength of the interpolative trends we observed, we chose Fe(II) complexes from the CSD to explore the potential of interpolation from previously synthesized complexes. For these 1202 complexes, the majority (661) are homoleptics, followed closely by binary complexes (407) and ternary complexes (129), with only six having more ligand types (Figure 2-4 and Appendix A, Table A-4). We restrict our analysis of chemical space interpolation to monodentate ligands from complexes of each symmetry type that only contain monodentate ligands to simplify the interpolation process because higher-denticity ligands impose geometric constraints that make them incompatible with certain symmetry types. Nevertheless, we note that there are a significant number of monodentate ligands that only appear in combination with higher-denticity ligands (Appendix A, Table A-4 and Table A-5). From the set of monodentate-only complexes, we identified 40 unique ligands already present in homoleptic complexes along with 48 additional unique ligands present in monodentate-only binary and ternary complexes for which we could assign a charge following the scheme introduced in Ref. 60. While the procedure is outlined in detail in Ref. 60, we reiterate that the procedure only assigns ligand charges if there are sufficient copies of that ligand in multiple complexes from which a consistent, single charge can be obtained. In practice, this leads to consistency with the octet rule as well.⁶⁰ Of the set of 88 ligands, only 56 are assigned a neutral charge and closed-shell electronic structure and deemed sterically feasible for homoleptic calculations [i.e., excluding only the neutral OP(Ph)₃ ligand].

We then computed the adiabatic spin-splitting energies and singlet HOMO levels of all 56 homoleptic Fe(II) complexes (see Sec. II). This set is strongly biased toward high-spin structures, with only a few ligands giving rise to low-spin ground states (Figure 2-9 and Appendix A, Figure A-5 and Table A-17). The majority of spin-splitting energies are in the range of –30 to 0 kcal/mol, whereas a minority of complexes are low-spin (Figure 2-9). Only one complex, Fe(II)(CO)₆, which was present in our study in Sec. III B, has a HOMO level deeper than –16 eV, and no complexes have intermediate HOMO levels (ca. –14 eV) while also sampling neardegenerate spin states (Figure 2-9).

While the preference for high spin could be attributed to the strong sensitivity of $\Delta E_{\text{H-L}}$ to the choice of the functional,^{85,86,87,88,89,90,91,92,93,94} reducing the amount of Hartree–Fock exchange (here, to 10% to exaggerate the effect in comparison to 15% recommended in Ref. 88) will not significantly alter the observation that there are regions of the HOMO level and $\Delta E_{\text{H-L}}$ values that homoleptic complexes rarely sample (Appendix A, Figure A-6).

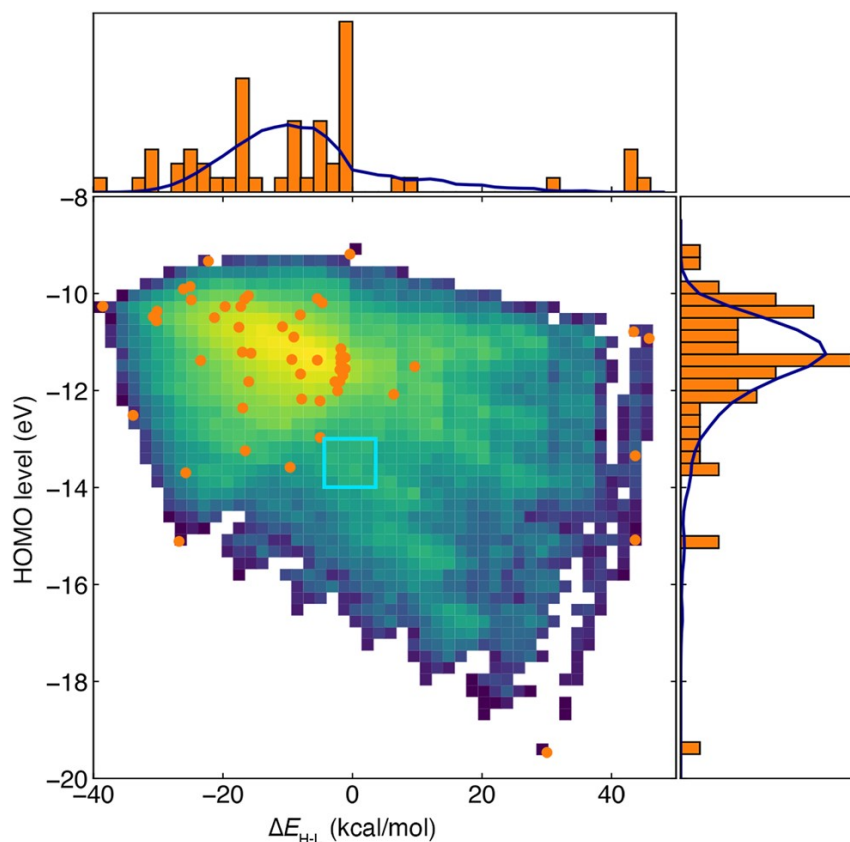


Figure 2-9. The $\Delta E_{\text{H-L}}$ (in kcal/mol) vs singlet HOMO level (in eV) for 56 homoleptic complexes (orange circles) and HO-only interpolation of all possible binary and ternary complexes colored by frequency from purple (low) to yellow (high). 1D histograms of each property are shown at top and right with bin widths of 2 kcal/mol and 0.25 eV, respectively, along with a kernel density estimate of the interpolated space shown as a dark blue line. A targeted zone of -4 to 4 kcal/mol for $\Delta E_{\text{H-L}}$ and -14.0 to -13.0 eV for the HOMO is annotated as a light blue square.

We next interpolated estimates for the 816 200 binary and ternary complexes from the 56 homoleptic complexes using the HO-only averaging scheme (Figure 2-9). By definition, this interpolation scheme provides rough estimates of which ligand

combinations will enrich $\Delta E_{\text{H-L}}$ /HOMO energetic pairings not observed in the homoleptic set (Figure 2-9). Nevertheless, this approach provides only a coarse estimate of the energetics in comparison to interpolation schemes that use knowledge of binary complex energetics. However, even for the most data-efficient approach of FS/MS-derived interpolation, a pool of 56 ligands would require 3080 explicit FS/MS calculations to achieve high-fidelity predictions for 816 200 complexes (see Table 2-1). Thus, we identified a way to use the HO-only interpolation to reduce the number of unique complexes required to achieve high fidelity within a target region.

First, we select a targeted region of $\Delta E_{\text{H-L}}$ values in the range of -4.0 to 4.0 kcal/mol and HOMO levels in the range of -14 to -13 eV. This region was selected because no homoleptic complexes were present in this range, but interpolation of the space predicted that heteroleptic complexes would be found there. These complexes could be of interest in chemical discovery applications that target spin-crossover (SCO) candidates (i.e., with near-degenerate spin states) with good oxidative stability (i.e., deep HOMO levels). We primarily select these two metrics as an illustrative example for simultaneous screening because few SCOs have deep HOMO levels. Next, we identify 12 common ligands from the parent homoleptic complexes that are predicted to give rise most frequently to binary and ternary complexes in the targeted zone on the basis of the simple HO-only model (Figure 2-10 and Appendix A, Table A-18 and Figure A-7). These ligands consist of common ligands from our original set (i.e., CO, H₂O, CH₃CN, and NH₃) but also introduce new chemistry (e.g., S-coordinating dimethylthioformamide, DMTF, and 2-chloropyrazine, ClPyz, Appendix A, Table A-18). From this set, only 132 additional FS/MS complexes need to be studied beyond the 12 homoleptic complexes we already computed to infer the 6776 additional HOMO level or $\Delta E_{\text{H-L}}$ properties at higher fidelity than HO-only averaging (see Table 2-1). We carried out geometry optimizations of these 132 complexes and used their properties to evaluate the revised interpolated HOMO level and $\Delta E_{\text{H-L}}$ values over this subset. Evaluating this subset also provides a validation of the accuracy of the homoleptic averaging over a larger set of ligands. Over this set, we observe that errors are comparable to the earlier tests (i.e., MAE of 4 kcal/mol), and homoleptic averaging is generally a good predictor of $\Delta E_{\text{H-L}}$ (Appendix A, Table A-19 and Figure A-8). For HOMO level predictions, there are more points that differ from the interpolated values, leading to higher MAEs (0.8 eV) than observed over our representative test ligands (Appendix A, Figure A-8 and Table A-19).

Nevertheless, the calculations on FS and MS complexes also highlight the limits of homoleptic averaging for seeking targeted properties. Of the 132 FS or MS

calculations carried out, three are found to be in our targeted zone, MS Fe(II) (MeCN)₃(NH₃)₃ and both FS and MS Fe(II)(CH₃CN)₃(MeOH)₃ (Appendix A, Table A-20). None of these three complexes were predicted to be in the zone from homoleptic averaging alone, with the MS Fe(II)(MeCN)₃(NH₃)₃ $\Delta E_{\text{H-L}}$ underestimated (predicted: -7.3 kcal/mol vs actual -3.6 kcal/mol), while FS/MS Fe(II)(CH₃CN)₃(MeOH)₃ $\Delta E_{\text{H-L}}$ was overestimated (predicted: 9.0 kcal/mol vs 2.9 and -0.8 kcal/mol, respectively, (Appendix A, Table A-20). For these same complexes, the homoleptic averaging predicted HOMO levels very well, within around 0.1 eV (i.e., lower error than for $\Delta E_{\text{H-L}}$, (Appendix A, Table A-20).

Returning to the properties predicted from FS/MS-interpolation, we identified a total of three additional binary complexes predicted to be in the targeted zone from FS/MS-interpolation but not in the targeted zone according to HO-only interpolation and we computed their properties (Fig. 10). These complexes are 5+1 Fe(II)(MeCN)₅(CO) and CS/TS Fe(II) (ClPyz)₄(CO)₂ (Figure 2-11 and Appendix A, Table A-20). Indeed, all three of these complexes have HOMO levels in the targeted zone, while the $\Delta E_{\text{H-L}}$ values are close to [4.9 – 5.7 kcal/mol for CS/TS Fe(II)(ClPyz)₄(CO)₂] or in the targeted zone [1.5 kcal/mol for 5+1 Fe(II)(MeCN)₅(CO), Appendix A, Table A-20]. For CS/TS Fe(II)(ClPyz)₄(CO)₂, homoleptic averaging had underestimated both the $\Delta E_{\text{H-L}}$ and HOMO level (i.e., out of the zone at -15.32 eV) and could not distinguish between CS and TS isomers (Appendix A, Table A-20). Although FS/MS interpolation slightly underestimated the $\Delta E_{\text{H-L}}$ values for CS/TS Fe(II)(ClPyz)₄(CO)₂, the errors are within what we had observed over other sets. In addition to shifting which compounds are predicted to fall within the target zone, the difference between the HO-only to FS/MS interpolation schemes causes some regions of property space to be predicted to be enriched, while others are predicted to be depleted (Figure 2-10).

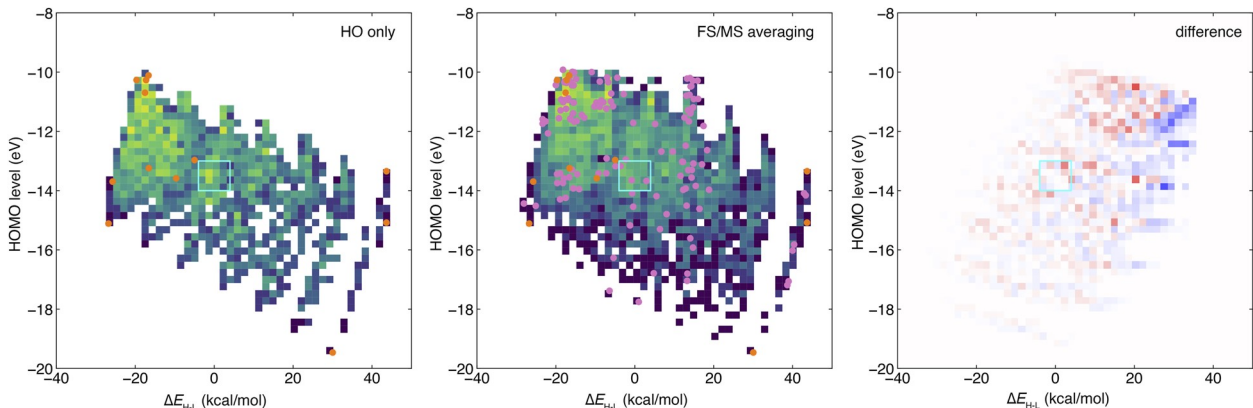


Figure 2-10. Left and middle: The $\Delta E_{\text{H-L}}$ (in kcal/mol) vs singlet HOMO level (in eV) for 12 homoleptic complexes (orange circles) and HO-only (left) or FS/MS-based (middle) interpolation of all possible binary and ternary complexes colored by frequency from purple (low) to yellow (high). The FS/MS complex energies are shown as pink circles. Right: The difference (i.e., HO-only minus FS/MS-based interpolation) of the two 2D histograms plotted from negative (blue, -81) to positive (red, 81). A targeted zone of -4 to 4 kcal/mol for $\Delta E_{\text{H-L}}$ and -14.0 to -13.0 eV for the HOMO is annotated as a light blue square.

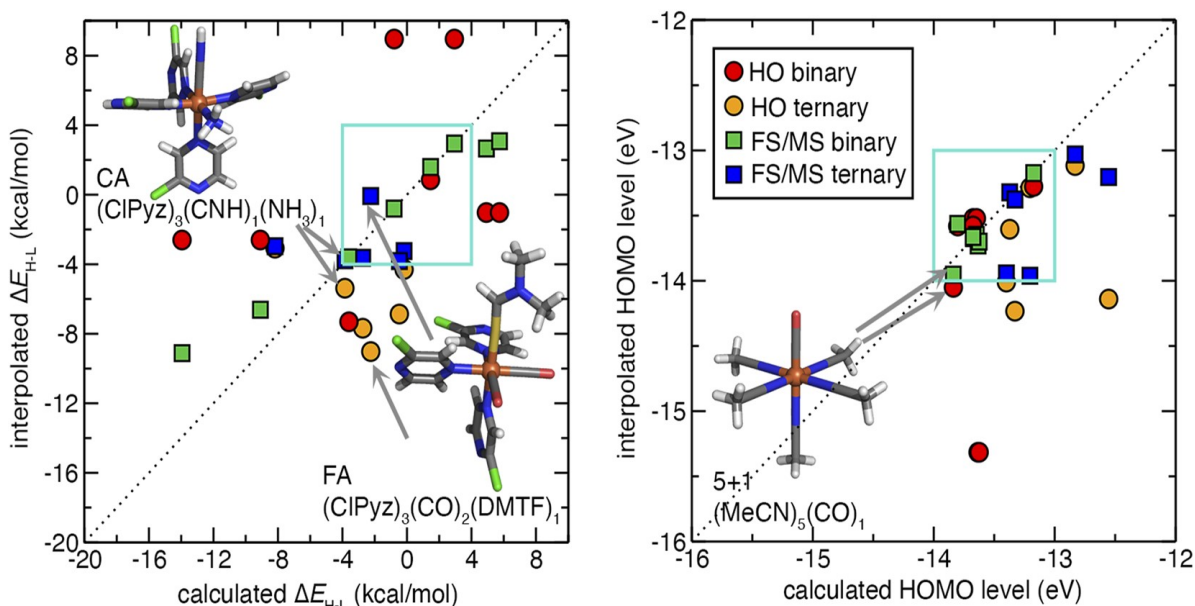


Figure 2-11. Properties of eight binary (red or green) and six ternary (orange or blue) complexes in the validation set for HO-interpolation (circles) and FS/MS-augmented interpolation (squares) for $\Delta E_{\text{H-L}}$ (in kcal/mol, left) and the HOMO level (in eV, right). The targeted zone for each quantity is shown as a turquoise square, and three representative complexes are shown in the inset with their symmetry class, and the associated points are indicated with gray arrows. Structures are colored as follows: brown for Fe, gray for C, blue for N, white for H, green for Cl, red for O, and yellow for S. A dotted parity line is also shown.

As a control, we also identified two complexes predicted to be in the zone by the HO-only interpolation but out of zone by FS/MS-interpolation, CS/Ts Fe(II)(MeOH)₄(CH₃CN)₂ (Appendix A, Table A-20). The homoleptic averaging predicts the $\Delta E_{\text{H-L}}$ value for both complexes to be -2.6 kcal/mol, while the calculated $\Delta E_{\text{H-L}}$ values are significantly lower (-9.1 and -13.9 kcal/mol). The FS/MS interpolation captures well the relative CS/Ts energetics in both this case and the in-zone example

and only slightly overestimates the $\Delta E_{\text{H-L}}$ value (Appendix A, Table A-20). Overall, FS/MS-interpolation MAEs for these binary complexes are low (2.5 kcal/mol for $\Delta E_{\text{H-L}}$ and 0.1 eV for HOMO levels) and less than half that observed from HO-only averaging (Appendix A, Table A-20).

As a more stringent test of our FS/MS interpolation scheme, we also selected six representative ternary complexes predicted to be in the targeted zone from FS/MS interpolation but out of the targeted zone when estimated with homoleptic interpolation (Figure 2-11 and Appendix A, Table A-21). All complexes were generally predicted to be more high-spin favoring by homoleptic averaging than predicted from FS/MS interpolation, and several were also predicted to have deeper HOMO levels than the targeted region (Appendix A, Table A-21). Over this set, explicit DFT calculations show that five of the complexes have $\Delta E_{\text{H-L}}$ values in the targeted zone and four have HOMO levels in the targeted zone (Appendix A, Table A-21). These include FA Fe(II)(ClPyz)₃(CO)₂(DMTF), which was predicted to be strongly HS by homoleptic averaging ($\Delta E_{\text{H-L}} = -9.0$ kcal/mol) but was much closer to the FS/MS-interpolated value (calculated $\Delta E_{\text{H-L}} = -2.2$ kcal/mol vs FS/MS $\Delta E_{\text{H-L}} = -0.1$ kcal/mol, Appendix A, Table A-21). For the worst-performing example, FA Fe(II)(DMTF)₃(CO)₂(H₂O), FS/MS interpolation overestimates $\Delta E_{\text{H-L}}$ by 5 kcal/mol (-3.0 kcal/mol vs -8.2 kcal/mol) and predicts a deeper HOMO level (-13.20 eV vs -12.55 eV, Appendix A, Table A-21). This could be due to the weak coordination of the metal by water, which leads to a more stabilized high-spin state. Performance of FS/MS interpolation on the remaining CA and TA complexes ranges from good ($\sim 2-3$ kcal/mol errors) to exceptional in the case of CA Fe(II)(ClPyz)₄(CNH)(NH₃) where errors on $\Delta E_{\text{H-L}}$ are below 0.1 kcal/mol and the HOMO level are around 0.05 eV (Appendix A, Table A-21). Overall errors are low from FS/MS interpolation at around 2.5 kcal/mol for $\Delta E_{\text{H-L}}$ and 0.4 eV for the HOMO level, roughly half their values from homoleptic averaging. These results demonstrate that coarse interpolation of large spaces with homoleptic complexes can be followed up by improved interpolation using selected FS/MS compounds to identify the most promising binary or ternary complexes for explicit calculation. Overall, the demonstrated approach represents a data-efficient strategy to infer properties across large compound spaces with systematically improvable fidelity.

2.4 Conclusions

A large-scale analysis of the mononuclear octahedral complexes deposited in the CSD revealed a propensity toward specific, higher symmetry classes. In addition, few

complexes contained more than three unique ligand types. To assess the relative diversity of these complexes compared to the enumerated chemical space, we obtained expressions for the theoretical number of complexes of the five binary and eight ternary symmetry classes for octahedral complexes. We showed that even for a relatively small number of neutral, monodentate ligands present in Fe(II) complexes, the total theoretical space of 816 200 binary and ternary complexes far exceeded those that had been characterized in the CSD.

An aim of identifying which uncharacterized compounds are most likely to be valuable or informative motivated our evaluation of interpolative schemes to determine the extent to which heteroleptic complex properties could be inferred from parent homoleptic complexes. Over representative test cases, we observed that a linear weighted averaging of homoleptic properties could reasonably (to ~ 4 kcal/mol for $\Delta E_{\text{H-L}}$ and 0.24 eV for the HOMO level) predict properties of binary and ternary heteroleptic complexes. We demonstrated a refinement of the approach to be able to distinguish isomers (e.g., CS vs TS or CA vs TA) by using expressions that also incorporated either CS/TS or FS/MS binary complexes at a slightly higher computational cost but with errors that were half as large (~ 2 kcal/mol for $\Delta E_{\text{H-L}}$ and 0.15 eV for the HOMO level). The most data-efficient approach required four FS/MS and three homoleptic energies (i.e., 7 total) to infer 18 binary and 29 ternary complex properties.

Finally, we demonstrated a two-stage discovery approach to leverage and validate our interpolative schemes. We first used 56 homoleptic Fe(II) complexes composed of neutral, closed-shell monodentate ligands to infer the properties of 816 200 binary or ternary complexes of these ligands using HO-only averaging. We then defined a targeted zone of the HOMO level and $\Delta E_{\text{H-L}}$ that contained none of the homoleptic complexes. To avoid explicit calculations for all (~ 3000) FS/MS complexes needed to achieve high fidelity over the full range of ligands, we then refined our analysis to the top 12 most frequently occurring ligands predicted to be in the targeted zone. From this set, we studied 66 each of FS and MS complexes to refine our interpolation of 6776 complexes. This approach helped us to identify 3 FS/MS complexes in the targeted zone that had not been predicted by homoleptic averaging alone. It also had a higher validation rate for binary and ternary complexes than homoleptic averaging, with all FS/MS-interpolation predicted complexes residing in the targeted zone or just outside it. Overall, errors for $\Delta E_{\text{H-L}}$ of around 5 kcal/mol with homoleptic averaging and 2 kcal/mol with FS/MS-interpolation are also comparable to prior machine learning (i.e., artificial neural network) model predictions on similar datasets.⁵¹ Thus, this approach represents a promising multi-stage strategy for efficient chemical space exploration at

low cost: an initial coarse interpolation from homoleptic complexes can be systematically refined by incorporating *cis* and *trans* isomer effects over a smaller subspace of ligands. While demonstrated here for magnetic and orbital energy properties, this approach is expected to have similar applicability in predicting other properties where ligands can be expected to behave in an approximately additive manner, such as in redox potentials or catalysis. This observed additivity could also be integrated into machine learning model property predictions or used synergistically to augment datasets for machine learning.

2.5 References

- ¹ Y. N. Shu and B. G. Levine, *J. Chem. Phys.* **142**(10), 104104 (2015).
- ² R. Gomez-Bombarelli, J. Aguilera-Iparraguirre, T. D. Hirzel, D. Duvenaud, D. Maclaurin, M. A. Blood-Forsythe, H. S. Chae, M. Einzinger, D. G. Ha, T. Wu, G. Markopoulos, S. Jeon, H. Kang, H. Miyazaki, M. Numata, S. Kim, W. L. Huang, S. I. Hong, M. Baldo, R. P. Adams, and A. Aspuru-Guzik, *Nat. Mater.* **15**(10), 1120+ (2016).
- ³ I. Y. Kanal, S. G. Owens, J. S. Bechtel, and G. R. Hutchison, *J. Phys. Chem. Lett.* **4**(10), 1613–1623 (2013).
- ⁴ K. D. Vogiatzis, M. V. Polynski, J. K. Kirkland, J. Townsend, A. Hashemi, C. Liu, and E. A. Pidko, *Chem. Rev.* **119**, 2453–2523 (2018).
- ⁵ M. Foscatto and V. R. Jensen, *ACS Catal.* **10**(3), 2354–2377 (2020).
- ⁶ S. Curtarolo, G. L. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, *Nat. Mater.* **12**(3), 191–201 (2013).
- ⁷ S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, *Comput. Mater. Sci.* **68**, 314–319 (2013).
- ⁸ J. K. Nørskov and T. Bligaard, *Angew. Chem., Int. Ed.* **52**(3), 776–777 (2013).
- ⁹ J. P. Janet, C. Duan, A. Nandy, F. Liu, and H. J. Kulik, *Acc. Chem. Res.* **54**(3), 532–545 (2021).
- ¹⁰ M. Ceriotti, C. Clementi, and O. A. von Lilienfeld, *Chem. Rev.* **121**(16), 9719–9721 (2021).
- ¹¹ J. A. Keith, V. Vassilev-Galindo, B. Q. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller, and A. Tkatchenko, *Chem. Rev.* **121**(16), 9816–9872 (2021).
- ¹² M. Ceriotti, *J. Chem. Phys.* **150**(15), 150901 (2019).

- ¹³ A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf, and M. Ceriotti, *ACS Cent. Sci.* **5**, 57–64 (2018).
- ¹⁴ K. Jorner, A. Tomberg, C. Bauer, C. Sköld, and P.-O. Norrby, *Nat. Rev. Chem.* **5**(4), 240–255 (2021).
- ¹⁵ M. F. Kasim and S. M. Vinko, *Phys. Rev. Lett.* **127**(12), 126403 (2021).
- ¹⁶ J. Westermayr, M. Gastegger, K. T. Schütt, and R. J. Maurer, *J. Chem. Phys.* **154**(23), 230903 (2021).
- ¹⁷ J. P. Coe, *J. Chem. Theory Comput.* **14**(11), 5739–5749 (2018).
- ¹⁸ A. S. Rosen, S. M. Iyer, D. Ray, Z. P. Yao, A. Aspuru-Guzik, L. Gagliardi, J. M. Notestein, and R. Q. Snurr, *Matter* **4**(5), 1578–1597 (2021).
- ¹⁹ T. Dimitrov, C. Kreisbeck, J. S. Becker, A. Aspuru-Guzik, and S. K. Saikin, *ACS Appl. Mater. Interfaces* **11**(28), 24825–24836 (2019).
- ²⁰ A. Glielmo, B. E. Husic, A. Rodriguez, C. Clementi, F. Noé, and A. Laio, *Chem. Rev.* **121**(16), 9722–9758 (2021).
- ²¹ L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. H. Hu, A. Palizhati, A. Sriram, B. Wood, J. Yoon, D. Parikh, C. L. Zitnick, and Z. Ulissi, *ACS Catal.* **11**(10), 6059–6072 (2021).
- ²² M. Christensen, L. P. E. Yunker, F. Adedeji, F. Häse, L. M. Roch, T. Gensch, G. D. Gomes, T. Zepel, M. S. Sigman, A. Aspuru-Guzik, and J. E. Hein, *Commun. Chem.* **4**(1), 112 (2021).
- ²³ T. Gensch, G. D. Gomes, P. Friederich, E. Peters, T. Gaudin, R. Pollice, K. Jorner, A. Nigam, M. Lindner-D'Addario, M. S. Sigman, and A. Aspuru-Guzik, *J. Am. Chem. Soc.* **144**(3), 1205–1217 (2022).
- ²⁴ A. M. Virshup, J. Contreras-García, P. Wipf, W. Yang, and D. N. Beratan, *J. Am. Chem. Soc.* **135**(19), 7296–7303 (2013).
- ²⁵ L. Ruddigkeit, R. van Deursen, L. C. Blum, and J.-L. Reymond, *J. Chem. Inf. Model.* **52**(11), 2864–2875 (2012).
- ²⁶ R. W. Sterner and J. J. Elser, *Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere* (Princeton University Press, 2002).
- ²⁷ H. J. M. Bowen, *Environmental Chemistry of the Elements* (Academic Press, 1979).
- ²⁸ T. Fink, H. Bruggesser, and J.-L. Reymond, *Angew. Chem., Int. Ed.* **44**(10), 1504–1508 (2005).

- ²⁹ M. J. Wester, S. N. Pollock, E. A. Coutsiyas, T. K. Allu, S. Muresan, and T. I. Oprea, *J. Chem. Inf. Model.* **48**(7), 1311–1324 (2008).
- ³⁰ P. G. Polishchuk, T. I. Madzhidov, and A. Varnek, *J. Comput.-Aided Mol. Des.* **27**(8), 675–679 (2013).
- ³¹ R. S. Bohacek, C. McMartin, and W. C. Guida, *Med. Res. Rev.* **16**(1), 3–50 (1996).
- ³² N. Fey, S. E. Harris, J. N. Harvey, and A. G. Orpen, *J. Chem. Inf. Model.* **46**(2), 912–929 (2006).
- ³³ N. Fey, A. C. Tsipis, S. E. Harris, J. N. Harvey, A. G. Orpen, and R. A. Mansson, *Chem. - Eur. J.* **12**(1), 291–302 (2006).
- ³⁴ D. J. Durand and N. Fey, *Chem. Rev.* **119**(11), 6561–6594 (2019).
- ³⁵ M. Foscatto, G. Occhipinti, V. Venkatraman, B. K. Alsberg, and V. R. Jensen, *J. Chem. Inf. Model.* **54**(3), 767–780 (2014).
- ³⁶ M. Foscatto, V. Venkatraman, G. Occhipinti, B. K. Alsberg, and V. R. Jensen, *J. Chem. Inf. Model.* **54**(7), 1919–1931 (2014).
- ³⁷ M. Foscatto, R. J. Deeth, and V. R. Jensen, *J. Chem. Inf. Model.* **55**(6), 1282–1290 (2015).
- ³⁸ M. Foscatto, B. J. Houghton, G. Occhipinti, R. J. Deeth, and V. R. Jensen, *J. Chem. Inf. Model.* **55**(9), 1844–1856 (2015).
- ³⁹ D. Balcells and B. B. Skjelstad, *J. Chem. Inf. Model.* **60**(12), 6135–6146 (2020).
- ⁴⁰ J. S. Smith, O. Isayev, and A. E. Roitberg, *Sci. Data* **4**, 170193 (2017).
- ⁴¹ R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. Von Lilienfeld, *Sci. Data* **1**, 140022 (2014).
- ⁴² C. A. Gaggioli, S. J. Stoneburner, C. J. Cramer, and L. Gagliardi, *ACS Catal.* **9**(9), 8481–8502 (2019).
- ⁴³ W. Jiang, N. J. DeYonker, and A. K. Wilson, *J. Chem. Theory Comput.* **8**(2), 460–468 (2012).
- ⁴⁴ F. Liu, C. Duan, and H. J. Kulik, *J. Phys. Chem. Lett.* **11**, 8067–8076 (2020).
- ⁴⁵ C. Duan, D. B. K. Chu, A. Nandy, and H. J. Kulik, *Chem. Sci.* **13**, 4962–4971 (2022).
- ⁴⁶ M. Feldt, Q. M. Phung, K. Pierloot, R. A. Mata, and J. N. Harvey, *J. Chem. Theory Comput.* **15**(2), 922–937 (2019).
- ⁴⁷ Q. M. Phung, M. Feldt, J. N. Harvey, and K. Pierloot, *J. Chem. Theory Comput.* **14**(5), 2446–2455 (2018).

- ⁴⁸ S. Gugler, J. P. Janet, and H. J. Kulik, *Mol. Syst. Des. Eng.* **5**, 139–152 (2020).
- ⁴⁹ A. Gaulton, A. Hersey, M. Nowotka, A. P. Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L. J. Bellis, E. Cibrián-Uhalte, M. Davies, N. Dedman, A. Karlsson, M. P. Magariños, J. P. Overington, G. Papadatos, I. Smit, and A. R. Leach, *Nucleic Acids Res.* **45**(D1), D945–D954 (2017).
- ⁵⁰ P. F. Bernath and S. McLeod, *J. Mol. Spectrosc.* **207**(2), 287 (2001).
- ⁵¹ J. P. Janet, C. Duan, T. Yang, A. Nandy, and H. J. Kulik, *Chem. Sci.* **10**, 7913–7922 (2019).
- ⁵² C. R. Groom, I. J. Bruno, M. P. Lightfoot, and S. C. Ward, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **72**(2), 171–179 (2016).
- ⁵³ J. Glerup, O. Moensted, and C. E. Schaeffer, *Inorg. Chem.* **15**(6), 1399–1407 (1976).
- ⁵⁴ R. J. Deeth, D. L. Foulis, and B. J. Williams-Hubbard, *Dalton Trans.*(20), 3949–3955 (2003).
- ⁵⁵ Y. Cytter, A. Nandy, A. Bajaj, and H. J. Kulik, *J. Phys. Chem. Lett.* **13**, 4549–4555 (2022).
- ⁵⁶ R. A. Friesner, E. H. Knoll, and Y. Cao, *J. Chem. Phys.* **125**(12), 124107 (2006).
- ⁵⁷ D. Rinaldo, L. Tian, J. N. Harvey, and R. A. Friesner, *J. Chem. Phys.* **129**(16), 164108 (2008).
- ⁵⁸ T. F. Hughes and R. A. Friesner, *J. Chem. Theory Comput.* **7**(1), 19–32 (2011).
- ⁵⁹ R. A. Friesner and S. V. Jerome, *Coord. Chem. Rev.* **344**, 205–213 (2017).
- ⁶⁰ C. Duan, A. J. Ladera, J. C.-L. Liu, M. G. Taylor, I. R. Ariyaratna, and H. J. Kulik, *J. Chem. Theory Comput.* **18**, 4836–4845 (2022).
- ⁶¹ K. U. Lao and J. M. Herbert, *J. Phys. Chem. A* **119**(2), 235–252 (2015).
- ⁶² M. S. Gordon, L. Slipchenko, H. Li, and J. H. Jensen, *Annu. Rep. Comput. Chem.* **3**, 177–193 (2007).
- ⁶³ F. Neese, A. Hansen, and D. G. Liakos, *J. Chem. Phys.* **131**(6), 064103 (2009).
- ⁶⁴ Q. Ma and H. J. Werner, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **8**(6), e1371 (2018).
- ⁶⁵ M. G. Taylor, T. Yang, S. Lin, A. Nandy, J. P. Janet, C. Duan, and H. J. Kulik, *J. Phys. Chem. A* **124**(16), 3286–3299 (2020).
- ⁶⁶ S. Vela, R. Laplaza, Y. Cho, and C. Corminboeuf, *npj Comput. Mater.* **8**(1), 188 (2022).

- ⁶⁷ S. Seritan, C. Bannwarth, B. S. Fales, E. G. Hohenstein, S. I. Kokkila-Schumacher, N. Luehr, J. W. Snyder, Jr., C. Song, A. V. Titov, and I. S. Ufimtsev, *J. Chem. Phys.* **152**(22), 224110 (2020).
- ⁶⁸ S. Seritan, C. Bannwarth, B. S. Fales, E. G. Hohenstein, C. M. Isborn, S. I. Kokkila-Schumacher, X. Li, F. Liu, N. Luehr, and J. W. Snyder, Jr., *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **11**(2), e1523 (2021).
- ⁶⁹ A. D. Becke, *J. Chem. Phys.* **98**(7), 5648–5652 (1993).
- ⁷⁰ C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785–789 (1988).
- ⁷¹ P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, *J. Phys. Chem.* **98**(45), 11623–11627 (1994).
- ⁷² P. J. Hay and W. R. Wadt, *J. Chem. Phys.* **82**(1), 270–283 (1985).
- ⁷³ R. Ditchfield, W. J. Hehre, and J. A. Pople, *J. Chem. Phys.* **54**(2), 724 (1971).
- ⁷⁴ V. R. Saunders and I. H. Hillier, *Int. J. Quantum Chem.* **7**(4), 699–705 (1973).
- ⁷⁵ L.-P. Wang and C. Song, *J. Chem. Phys.* **144**(21), 214108 (2016).
- ⁷⁶ E. I. Ioannidis, T. Z. H. Gani, and H. J. Kulik, *J. Comput. Chem.* **37**, 2106–2117 (2016).
- ⁷⁷ See <https://github.com/hjkgrp/molsimplify> for the KulikGroup.
- ⁷⁸ N. M. O’Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, *J. Cheminf.* **3**, 33 (2011).
- ⁷⁹ N. M. O’Boyle, C. Morley, and G. R. Hutchison, *Chem. Cent. J.* **2**, 5 (2008).
- ⁸⁰ N. Arunachalam, S. Gugler, M. G. Taylor, C. Duan, A. Nandy, J. P. Janet, R. Meyer, J. Oldenstaedt, D. B. K. Chu and H. J. Kulik (2022). “Data set for Ligand additivity relationships enable efficient exploration of transition metal chemical space,” Zenodo. <https://doi.org/10.5281/zenodo.7224793>
- ⁸¹ G. Pólya and R. C. Read, *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds* (Springer, New York, NY, 1987).
- ⁸² I. Baraldi, C. Fiori and D. Vanossi, *J. Math. Chem.* **25**, 23–30 (1999).
- ⁸³ B. A. Kennedy, D. A. McQuarrie, and C. H. Brubaker, *Inorg. Chem.* **3**(2), 265–268 (1964).
- ⁸⁴ A. Nandy, D. B. K. Chu, D. R. Harper, C. Duan, N. Arunachalam, Y. Cytter, and H. J. Kulik, *Phys. Chem. Chem. Phys.* **22**, 19326–19341 (2020).
- ⁸⁵ M. Reiher, *Inorg. Chem.* **41**(25), 6928–6935 (2002).

- ⁸⁶ G. Ganzenmüller, N. Berkäine, A. Fouqueau, M. E. Casida, and M. Reiher, *J. Chem. Phys.* **122**(23), 234321 (2005).
- ⁸⁷ M. Reiher, O. Salomon, and B. A. Hess, *Theor. Chem. Acc.* **107**(1), 48–55 (2001).
- ⁸⁸ O. Salomon, M. Reiher, and B. A. Hess, *J. Chem. Phys.* **117**(10), 4729–4737 (2002).
- ⁸⁹ E. I. Ioannidis and H. J. Kulik, *J. Chem. Phys.* **143**(3), 034104 (2015).
- ⁹⁰ D. N. Bowman and E. Jakubikova, *Inorg. Chem.* **51**(11), 6011–6019 (2012).
- ⁹¹ A. Droghetti, D. Alfè, and S. Sanvito, *J. Chem. Phys.* **137**(12), 124303 (2012).
- ⁹² D. M. Smith, M. Dupuis, and T. Straatsma, *Mol. Phys.* **103**(2–3), 273–278 (2005).
- ⁹³ P. Verma, Z. Varga, J. E. Klein, C. J. Cramer, L. Que, and D. G. Truhlar, *Phys. Chem. Chem. Phys.* **19**(20), 13049–13069 (2017).
- ⁹⁴ K. P. Kepp, *Inorg. Chem.* **55**(6), 2717–2727 (2016).

3 Isovalent transition metal complex structures demonstrate additional trends in chemical space

Aspects of this work related to 3d/4d transition metal complexes have previously appeared in D.R. Harper, A. Nandy, N. Arunachalam, C. Duan, J.P. Janet, and H.J. Kulik, Representations and strategies for transferable machine learning improve model performance in chemical discovery, *The Journal of Chemical Physics* **156**, 074101 (2022) and A. Nandy, D.B. Chu, D.R. Harper, C. Duan, N. Arunachalam, Y. Cytter, and H.J. Kulik, Large-scale comparison of 3d and 4d transition metal complexes illuminates the reduced effect of exchange on second-row spin-state energetics, *Physical Chemistry Chemical Physics* **22**, 19326 (2020), and are cited accordingly.

3.1 Introduction

In 2019, our group was interested in exploring second row transition metals because their database coverage was sparse compared to 1st row transition metals. A systematic exploration of complexes containing 2nd row metals could be interesting because it can highlight new ways to think about molecular representations (e.g. for machine learning), and bring new types of chemistry into light.

As a first step towards studying 2nd row transition metal complexes, we looked at low, intermediate, and high-spin states for Mo, Tc, Ru, and Rh centers with oxidation states II and III. We chose these because they were one period below the existing metals that mAD previously had full support for, i.e. Cr, Mn, Fe, and Co. Analogously, we were interested in the effect of 2p vs 3p ligands, and performed an analysis of property changes between complexes with 3d or 4d metals combined with 2p or 3p ligands.

3.2 Methods

3.2.1 Data set construction – 3d/4d

We studied the effect of HF exchange on the spin-state ordering of octahedral TMCs with a single mid-row transition-metal center. We compared properties of TMCs comprised of first-row (i.e., 3d valence) Cr, Mn, Fe, and Co to second-row (i.e., 4d valence) Mo, Tc, Ru, and Rh (Figure 3-1). In all cases, we calculated properties of metal centers in formal M(II) or M(III) oxidation states to ensure differences in spin state correspond to differences in d orbital occupations (Figure 3-1). The TMCs were evaluated in up to three spin states: low-spin (LS), intermediate-spin (IS), and high-spin (HS), where we defined the IS and HS states as those that differ from the LS state (i.e., with ls unpaired electrons) by two more (i.e., $ls+2$) or four more (i.e., $ls+4$) unpaired electrons, respectively. We then computed the gas phase, adiabatic spin-splitting energy: between the LS and HS states, ΔE_{H-L} , as well as between the IS state and either the LS or the HS state (i.e., ΔE_{H-I} and ΔE_{I-L}). The nominally d^3 Cr(III)/Mo(III) and d^7 Co(II)/Rh(II) were evaluated only in LS doublet and IS quartet spin multiplicities (i.e., also the highest accessible spin state), and only ΔE_{I-L} was computed (Figure 3-1). The d^5 metals (i.e., Mn(II)/Tc(II) or Fe(III)/Ru(III)) were studied in LS doublet, IS quartet, and HS sextet states (Figure 1). Analogously, d^4 (i.e., Mn(III)/Tc(III) or Cr(II)/Mo(II)) and d^6 (i.e., Fe(II)/Ru(II) or Co(III)/Rh(III)) metals were calculated in LS singlet, IS triplet, and HS quintet states (Figure 3-1).

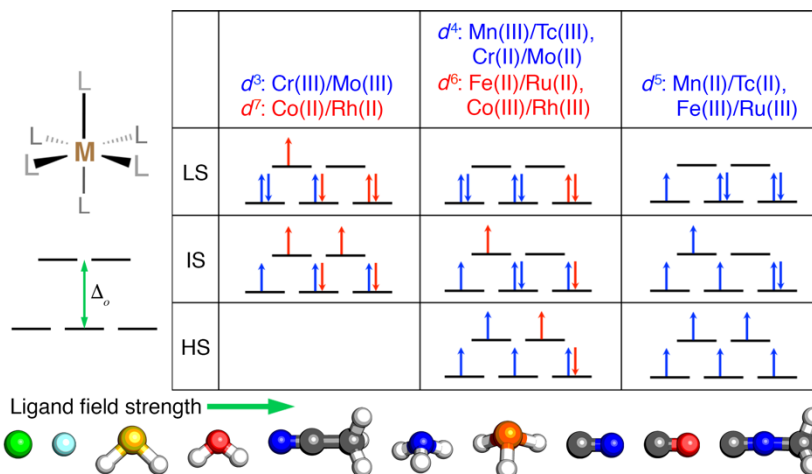


Figure 3-1. (top) Qualitative diagrams of electron configurations in low-spin (LS), intermediate-spin (IS), and high-spin (HS) states for the mononuclear octahedral transition-metal complexes studied in this work (schematically shown at left). For both

d^3/d^7 and d^4/d^6 M(II) or M(III) complexes, the additional electrons for the later transition metal are shown in red, and the electrons that apply to both states are shown in blue. The d^3 or d^7 complexes do not have a defined HS state. (bottom) The ten main monodentate ligands studied in this work ordered by their increasing ligand field strength, which tunes the octahedral field splitting (schematically shown at left). Atoms in the ball-and-stick representation are colored as follows: H in white, C in gray, N in blue, O in red, F in light blue, P in orange, S in yellow, and Cl in green.

We calculated properties of complexes formed from combinations of ten small, monodentate ligands that spanned ligand field strengths and coordinating element identities. Negatively charged halides (Cl⁻ and F⁻) ions are known¹ to have among the weakest field strength, while several others (i.e., phosphine, carbonyl, and cyanide) have among the highest field strengths (Figure 3-1). Intermediate behavior is expected of the remaining (i.e., water, ammonia, hydrogen sulfide, acetonitrile, and methyl isocyanide) ligands (Figure 3-1). In addition to homoleptic complexes, heteroleptic complexes were formed from up to two ligands (i.e., L₁ and L₂). Both M(L₁)₄(L₂)₂ TMCs with the two minority L₂ ligands either trans (i.e., aligned 180° in the TMC) or cis (i.e., 90° in the TMC) were studied along with M(L₁)₅(L₂) TMCs.

3.2.2 Electronic structure calculations – 3d/4d

Calculations on mononuclear, octahedral TMCs studied in this work followed an established protocol.^{2,3} All initial structures were generated using molSimplify^{4,5,6}, which employs OpenBabel^{7,8} as a backend for ligand structure generation. These calculations were automated and checked for fidelity with molSimplify automatic design (mAD)^{2,9}. All TMCs were geometry optimized with DFT using a development version of TeraChem.^{10,11} For the geometry optimizations, the standard B3LYP^{12,13,14} global hybrid functional was employed along with modified forms in which the Hartree–Fock exchange fraction (a_{HF}) was varied from its default value of 0.20 to as low as $a_{\text{HF}} = 0.00$ (i.e., a pure BLYP GGA) or as high as $a_{\text{HF}} = 0.30$ in increments of 0.05 while holding the LDA/GGA exchange ratio fixed^{15,16}. As in 3d TMCs^{15,17,18}, sensitivities from a_{HF} variations using PBE as the GGA xc are comparable on a representative 4d TMC (Appendix B, Table B-1). All calculations employed the LANL2DZ¹⁹ effective core potential for transition metals and the 6-31G* basis for all other atoms. Only singlet calculations were carried out in a spin-restricted formalism, with all other spin multiplicities carried out unrestricted. Level shifting²⁰ was employed to aid self-

consistent field (SCF) convergence with the majority spin and minority spin virtual orbitals each shifted by 0.25 Ha. The default SCF convergence threshold of 3×10^{-5} for the largest component of the DIIS vector was employed. Geometry optimizations were carried out in translation rotation internal coordinates²¹ using the L-BFGS algorithm to default tolerances for the gradient of 4.5×10^{-4} hartree/bohr and energy difference between steps of 10^{-6} hartree. Geometric properties from a representative case optimized with a larger triple- ζ basis set and the resulting exchange sensitivities preserves trends Appendix B, Table B-1).

For the mAD calculation workflow, calculations were run for 24-hour increments and resubmitted for up to five additional runs. At each resubmission, mAD applies loose geometric criteria² and abandons any calculations that fail these checks (Appendix B, Table B-2). In this workflow, the B3LYP ($a_{\text{HF}} = 0.20$) geometry optimization was carried out first. If the B3LYP geometry optimization converged, we used the converged structure and wavefunction to initialize geometry at the adjacent increased (i.e., 0.25) or decreased (i.e., 0.15) a_{HF} values, as in prior work²². If these calculations converged, their structures and wavefunctions were then used for the next adjacent (e.g., increased to 0.30 or decreased to 0.10) a_{HF} value geometry optimizations. However, if an optimization failed to converge, the next a_{HF} value was not attempted.

For all converged calculations, automated data fidelity checks were employed based on refinements of prior geometric² and electronic structure criteria^{141-142, 155} (Appendix B, Table B-2). Specifically, complexes were retained if their structure was deemed to be intact based on tighter geometric criteria than were employed during the optimization (Appendix B, Table B-3). The electronic structure criteria required that the deviation of the \hat{S}^2 expectation value from its anticipated value (i.e., $S(S+1)$) was below 1 and the Mulliken metal atomic spin density was within $1 \mu_{\text{B}}$ of the total spin of the molecule (Appendix B, Table B-3).

Linearized exchange sensitivities, S , were obtained from linear fits of the dependence of the relevant property (e.g., $\Delta E_{\text{H-L}}$) on a_{HF} . The resulting sensitivity (e.g., $S(\Delta E_{\text{H-L}})$) is reported as the change in property over the range from $a_{\text{HF}} = 0.0$ to 1.0, which we refer to as HFX as in prior work^{15,22,23}. As long as a single, qualitatively consistent electronic state has been converged over all points, this linear approximation is known to be good for a range of properties, both energetic (e.g., spin splitting^{15,22,23,24,25,26} and reaction energies^{15,27}) and electronic^{28,29,30} in nature. To ensure that linearized exchange sensitivities could be quantitatively obtained from collected data points, we applied a series of constraints and filtering steps; this procedure and its

results are detailed in Appendix B, Table B-4. R^2 values, computed sensitivities, and reasons for eliminating points or sensitivities are provided in Appendix B.

Potential energy curves (PECs) of $3d$ Fe(II) and $4d$ Ru(II) in LS singlet and HS quintet states were obtained for homoleptic complexes of He atom and CO ligands also using TeraChem and B3LYP/LACVP* with modified a_{HF} values. These PECs were obtained by rigidly shifting all six He atom (CO) ligands from distances as short as 1.70 Å (1.80 Å) to as long as 2.90 Å (2.80 Å) in 0.01 Å (0.02 Å) increments, with calculations at longer bond lengths starting from the converged wavefunction at the shorter bond lengths. The He atom was selected following recent work³¹ that showed it is a representative weak field ligand, and a single He atom is simpler to translate during PEC evaluation than a typical non-linear weak-field ligand (i.e., H₂O or NH₃). The CO bond length was fixed to 1.125 Å, its value in relaxed TMCs. These calculations were repeated for a_{HF} fractions increasing from 0.0 to 0.45 in increments of 0.05, with wavefunctions always initializing from shorter bond lengths and lower a_{HF} values.

3.2.3 Data set construction and electronic structure calculations – 2p/3p

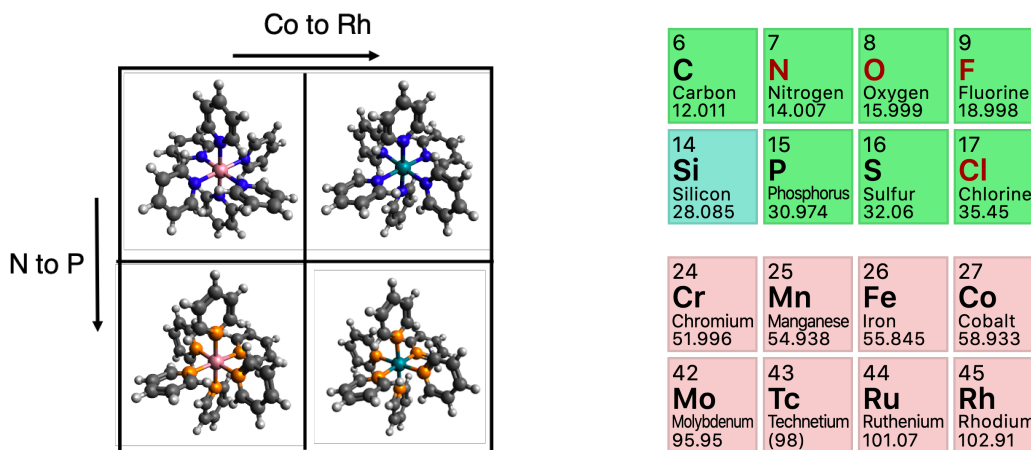


Figure 3-2. Isovalent mutation example and isovalent counterparts

After the data in section 3.2.1 was generated, all attempted mononuclear octahedral homoleptic DFT-optimized complexes (where $a_{\text{HF}} = 0.2$) in our previously reported database of TMCs were selected, resulting in 2,337 complexes. Each of these complexes was modified to produce three derivative complexes by perturbing the metal identity and connecting atom identities; for example, a $2p$ ligand connecting atom would be converted to a $3p$ atom and vice versa, and a $3d$ transition metal would be converted to a $4d$ transition metal and vice versa – this resulted in 9,348 complexes. Then, to

identify complexes with usable initial structures, only those with a parent complex passing previously reported geometry and structure criteria were retained, leading to 7,944 complexes. Then, complexes in this subset with duplicate identities (defined as identical RACs, spin states, and oxidation states) were removed, resulting in 4,602 complexes. Finally, TMCs whose parent complex belongs to the previously reported MD2 subset were retained, resulting in 1,301 complexes. The result of this filtering is shown in Figure 3-3.

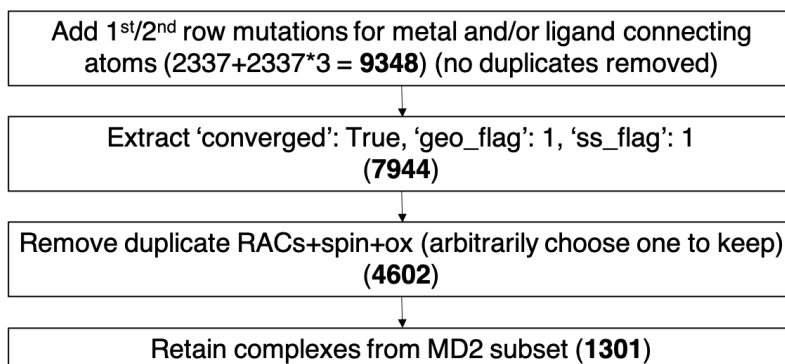


Figure 3-3. Construction of dataset for isovalent mutations

DFT calculations were then performed on this MD2 subset at $a_{\text{HF}} = 0.2$ using the same procedure outlined section 3.2.2, resulting in values for properties such as $\Delta E_{\text{H-L}}$, HOMO levels, and normalized bond lengths for each complex.

3.3 Results and discussion

3.3.1 3d vs. 4d transition metal complexes

Previous work by the Kulik group using the 3d/4d data described above has been used to demonstrate several trends relevant to exploration of 4d TMC chemical space. For example, as shown in Figure 3-4 (reproduced from the previous work), 4d TMCs tend to be more LS-shifted than 3d TMCs and have lower exchange sensitivity as well. Other observations made include that (a) strong-field ligands obey additivity rules for both $\Delta E_{\text{H-L}}$ and $S(\Delta E_{\text{H-L}})$, (b) 4d TMC energetics can deviate by up to 10 kcal/mol per 10% change in Hartree-Fock exchange, and (c) 4d TMCs are less likely than 3d TMCs to have a change in spin state ordering over the range of Hartree-Fock exchange fractions studied.

The results of this study indicate that, due to LS-shifting and reduced exchange sensitivity, $4d$ TMCs are less likely to be considered spin-crossover complexes than $3d$ TMCs using conventional DFT tuning approaches. However, if we do wish to explore the space of $4d$ TMCs anyway, we can use ligand additivity to efficiently explore $4d$ TMCs space by only studying homoleptic complexes, as suggested by previous work on ligand field asymmetry outlined in Chapter 2.

Other work by the Kulik group using $3d/4d$ data described above as part of a larger data set has shown how representation vectors can be modified to improve transfer learning across rows of the periodic table. In Figure 3-5 (reproduced from the previous work), the $3d/4d$ data was used to train machine learning models to extrapolate $4d$ TMC electronic properties from $3d$ data using a very small amount of $4d$ data (i.e., 20 data points). This work demonstrated that a tailored representation (eRAC-185) incorporating a heuristic group number property to the RAC vector can improve model performance. Combined with the inclusion of a small amount of $4d$ data, extrapolation performance improved significantly.

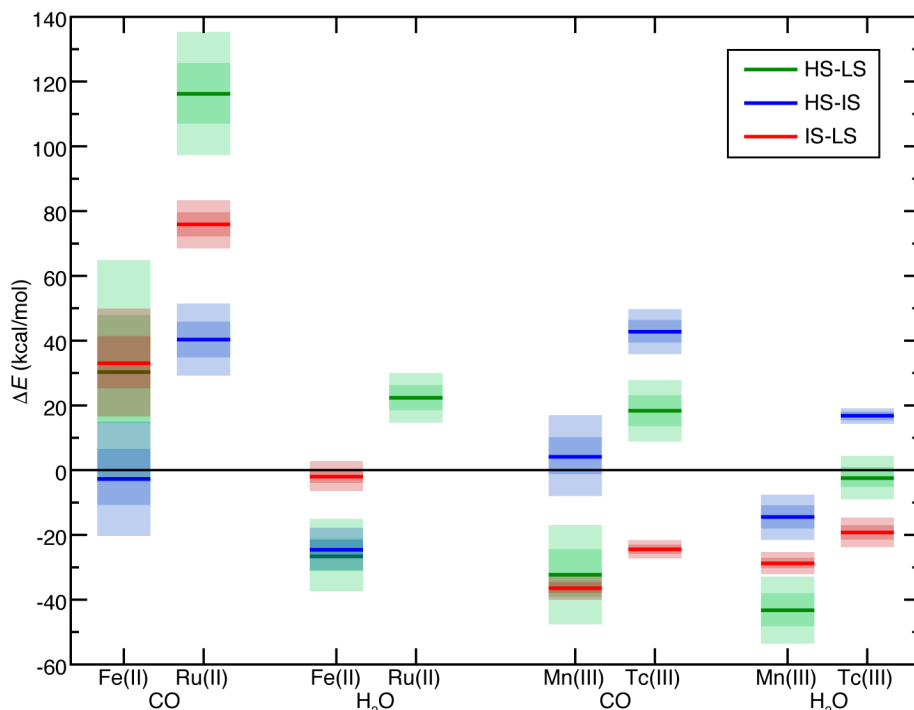


Figure 3-4. Comparisons of adiabatic spin splitting (ΔE , in kcal/mol) for HS-LS (green lines and shading), HS-IS (blue lines and shading) and IS-LS (red lines and shading) for pairs of homoleptic TMCs grouped first by the isovalent $3d$ and $4d$ metals (i.e., Fe(II)/Ru(II) vs Mn(III)/Tc(III)) and then by ligand (i.e., CO vs H₂O), as indicated on

the x -axis. The solid lines correspond to values at $a_{\text{HF}} = 0.2$, the inner translucent shaded regions correspond to the $a_{\text{HF}} = 0.1\text{--}0.3$ range, and the outer translucent shaded regions correspond to $a_{\text{HF}} = 0.0\text{--}0.4$. A zero axis is shown to indicate where ordering changes for any pair of states. The Ru(II)(H₂O)₆ IS state was eliminated during filtering steps, and so its HS-IS or IS-LS data is unavailable.

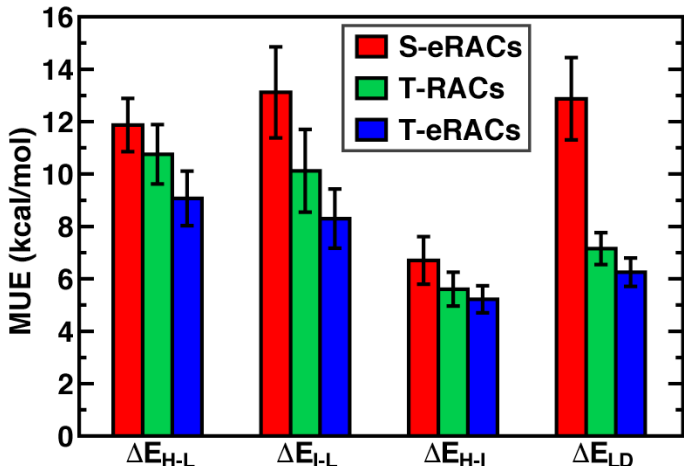


Figure 3-5. Mean unsigned error (MUE, in kcal/mol) for the prediction of $4d$ TMC properties with only 20 $4d$ data points with eRAC-185 (red) or by including $3d$ data with 20 $4d$ data points and using either the RAC-155 (green) or eRAC-185 (blue) on the isovalent metal pairing data set. The colored bars represent the average from an ensemble of 25 feature-selected KRR models, and the error bars are the standard deviation of the ensemble.

3.3.2 Incorporating the effect of 2p vs. 3p ligands

Comparison of $\Delta E_{\text{H-L}}$, HOMO levels, and normalized bond lengths across metal and ligand connecting atoms revealed that the effects of individual atom perturbations are uncorrelated but act additively for these properties. To clarify what this means, we define a property P to represent a quantity such as $\Delta E_{\text{H-L}}$, HOMO level, or normalized bond lengths, and define five types of perturbations as follows, where the subscript “ligand” indicates a change in ligand connecting atom identity from $2p$ to $3p$ and the subscript “metal” indicates a change in metal identity from $3d$ to $4d$:

$$\begin{aligned} \Delta P_{\text{ligand},3d} &= P_{3p,3d} - P_{2p,3d} \\ \Delta P_{\text{ligand},4d} &= P_{3p,4d} - P_{2p,4d} \\ \Delta P_{2p,\text{metal}} &= P_{2p,4d} - P_{2p,3d} \\ \Delta P_{3p,\text{metal}} &= P_{3p,4d} - P_{3p,3d} \end{aligned}$$

$$\Delta P_{ligand,metal} = P_{3p,4d} - P_{2p,3d}$$

We find that, although individual perturbations are not correlated (e.g., there is no correlation between $\Delta P_{ligand,3d}$ and $\Delta P_{ligand,4d}$), combined perturbations act additively (i.e., $\Delta P_{ligand,metal} \propto (\Delta P_{ligand,3d} + \Delta P_{2p,metal})$), as shown in Figure 3-6. There exist several examples that lie well off the correlation trend; these examples may be of interest for further investigation.

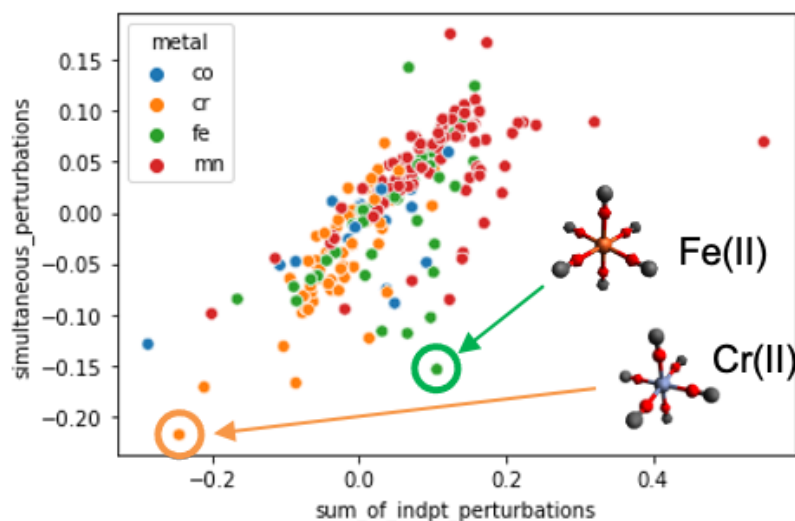


Figure 3-6. Effect of simultaneous metal-ligand perturbation vs. sum of independent perturbations on normalized bond length. The illustrated TMCs contain 3d metals and 2p ligands.

3.4 Conclusion

This investigation of isoivalent transition metal complex structures incorporating 2p and 3p ligands, as well as 3d and 4d metals, has provided insights into underlying chemical trends in TMC design across rows of the periodic table. By expanding our scope to include new elements, we have uncovered additional structure-property relationships and chemical behaviors that describe the behavior of TMCs involving components beyond a limited range of ligands and metals. These findings not only enhance our understanding of the complex interplay between ligands, metals, and their resulting properties but also offer a more comprehensive perspective on the chemical space of transition metal complexes.

The results from this work can be (and have been) utilized to inform and improve machine learning models for property prediction in transition metal complexes. By incorporating the trends and relationships identified through the examination of isovalent structures, we can refine the predictive capabilities of these models, ultimately enabling more accurate and efficient exploration of inorganic chemical space. This, in turn, can accelerate the discovery of novel transition metal complexes with desirable properties.

3.5 References

- ¹ Tsuchida, R., Absorption Spectra of Co-ordination Compounds. I. *BCSJ* **1938**, *13*, 388-400.
- ² 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*, 13973-13986.
- ³ Duan, C.; Janet, J. P.; Liu, F.; Nandy, A.; Kulik, H. J., Learning from Failure: Predicting Electronic Structure Calculation Outcomes with Machine Learning Models. *Journal of Chemical Theory and Computation* **2019**, *15*, 2331-2345.
- ⁴ KulikGroup molSimplify & molSimplify Automatic Design. <https://github.com/hjkgrp/molsimplify> (accessed June 1, 2020).
- ⁵ 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.
- ⁶ KulikGroup molSimplify Documentation. <http://molsimplify.mit.edu> (accessed June 1, 2020).
- ⁷ O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R., Open Babel: An Open Chemical Toolbox. *J. Cheminform* **2011**, *3*, 33.
- ⁸ O'Boyle, N. M.; Morley, C.; Hutchison, G. R., Pybel: A Python Wrapper for the Openbabel Cheminformatics Toolkit. *Chem Cent J* **2008**, *2*, 5.
- ⁹ Janet, J. P.; Chan, L.; Kulik, H. J., Accelerating Chemical Discovery with Machine Learning: Simulated Evolution of Spin Crossover Complexes with an Artificial Neural Network. *The Journal of Physical Chemistry Letters* **2018**, *9*, 1064-1071.

- ¹⁰ Ufimtsev, I. S.; Martinez, T. J., Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *Journal of Chemical Theory and Computation* **2009**, *5*, 2619-2628.
- ¹¹ Petachem. <http://www.petachem.com>. (accessed June 1, 2020).
- ¹² Becke, A. D., Density-Functional Thermochemistry. III. The Role of Exact Exchange. *Journal of Chemical Physics* **1993**, *98*, 5648-5652.
- ¹³ Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Physical Review B* **1988**, *37*, 785--789.
- ¹⁴ 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.
- ¹⁵ Ioannidis, E. I.; Kulik, H. J., Towards Quantifying the Role of Exact Exchange in Predictions of Transition Metal Complex Properties. *The Journal of Chemical Physics* **2015**, *143*, 034104.
- ¹⁶ Janet, J. P.; Kulik, H. J., Predicting Electronic Structure Properties of Transition Metal Complexes with Neural Networks. *Chemical Science* **2017**, *8*, 5137-5152.
- ¹⁷ Ioannidis, E. I.; Kulik, H. J., Ligand-Field-Dependent Behavior of Meta-GGA Exchange in Transition-Metal Complex Spin-State Ordering. *The Journal of Physical Chemistry A* **2017**, *121*, 874-884.
- ¹⁸ Liu, F.; Yang, T.; Yang, J.; Xu, E.; Bajaj, A.; Kulik, H. J., Bridging the Homogeneous-Heterogeneous Divide: Modeling Spin and Reactivity in Single Atom Catalysis. *Frontiers in Chemistry* **2019**, *7*, 219.
- ¹⁹ Hay, P. J.; Wadt, W. R., Ab Initio Effective Core Potentials for Molecular Calculations. Potentials for the Transition Metal Atoms Sc to Hg. *Journal of Chemical Physics* **1985**, *82*, 270-283.
- ²⁰ Saunders, V. R. H., I. H., A "Level-Shifting" Method for Converging Closed Shell Hartree-Fock Wave Functions. *Int J Quantum Chem* **1973**, *7*, 699-705.
- ²¹ Wang, L.-P.; Song, C., Geometry Optimization Made Simple with Translation and Rotation Coordinates. *The Journal of Chemical Physics* **2016**, *144*, 214108.
- ²² Janet, J. P.; Kulik, H. J., Predicting Electronic Structure Properties of Transition Metal Complexes with Neural Networks. *Chemical Science* **2017**, *8*, 5137-5152.

- ²³ Gani, T. Z. H.; Kulik, H. J., Unifying Exchange Sensitivity in Transition Metal Spin-State Ordering and Catalysis through Bond Valence Metrics *Journal of Chemical Theory and Computation* **2017**, *13*, 5443-5457.
- ²⁴ Ganzenmüller, G.; Berkaine, N.; Fouqueau, A.; Casida, M. E.; Reiher, M., Comparison of Density Functionals for Differences between the High- (T2g5) and Low- (A1g1) Spin States of Iron(II) Compounds. IV. Results for the Ferrous Complexes [Fe(L)(‘NHS4’)]. *The Journal of Chemical Physics* **2005**, *122*, 234321.
- ²⁵ Droghetti, A.; Alfè, D.; Sanvito, S., Assessment of Density Functional Theory for Iron (II) Molecules across the Spin-Crossover Transition. *The Journal of chemical physics* **2012**, *137*, 124303.
- ²⁶ Reiher, M.; Salomon, O.; Hess, B. A., Reparameterization of Hybrid Functionals Based on Energy Differences of States of Different Multiplicity. *Theoretical Chemistry Accounts* **2001**, *107*, 48-55.
- ²⁷ Zhao, Q.; Kulik, H. J., Stable Surfaces That Bind Too Tightly: Can Range Separated Hybrids or DFT+U Improve Paradoxical Descriptions of Surface Chemistry? *The Journal of Physical Chemistry Letters* **2019**, *10*, 5090-5098.
- ²⁸ Gani, T. Z. H.; Kulik, H. J., Where Does the Density Localize? Convergent Behavior for Global Hybrids, Range Separation, and DFT+U *Journal of Chemical Theory and Computation* **2016**, *12*, 5931-5945.
- ²⁹ Liu, F.; Kulik, H. J., Impact of Approximate DFT Density Delocalization Error on Potential Energy Surfaces in Transition Metal Chemistry. *Journal of Chemical Theory and Computation* **2020**, *16*, 264-277.
- ³⁰ Zhao, Q.; Kulik, H. J., Where Does the Density Localize in the Solid State? Divergent Behavior for Hybrids and DFT+U. *Journal of Chemical Theory and Computation* **2018**, *14*, 670-683.
- ³¹ Feldt, M.; Phung, Q. M.; Pierloot, K.; Mata, R. A.; Harvey, J. N., Limits of Coupled-Cluster Calculations for Non-Heme Iron Complexes. *Journal of chemical theory and computation* **2019**, *15*, 922-937.

4 Accelerating training data curation for autonomous workflows through parallel computing

This work has previously appeared in the documentation for the UFF.jl package, which can be accessed at <https://github.com/naveenarun/UFF.jl>.

4.1 Introduction

Research and development in chemistry can often be expensive or hazardous due to labor costs, chemical hazards, and conditions such as high pressures and temperatures. In the past, chemists have turned to physical simulation of matter at the atomic scale as an alternative to performing experiments in a laboratory setting. The adoption of so-called *in silico* techniques has led to the development high throughput virtual screening (HTVS), which allows the chemist to filter a large library of molecules for promising leads by performing separate *in silico* chemical experiments or characterizations for each molecule in the library. As suggested by Figure 1, these techniques have led to the discovery of new therapeutics¹ and materials², and will result in faster generation of novel matter in the future as computational tools and modeling techniques continue to improve over time.

In practice, physical simulations of matter suffer from a tradeoff between time efficiency, accuracy, and system size given a fixed set of computational resources. To resolve this issue, a hierarchy of methods have been developed for systems of different scales. For example, single atoms and small molecules can be simulated to a high degree of chemical accuracy using quantum mechanical methods such as density functional theory (DFT) and coupled cluster (CC) theory; these methods are numerically expensive as they attempt to directly solve the Schrödinger equation for multiple bodies. On the other hand, the largest length and time scales can be simulated using classical mechanics; for example, a rocket's orbit in space can be simulated accurately using classical laws of motion. However, at intermediate length and time scales (nanometers to micrometers; picoseconds to milliseconds), the ideal choice between quantum mechanical

and classical simulation becomes unclear. Ideally, we would like to perform physical simulations taking advantage of both, while taking maximum advantage of the computing resources available. To do this, we turn to molecular mechanics (MM) methods implemented in a parallel computing environment.

Molecular mechanics (MM) methods are a class of simulation methods that model molecules as classical systems; for example, one possible MM approach for simulating 1000 water molecules for 1 ns might involve setting up 1000 ball-and-spring structures with a specified spring constant for the O-H bond and another specified spring constant for the O-H-O angle, and then allowing the structures to interact with each other at a distance by representing long-range interaction energies with the Lennard-Jones potential (which tends to accurately describe the behavior of noble gases). Because all the forces in the system are specified by the imposed potential energy functions, the positions of the atoms in the system can then be integrated over time, e.g. for 1000 timesteps of 1 ps each. Because these tasks (setting up long-distance potentials, imposing bond-bond and angle potentials, integrating positions over time) are so common in molecular mechanics simulations, there exist dozens of software suites that efficiently perform these tasks (e.g. LAMMPS³ and NAMD⁴), each of which are tailored to different application domains (e.g. inorganic vs organic, small-molecule vs. large-molecule) and architectures (e.g. GPU vs non-GPU).

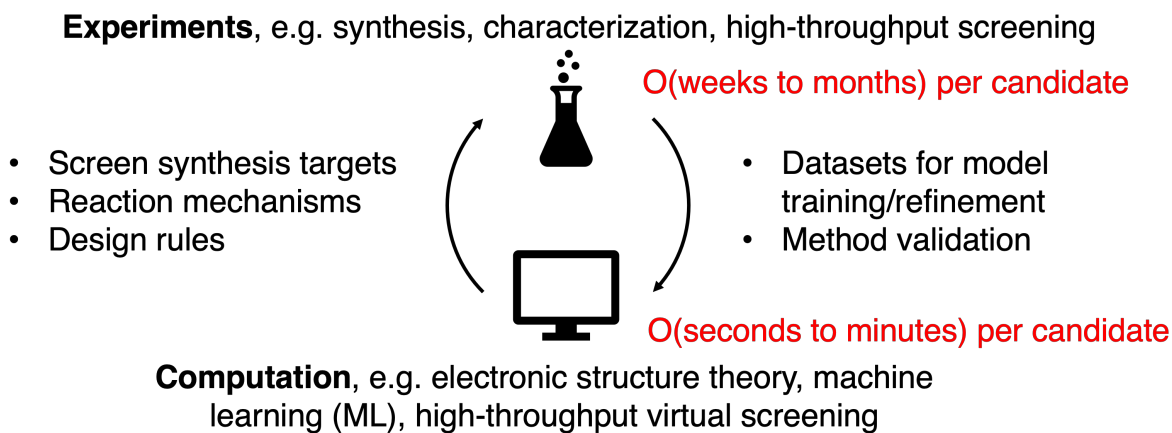


Figure 4-1. Relationship between computation and experiment

Although many molecular mechanics software suites provide a wide range of powerful tools, many could benefit from being scriptable, opensource, and taking fuller advantage of CPU and/or GPU parallelism. The CESMIX software suite⁵, combined with JuliaMolSim⁶, attempt to address this need by providing molecular simulation tools

that can be directly set up and called within Julia in a parallel environment. Currently, the two codebases provide tools for both quantum mechanical and classical simulations of atoms and molecules. For MM simulations, Molly.jl can be used to load in OpenMM force field files⁷ (e.g. for CHARMM, AMBER) or GROMACS force field files (e.g. OPLS⁸) to specify various intra- and intermolecular potential energy functions. However, it is currently unclear how one should introduce a non-OpenMM force field (or a learned force field) into either suite; on the one hand, the CESMIX suite tends to be specialized in atomistic simulations, while on the other hand, JuliaMolSim provides rigid tooling for reading in force field parameters. Thus, the focus of this project is to introduce a standalone tool called UFF.jl to interface with both suites to represent and load force field parameters into MM simulations.

The need for such a class becomes apparent when working with the universal force field (UFF)⁹, which is not associated with an OpenMM or GROMACS force field file. UFF is an empirical potential that provides parameters for bond strengths, long-distance interactions, angle forces, torsional barriers, and more across the entire periodic table. Because it provides parameters across the entire periodic table, it is useful for the simulation of metal-containing molecules such as coordination complexes (like the one shown in Figure 2, which consist of transition metals surrounded by organic ligands. In HTVS environments, the force field can be used to generate a guess of a molecule’s 3D structure before a higher level of theory such as DFT is used to refine the 3D structure. Because UFF is useful on a practical level for many simulation tasks, implementations can be found in a wide range of existing software packages such as LAMMPS and NAMD, but as of March 2022 the only UFF implementation in the Julia ecosystem is in PorousMaterials.jl¹⁰, with support limited to long-range atomic interactions. The lack of molecular implementations of UFF is likely due to the absence of tooling for integrating custom force field parameters into Julia’s MM ecosystem. This project aims to address this issue by providing an implementation of UFF that can be directly loaded into a parent molecular mechanics framework, and allowing users to perform UFF and custom force field simulations in a way that takes advantage of the straightforward differentiability and parallelism provided by the Julia environment.

4.2 Methods

Within UFF, the potential energy of a molecule is represented as a sum of two-body, three-body, and four-body interactions, some of which are bonded interactions and

some of which are nonbonded. The potential energy can be summarized by the following formula:

$$E = E_R + E_\theta + E_\phi + E_\omega + E_{vdw} + E_{el}$$

(Equation 4-1)

E_R represents the bond-stretching interaction; E_θ , E_ϕ , and E_ω represent angle bending, dihedral angle torsion, and inversion respectively. Two nonbonded interactions, E_{vdw} and E_{el} , are also included, and they represent the van der Waals interaction and electrostatic interactions respectively. All terms in this expression can be derived from a set of 9 parameters per atom, summarized in the table above. In this work, we implement the two-, three-, and four-body bonded terms E_R , E_θ , and E_ϕ , as well as the two-body nonbonded terms E_{vdw} and E_{el} .

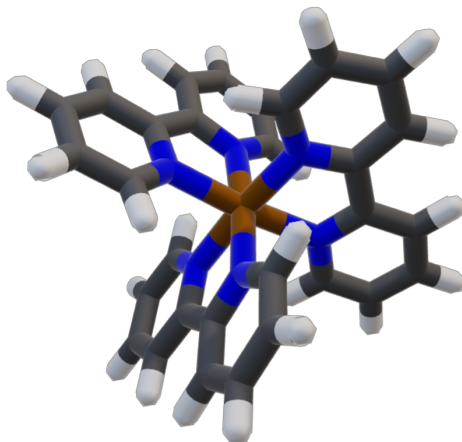


Figure 4-2. Example of a transition metal complex, which contains a central transition metal surrounded by organic ligands.

Table 4-1. Atomic parameters provided by UFF.

Symbol	Description	Units
r_I	UFF atomic radius	Angstroms
θ_0	equilibrium bond angle associated with atom	degrees
X_I	atomic van der Waals radius	Angstroms

D_I	atomic van der Waals energy	kcal/mol
ζ	van der Waals shape term used to derive x_I and D_I	unitless
Z_I	effective atomic charge	charge
V_I	sp^3 torsional barrier coefficient	kcal/mol
U_I	sp^2 torsional barrier coefficient	kcal/mol
χ	GMP ¹¹ electronegativity correction parameter	unitless

4.2.1 Bond-Stretching

The potential function for the bond-stretching interaction between atoms I and J is expressed as a harmonic oscillator by the formula

$$E_R = \frac{1}{2} k_{IJ} (r - r_{IJ})^2$$

(Equation 4-2)

Where

$$r_{IJ} = r_I + r_J + r_{BO} + r_{EN}$$

(Equation 4-3)

for which the bond order correction (r_{BO}) is

$$r_{BO} = -\lambda(r_I + r_J)\ln(n)$$

(Equation 4-4)

and the electronegativity correction (r_{EN}) is

$$r_{EN} = \frac{r_I r_J (\sqrt{\chi_I} - \sqrt{\chi_J})^2}{\chi_I r_I + \chi_J r_J}$$

(Equation 4-5)

In the above expressions, λ is a constant, $\lambda = 0.1332$ (empirically determined from propane/propene/propyne) and n is a bond order.

The force constant k_{IJ} can be evaluated using the empirical expression

$$k_{IJ} = 664.12 \frac{Z_I Z_J}{r_{IJ}^3}$$

(Equation 4-6)

4.2.2 Angular bending

Next, the angular bending term for atoms I , J , and K , where J is the central atom, can be calculated as

$$E_\theta = K_{IJK}(C_0 + C_1 \cos \theta + C_2 \cos 2\theta)$$

(Equation 4-7)

where

$$\begin{aligned} C_2 &= 1/(4 \sin^2 \theta) \\ C_1 &= -4C_2 \cos \theta_0 \\ C_0 &= C_2(2 \cos^2 \theta_0 + 1) \end{aligned}$$

for general nonlinear bonds. However, for linear, trigonal-planar, square-planar, and octahedral geometries, $n = 1, 2, 3, 4$ should be used respectively in the following formula instead:

$$E_\theta = \frac{K_{IJK}}{n^2} (1 - \cos n\theta)$$

(Equation 4-8)

In both expressions, K_{IJK} is defined as

$$K_{IJK} = \beta \frac{Z_I Z_K}{r_{IK}^5} r_{IJ} r_{JK} (3r_{IJ} r_{JK} (1 - \cos^2 \theta_0) - r_{IK}^2 \cos \theta_0)$$

(Equation 4-9)

where

$$r_{IK}^2 = r_{IJ}^2 + r_{JK}^2 - 2r_{IJ} r_{JK} \cos \theta$$

(Equation 4-10)

and

$$\beta = \frac{664.12}{r_{IJ}r_{JK}}$$

(Equation 4-11)

4.2.3 Dihedral torsion

The torsional term E_ϕ is calculated as

$$E_\phi = \frac{1}{2}V_\phi(1 - \cos n\phi_0 \cos n\phi)$$

(Equation 4-12)

where n and ϕ_0 vary based on the nature of the central two atoms (for example, for an sp^3-sp^3 bond, $n = 3$ and $\phi_0 = 180^\circ$). For two sp^3 centers, the torsional barrier V_ϕ is

$$V_{sp^3} = \sqrt{V_J V_K}$$

(Equation 4-13)

whereas the barrier for a bond involving an sp^2 center is

$$V_{sp^2} = 5\sqrt{U_J U_K}(1 + 4.18 \ln BO_{JK})$$

Equation 4-14

where BO_{JK} is the bond order between atoms J and K .

4.2.4 Inversion

To calculate the inversion term for atom I bonded to atoms J , K , and L , the following formula can be used:

$$E_\omega = K_{IJKL}(C_0 + C_1 \cos \omega_{IJKL} + C_2 \cos 2\omega_{IJKL})$$

(Equation 4-15)

where ω_{IJKL} is the angle between IL and the IJK plane.

The original UFF paper does not provide much guidance on how to derive K_{IJKL} , and specifies that for atom types other than C, N, O, P, As, Sb, and Bi, the inversion term is 0. Thus, we omit the inversion term for the time being.

4.2.5 van der Waals interaction

The form of E_{vdw} is the Lennard-Jones 12-6 potential,

$$E_{vdw} = D_{IJ} \left(-2 \left(\frac{x_{IJ}}{x} \right)^6 + \left(\frac{x_{IJ}}{x} \right)^{12} \right)$$

(Equation 4-16)

where x is the distance between two atoms. In this formula, x_{IJ} and D_{IJ} are calculated as the geometric means of their atomic counterparts, i.e.

$$x_{IJ} = \sqrt{x_I x_J}$$

(Equation 4-17)

$$D_{IJ} = \sqrt{D_I D_J}$$

(Equation 4-18)

In the UFF formulation, the terms in the 12-6 potential are derived by matching ζ terms in the expression

$$E_{vdw} = \left(D_{IJ} \left(\frac{6}{\zeta - 6} \right) e^\zeta \right) e^{-\zeta \left(\frac{x}{x_{IJ}} \right)} - \frac{\left(D_{IJ} \left(\frac{6}{\zeta - 6} \right) x_{IJ}^6 \right)}{x^6}$$

(Equation 4-19)

to the form of the 12-6 potential.

4.2.6 Electrostatic Interactions.

The electrostatic contribution to the energy is expressed as

$$E_{el} = 332.0637 \frac{Q_I Q_J}{\epsilon R_{IJ}}$$

(Equation 4-20)

where ϵ is 1 and R_{IJ} is the distance between atoms I and J .

Note that in UFF, both types of nonbonded interactions (van der Waals and electrostatic) are excluded for atoms that are up to 2 bonds away from each other. The functional forms of the bonded potentials are shown in Figure 3.

4.3 Results and discussion

A thorough performance analysis of the code was performed to determine the efficiency and scaling behavior of parameter generation. Two types of scaling were studied: system size, i.e. the number of unique atom types and degree of parallelization. Additionally, the speed of empirical simulation was tested relative to degree of parallelization.

Several UFF potentials were plotted to ensure that results were chemically reasonable. The results are shown below.

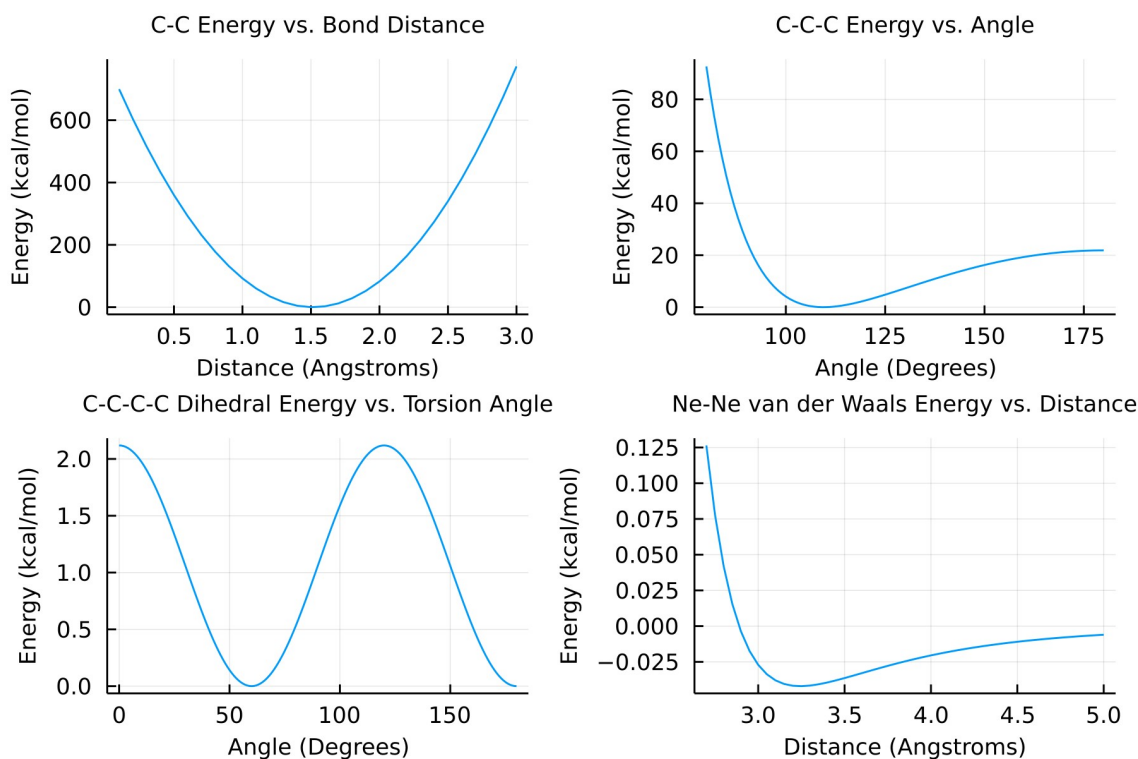


Figure 4-3. Calculated forms of the various potentials provided by UFF.

As expected, the optimal C-C bond distance is approximately 1.5 Angstroms, the optimal C-C-C angle is about 109.5 degrees, and the optimal dihedral angles for a C-C-C-C chain are either 60 degrees or 180 degrees. Additionally, the van der Waals potential energy surface for Ne atoms has the expected functional form. These results

are chemically reasonable, and demonstrate that the implementation can be meaningfully used for optimizing chemical structures.

4.3.1 Performance analysis of parameter generation time vs. system size

For a system containing N atoms, there are $N(N + 1)$ bond-stretching parameters to generate ($\binom{N+2-1}{2}$ unique pairs, each of which requires 2 parameters: a force constant and an equilibrium bond distance), $2N(N^2 - N + 2)$ angular bending parameters ($3\binom{N}{3} + 2\binom{N}{2} + N$ ways to create a unique chain of 3 atoms, and for each unique 3-atom bond there are three Fourier coefficients and one force constant), and $\frac{N(N+1)}{2}$ dihedral torsion parameters (one barrier coefficient per central pair of atoms). Using similar reasoning, we find that there are $N(N + 1)$ van der Waals parameters to generate and $\frac{N(N+1)}{2}$ electrostatic parameters to generate.

Thus, the total number of fixed parameters to generate is:

$$\begin{aligned} & \underbrace{N(N+1)}_{\text{bond-stretching}} + \underbrace{2N(N^2 - N + 2)}_{\text{angle-bending}} + \underbrace{\frac{N(N+1)}{2}}_{\text{torsional}} + \underbrace{N(N+1)}_{\text{van der Waals}} + \underbrace{\frac{N(N+1)}{2}}_{\text{electrostatic}} \\ &= 2N^3 + N^2 + 7N \\ &\approx O(N^3) \end{aligned}$$

Figure 4 illustrates the effect of system size (i.e. number of unique atom types) on performance.

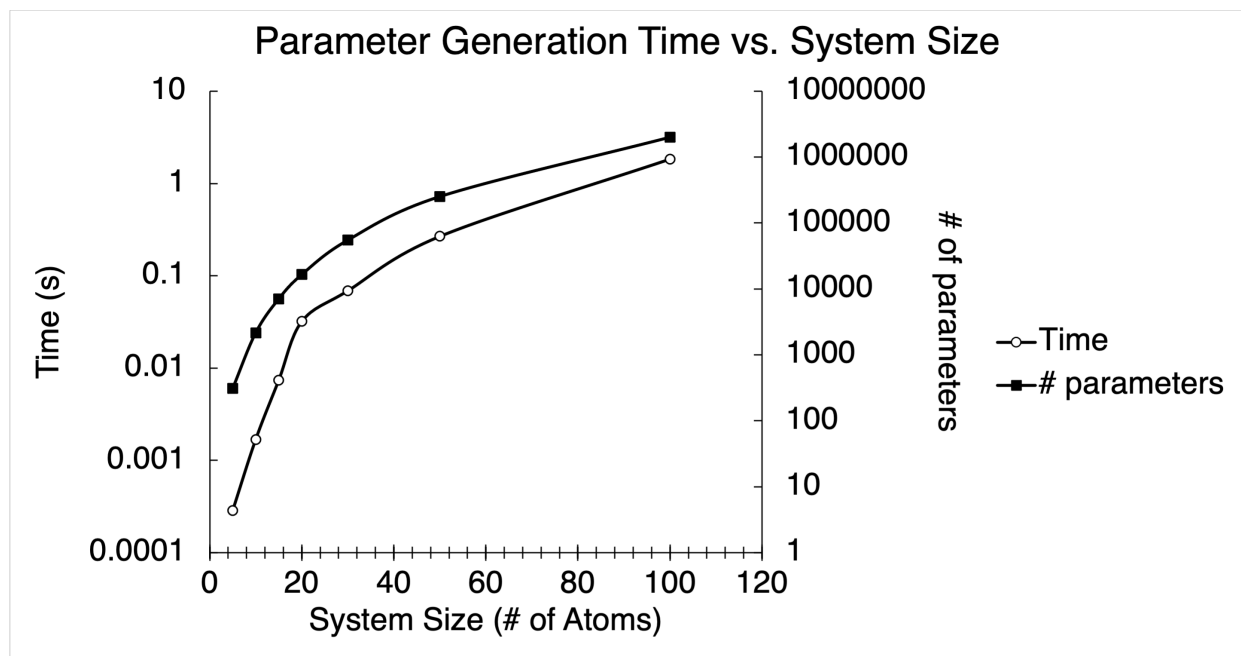


Figure 4-4. Plot illustrating how the time needed to generate a full set of parameters for N unique atom types scales with N .

The time needed to generate all parameters for N unique atom types scales approximately as N^3 . Parameter generation does not exceed 1 second for system sizes under about 90. The usefulness of Julia's performance is illustrated here by the fact that about 2 million parameters were generated in about 1.8 seconds for the system size of 100 - especially considering the many floating point operations involved in the calculation of each parameter and the fact that the code ran on a single thread on a Macbook. In order to explore how far the code's performance can be enhanced, a study of performance under multithreading was performed as well, as described in the next section.

4.3.2 Performance analysis of parameter generation time vs. degree of parallelization

All parameters were generated for system sizes ranging from 5 to 128 using both a threaded and nonthreaded implementation. The results are shown in Figure 5.

From these results, it is clear that for small system sizes, the overhead of creating threads and allocating tasks exceeds to time needed to actually calculate all parameters, meaning that for systems with up to about 25 unique atom types, it is preferable to generate parameters in a single thread. However, for larger system sizes, there are significant benefits to parallelism: for a system with 128 unique atom types, performing tasks on 4 threads results in a 2.6 times speedup, indicating that the time to calculate parameters far exceeds the time needed to create and manage threads.

Thus, for large system sizes, it is useful to parallelize parameter generation; however for small system sizes, implementing parallelization is not as important and can even be detrimental to performance. System sizes of about 25 unique atoms perform equally well under threaded and non-threaded implementations.

4.3.3 Local optimization of butane using forward-mode automatic differentiation

Forward-mode automatic differentiation (AD) was employed to generate a local minimum structure for butane. The optimized structure from Avogadro was used as an initial structure, and Julia's ForwardDiff package was used to generate gradients for the

length-12 coordinate vector (4 atoms with 3 coordinates per atom). The result was optimization to a structure with an energy of 1.01 kcal/mol, which is lower than the original energy of 1.52 kcal/mol. The results of the optimization trajectory are shown in Figure 6.

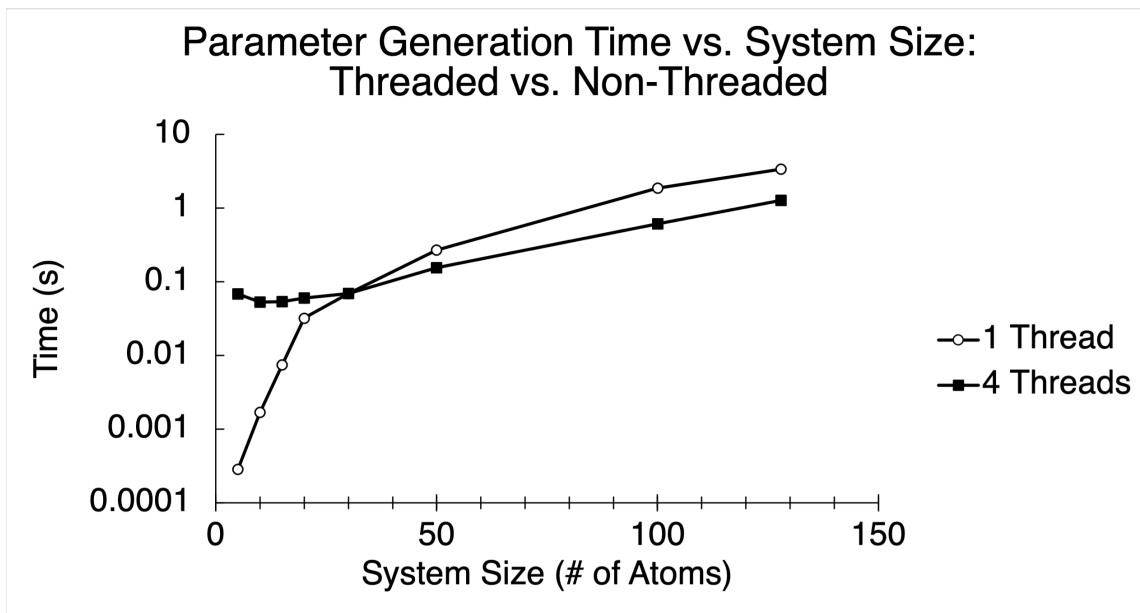


Figure 4-5. Performance of UFF.jl for generating full parameter sets in parallel and non-parallel environments.

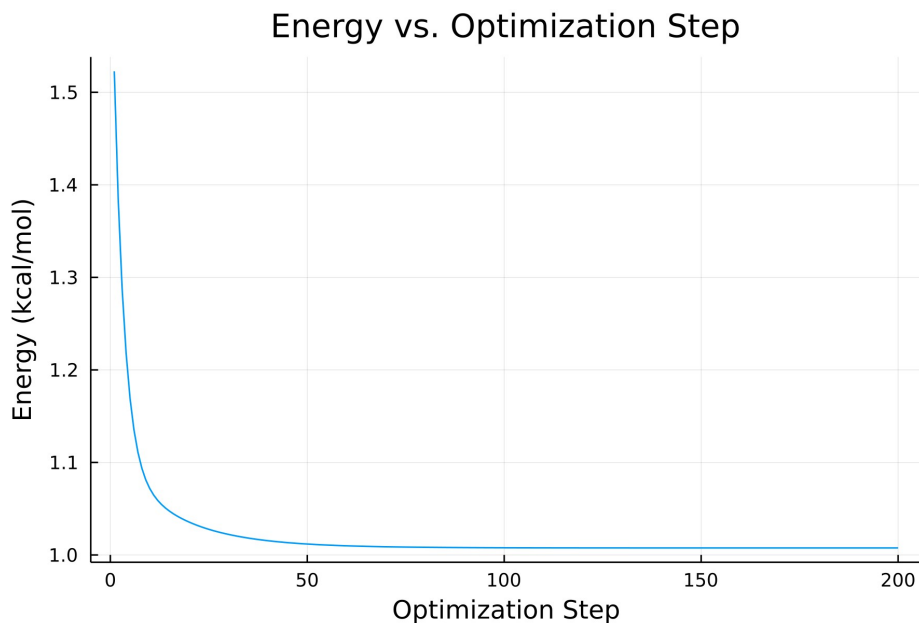


Figure 4-6. Optimization trajectory for butane using forward-mode automatic differentiation to obtain gradients.

4.3.4 Global optimization of butane using Optim.jl

In order to test the implementation of UFF in Julia, a test case was performed where butane was optimized against its UFF energy. A planar input structure was drawn in Avogadro and optimized within the program; both structures can be seen in Figure 7. The structure optimized by Avogadro settled to a local minimum with a dihedral angle of 60 degrees; however, a global minimum of 180 degrees exists because the van der Waals interaction causes a repulsive force between the first and fourth carbon. Optim.jl was used to generate a globally optimal structure based on the coordinates of the four carbons, resulting in a planar molecule with a dihedral angle of 180 degrees. Thus, the Julia implementation was able to successfully find an optimal structure for butane. Within a structured molecular mechanics framework, therefore, the implementation can be employed for large-scale simulations with many more atoms and atom types.

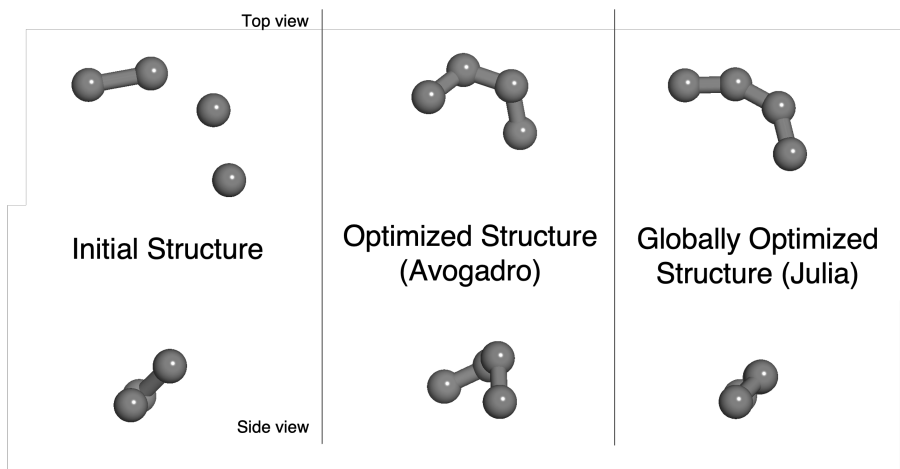


Figure 4-7. Structures of butane before and after optimization.

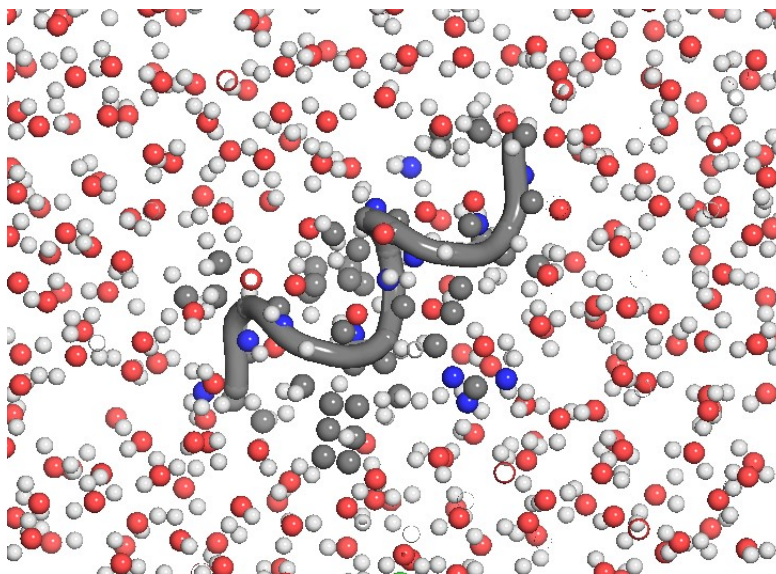


Figure 4-8. Example of a larger-scale simulation involving a protein in H_2O solvent.

4.4 Conclusion

The Universal Force Field (UFF) was implemented in Julia and tested on a range of sample systems. Performance analysis showed that the time needed to generate a parameter set for N unique atom types scales as N^3 , and that a multithreaded approach to parameter generation works best for larger N . For smaller N , a single-threaded approach is much faster due to the lack of overhead for thread setup and management; however, the time savings are significant for larger N . Quite notably, without the use of explicit force calculations or a parent molecular mechanics framework, it was possible to optimize the structure of butane via global optimization and via local gradient-based optimization, the latter of which was performed via forward-mode automatic differentiation. These feats were made relatively straightforward due to features unique to Julia, including the macro system which allows for straightforward multithreading and automatic differentiation. Future work will focus on integrating the UFF code into molecular mechanics codes, such as those in the CESMIX suite.

4.4.1 Future work

The next step for this implementation is to integrate it into the CESMIX suite (or any other molecular simulation package for Julia) so that the UFF parameters can be used in large-scale simulations. For example, in larger systems, such as the solvated protein shown in Figure 8, UFF parameters can be calculated for all chains of atoms present, and a simulation package can decide which sets of atoms should be included in energy calculations.

Additionally, it would be of interest to employ UFF parameters in a distributed computing environment, for example by dividing the energy calculation tasks for various sets of atoms among different nodes in a computer cluster. These are tasks that a wrapper package should handle, agnostic of the underlying force field.

The UFF force field comes with many exceptions and edge cases, which need to be implemented meticulously. The implementation used in this work ignores many edge cases and uses the most general formulations available; thus for future work it is important to include exceptions and edge cases in the code. Additionally, many modifications to UFF exist, and it would be of interest to provide flags so that users can select which variant(s) they would like to use.

4.5 References

- ¹ A. Dhasmana, S. Raza, R. Jahan, M. Lohani, and J. M. Arif, *High-throughput virtual screening (HTVS) of natural compounds and exploration of their biomolecular mechanisms*, in *New Look to Phytomedicine*, Elsevier, 2019, pp. 523–548, <https://doi.org/10.1016/b978-0-12-814619-4.00020-3>, <https://doi.org/10.1016/b978-0-12-814619-4.00020-3>.
- ² Omer H. Omar, M. del Cueto, T. Nematiram, and A. Troisi, *High-throughput virtual screening for organic electronics: a comparative study of alternative strategies*, *Journal of Materials Chemistry C*, 9 (2021), pp. 13557–13583, <https://doi.org/10.1039/d1tc03256a>, <https://doi.org/10.1039/d1tc03256a>.
- ³ A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J.

- Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, *LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales*, Computer Physics Communications, 271 (2022), p. 108171, <https://doi.org/10.1016/j.cpc.2021.108171>, <https://doi.org/10.1016/j.cpc.2021.108171>.
- ⁴ J. C. Phillips, D. J. Hardy, J. D. C. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Henin, W. Jiang, R. McGreevy, M. C. R. Melo, B. K. Radak, R. D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L. V. Kale, K. Schulten, C. Chipot, and E. Tajkhorshid, *Scalable molecular dynamics on CPU and GPU architectures with NAMD*, The Journal of Chemical Physics, 153 (2020), p. 044130, <https://doi.org/10.1063/5.0014475>, <https://doi.org/10.1063/5.0014475>.
- ⁵ <https://github.com/cesmix-mit/> (accessed 2022).
- ⁶ <https://github.com/JuliaMolSim/> (accessed 2022).
- ⁷ <https://github.com/openmm/openmmforcefields> (accessed 2022).
- ⁸ P. Bauer, B. Hess, and E. Lindahl, *Gromacs 2022.1 manual* (2022), <https://doi.org/10.5281/ZENODO.6451567>, <https://zenodo.org/record/6451567>.
- ⁹ A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations, Journal of the American Chemical Society, 114 (1992), pp. 10024–10035, <https://doi.org/10.1021/ja00051a040>, <https://doi.org/10.1021/ja00051a040>.
- ¹⁰ C. Simon, A. H. York, Sturluson, C. Simon, A. Henle, N. Gantzler, M. T. Huynh, CalebLaird, J. TagBot, and M. Piibeleht, *Simonensemble/porousmaterials.jl: v0.4.0*, 2022, <https://doi.org/10.5281/ZENODO.1400838>, <https://zenodo.org/record/1400838>.
- ¹¹ A. K. Rappe and W. A. Goddard, *Charge equilibration for molecular dynamics simulations*, The Journal of Physical Chemistry, 95 (1991), pp. 3358–3363, <https://doi.org/10.1021/j100161a070>, <https://doi.org/10.1021/j100161a070>.

5 Design tools for transition metal complexes and metal organic frameworks allow users to explore and provide feedback on machine learning models

Aspects of this work related to the molSimplify Lite and MOFSimplify web tools have previously appeared in C. Duan, A. Nandy, R. Meyer, N. Arunachalam, and H.J. Kulik, A transferable recommender approach for selecting the best density functional approximations in chemical discovery, *Nature Computational Science* **3**, 38 (2022) and A. Nandy, G. Terrones, N. Arunachalam, C. Duan, D.W. Kastner, and H.J. Kulik, MOFSimplify, machine learning models with extracted stability data of three thousand metal-organic frameworks, *Scientific Data* **9**, (2022) respectively. Text from the *Scientific Data* paper is included in the methods section of this chapter, and is cited accordingly.

5.1 Introduction

As the field of inorganic chemical design advances, the development of effective and user-friendly tools becomes increasingly essential for researchers to efficiently explore the vast chemical space of transition metal complexes (TMCs) and metal-organic frameworks (MOFs). Machine learning models have emerged as powerful tools to predict the properties and behavior of these compounds, assisting in the identification of promising candidates for various applications. However, the success of these models depends on their ability to adapt and improve based on user feedback, ensuring that they remain relevant and accurate in an ever-evolving research landscape.

In this chapter, we present the development of two interactive design tools, molSimplify Lite and MOFSimplify, which have been designed to facilitate the exploration of TMC and MOF chemical space while allowing users to provide feedback

on machine learning models. These web-based platforms enable researchers to design, visualize, and assess the properties of TMCs and MOFs in a user-friendly environment, allowing computational and experimental researchers to access machine learning models more easily. By collecting user feedback for use in the development and refinement of these machine learning models, these tools contribute to the iterative improvement of predictive capabilities, ensuring that the models remain useful and effective for future research endeavors.

5.2 Methods

5.2.1 Frontend and backend components for molSimplify Lite

Bootstrap with jQuery was chosen as a frontend framework due to ease of prototyping and development. An off-the-shelf, adaptive, open-source theme¹ was used to create the homepage, as shown in Figure 5-1.

molSimplify About Source Code Installation Instructions Tutorials Documentation How to Cite

molSimplify Lite

 (powered by molSimplify from the Kulik group at MIT)

Predict and view structures and spin properties of octahedral transition metal complexes in your browser below!

For more geometries, properties, and functionality, check out the full molSimplify utility. Learn more in the [About](#) tab.

Metal
Iron

Axial Ligand 1 (selection or SMILES string)
water

Axial Ligand 2 (selection or SMILES string)
water

Equatorial Ligand (selection or SMILES string)
water

Oxidation State
II

Show/hide advanced parameters

Generate! Download XYZ File Predict ΔE_{H-L} Plot ΔE_{H-L}
PCA ΔE_{H-L} Recommend DFA for ΔE_{H-L}

Load example molecule Reset zoom level

Site developed by Naveen Arunachalam and maintained by the Kulik Group at MIT

Figure 5-1. Current frontend of molSimplify Lite

To create the interactive 3D model, 3Dmol² was used (this is the same viewer we typically use in Jupyter notebooks). The xyz file is loaded into a 3Dmol class object and is then rendered in the browser automatically.

The backend of molSimplify Lite is contained in app.py in <https://github.com/hjkgrp/molSimplifyLite>. Requests involving calls to Python code (e.g. using molSimplify) are performed using Flask; for example, the backend performs homoleptic structure generation when a POST request containing molSimplify options is passed to molsimplify.mit.edu/generate. At molsimplify.mit.edu/index, the homepage is served, as detailed in the usage notes in the next section.

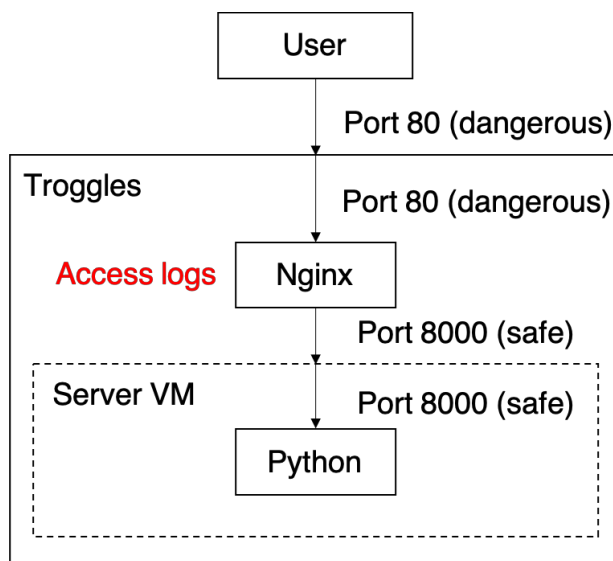


Figure 5-2. Hosting Structure of molSimplify Lite

5.2.2 Form and database components for MOFSimplify

The frontend and backend components of MOFSimplify were inherited from molSimplify Lite. In order to receive user feedback, we developed five separate feedback forms placed throughout the site. To record feedback, the forms on the MOFSimplify website were connected to a MongoDB instance hosted on the same server on the backend, which can be accessed by MOFSimplify Lite via the internet. For files uploaded via the forms, a maximum limit of 2 MB was implemented.

When a user presses submit, a packet of data containing various form fields is submitted to the MongoDB instance on the host machine. The form fields are explicitly named 'feedback_form_name', 'rating', 'email', 'reason', 'comments', 'cif_file_name',

and 'structure'. These fieldnames correspond to fields available in the online feedback form as shown below. The 'feedback_form_name' field indicates which of the five feedback forms the entry corresponds to, which allows us to use similar backend code for all five frontend forms to send requests to the MongoDB server.

Solvent removal stability ANN nearest neighbors in latent space

1. LAZXOB_clean

Show **Download**

Visualize **Get components**

Current nearest neighbor: LAZXOB_clean
 Current nearest neighbor's latent distance: 3.3013945
 Nearest neighbor DOI: 10.1021/cg300518k
 Experimentally observed to be stable upon solvent removal? Yes

Rate this existing data:

Not Rated

Email (.edu preferred) *

Reason for rating...

Comments

Upload File (optional; .pdf, .jpg, .png, .tiff; 20 ME) Browse

Submit

Thermal stability ANN nearest neighbors in latent space

1. SONVEX_neutral_b

Show **Download**

Visualize **Get components**

Current nearest neighbor: SONVEX_neutral_b
 Current nearest neighbor's latent distance: 1.3311056
 Circle in TGA plot marks the breakdown temperature.
 Nearest neighbor DOI: 10.1021/ja807357r
 Experimentally observed thermal breakdown temperature: 348.3°C

Rate this existing data:

Not Rated

Email (.edu preferred) *

Reason for rating...

Comments

Upload File (optional; .pdf, .jpg, .png, .tiff; 20 ME) Browse

Submit

Figure 5-3. Example feedback forms on MOFSimplify website.

Main
Visualization
Component Analysis
Data Upload

Upload information on your MOF!

Email (.edu preferred) *

Comments

Upload MOF (.cif; Required) Browse

Upload TGA Trace (.pdf, .jpg, .png, .tiff; 20 MB max; Required) Browse

Submit

Figure 5-4. Data upload form on MOFSimplify website.

An overall schematic of the database backup system is shown below.

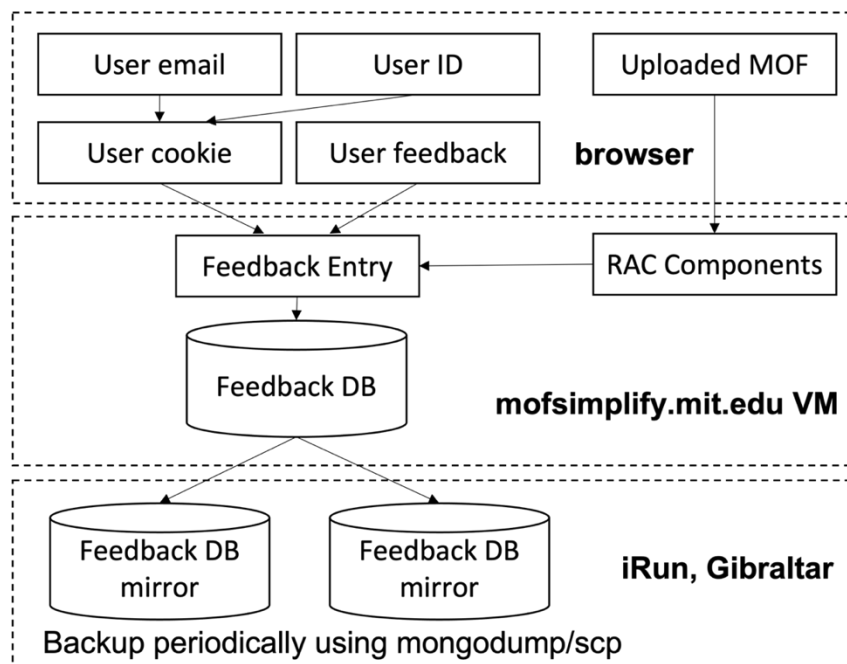


Figure 5-5. Database system diagram. iRun and Gibraltar refer to servers owned by the Kulik group. The mofsimply VM is hosted on the iRun server.

In order to improve the usability of the website, we performed testing via recorded sessions where we asked group members to perform specified tasks. During we identified that a major roadblock for uploading data was that users had to resubmit all form data in order to give feedback on multiple predictions (e.g., thermal stability and solvent removal stability). To address this issue, we modified the JavaScript code that is run after the form is submitted so that previously submitted data is retained and the page does not refresh; this made it significantly easier for users to upload feedback on multiple predictions.

5.3 Results and discussion

5.3.1 Usage notes: molSimplify Lite

As shown in Figure 5-1, the interface for molSimplify Lite provides text fields for specifying the metal, axial ligands (either from a drop-down selection or by entering SMILES strings), equatorial ligand, and oxidation state of a TMC, allowing for a wide range of customizable complex configurations.

For users seeking finer control over the generated structures, the “show/hide advanced parameters” button reveals additional options such as the Hartree-Fock

exchange fraction, which influences the predicted spin-splitting energy ΔE_{H-L} and spin multiplicity, which determines whether a low-spin or high-spin structure is generated. Once the desired parameters are entered, users can click "Generate!" to create the specified complex.

Upon generation, several new options become available to users, including the ability to download an .xyz file of the generated molecule, predict the ΔE_{H-L} value using molSimplify, and receive a recommendation for an appropriate density functional approximation. Furthermore, users can interactively explore their generated complex in the context of the training data by clicking "Plot ΔE_{H-L} " to visualize the spin-splitting energy comparison or "PCA ΔE_{H-L} " to access an interactive PCA plot of the RAC vector representation. These features allow users to gain a deeper understanding of their designed complexes and how they relate to other structures within TMC design space.

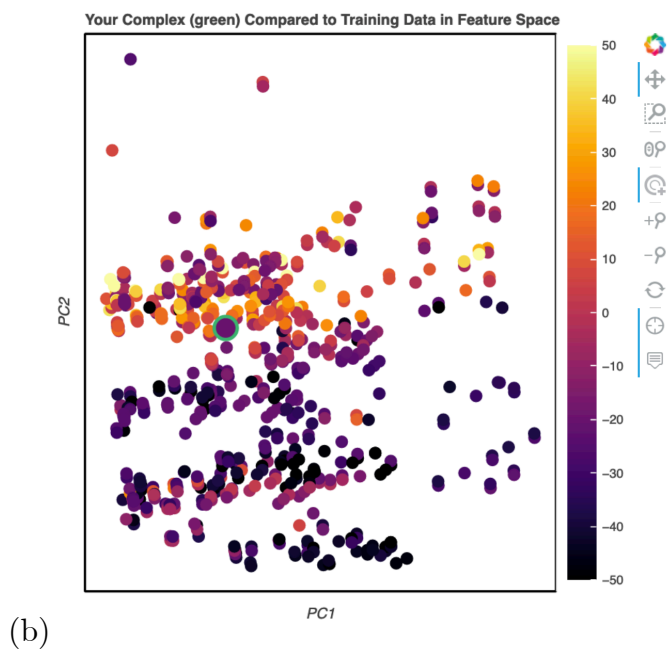
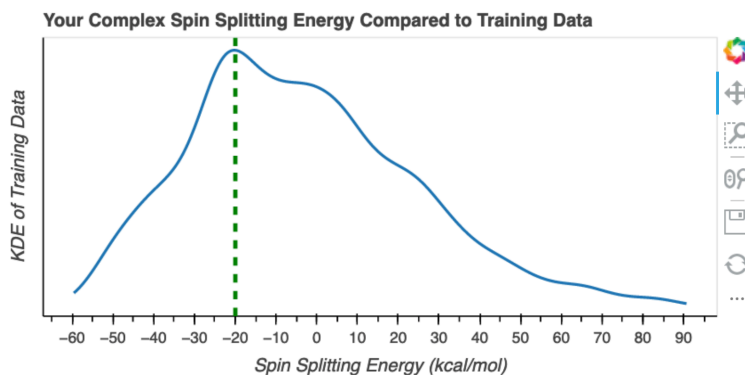


Figure 5-6. a) A plot of the predicted $\Delta E_{\text{H-L}}$ for $\text{Fe}(\text{NH}_3)(\text{H}_2\text{O})_6$ with oxidation state 2 compared to other complexes in molSimplify’s training data. b) PCA showing the same complex in green among the training data in RAC space.

Recommended density functional approximation (DFA)
 We recommend "m06" to be used for computing the vertical spin splitting of the selected complex. The deviation of the recommended DFA and our reference data (DLPNO-CCSD(T)) is predicted to be 0.49 kcal/mol.

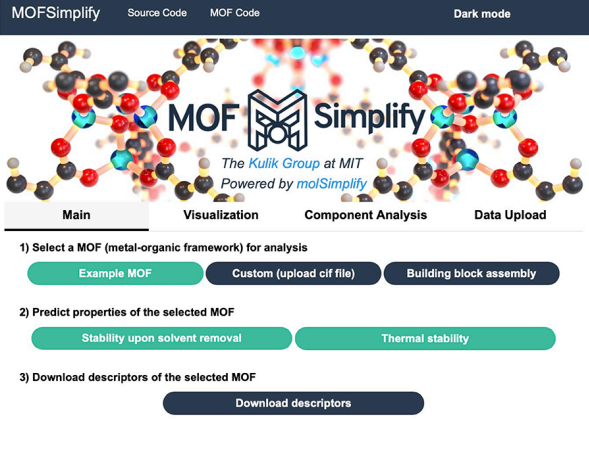
Ranked DFAs
 Below is a sorted list reporting the predicted DFA errors from our recommender:

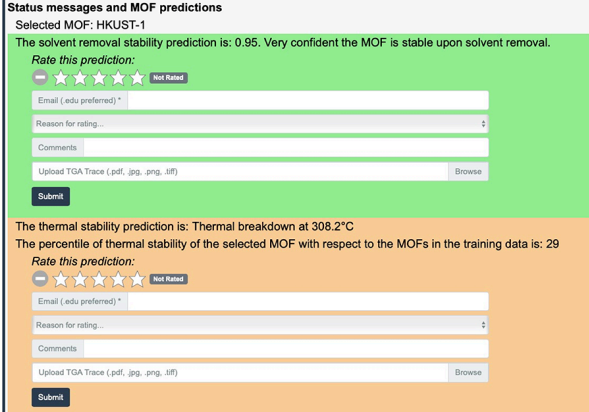
Functional	Deviation (kcal/mol)
m06	0.49
m06-l_hfx_30	0.92
dsd-blyp-d3bj	1.91
m06-l_hfx_40	1.99
mn15-l_hfx_40	2.16
dsd-pbep86-d3bj	2.39
pbe_hfx_30	2.67
b2gpplyp	2.83
scan_hfx_40	3.22
blyp_hfx_50	3.62
mn15-l_hfx_50	4.29
mn15-l_hfx_30	4.55
pbe_hfx_40	4.85
m06-l_hfx_50	4.90
m06-2x	5.47
pbe0-dh	5.59
dsd-pbeb95-d3bj	5.70

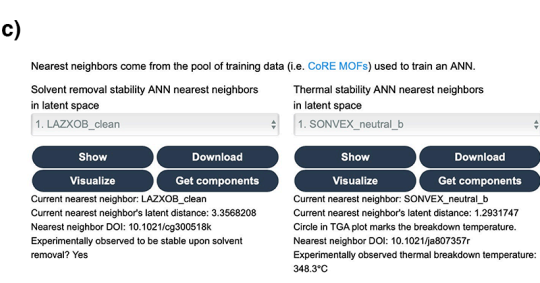
Figure 5-7. Example DFA recommendation for $\Delta E_{\text{H-L}}$ for $\text{Fe}(\text{NH}_3)(\text{H}_2\text{O})_6$. The website ranks DFAs by predicted error relative to DLPNO-CCSD(T).

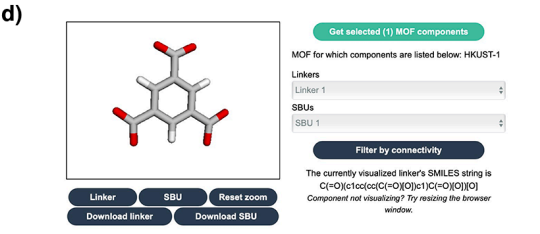
5.3.2 Usage notes: MOFSimplify³

The MOFSimplify interface encourages community engagement by enabling the user to add new MOF data to our database by uploading MOF CIF files and TGA traces in the Data Upload tab. MOFSimplify also lets the user indicate whether they agree with an ANN prediction or curated experimental data and support their position by uploading a TGA trace (Figure 5-8). These TGA traces will be digitized by us to extract T_d data in a manner consistent with our previous thermal stability data. User input will be used to improve our ANN models through community-based active learning. Users can opt out of uploading data or providing feedback. If users wish to remove data after the fact, an email form is provided for removal requests.

a)  MOFSimplify Source Code MOF Code Dark mode
The Kulik Group at MIT
Powered by molSimplify
1) Select a MOF (metal-organic framework) for analysis
Example MOF Custom (upload cif file) Building block assembly
2) Predict properties of the selected MOF
Stability upon solvent removal Thermal stability
3) Download descriptors of the selected MOF
Download descriptors

b)  Status messages and MOF predictions
Selected MOF: HKUST-1
The solvent removal stability prediction is: 0.95. Very confident the MOF is stable upon solvent removal.
Rate this prediction:
Reason for rating...
Comments
Upload TGA Trace (pdf, jpg, png, tiff)
The thermal stability prediction is: Thermal breakdown at 308.2°C
The percentile of thermal stability of the selected MOF with respect to the MOFs in the training data is: 29
Rate this prediction:
Reason for rating...
Comments
Upload TGA Trace (pdf, jpg, png, tiff)

c)  Nearest neighbors come from the pool of training data (i.e. CoRE MOFs) used to train an ANN.
Solvent removal stability ANN nearest neighbors in latent space
Thermal stability ANN nearest neighbors in latent space
Current nearest neighbor: LAZXOB_clean
Current nearest neighbor's latent distance: 3.3568208
Nearest neighbor DOI: 10.1021/cg300518k
Experimentally observed to be stable upon solvent removal? Yes
Current nearest neighbor: SONVEX_neutral_b
Current nearest neighbor's latent distance: 1.2931747
Circle in TGA plot marks the breakdown temperature.
Nearest neighbor DOI: 10.1021/ja807357r
Experimentally observed thermal breakdown temperature: 348.3°C

d)  Get selected (1) MOF components
MOF for which components are listed below: HKUST-1
Linkers
SBUs
Filter by connectivity
The currently visualized linker's SMILES string is
Component not visualizing? Try resizing the browser window.

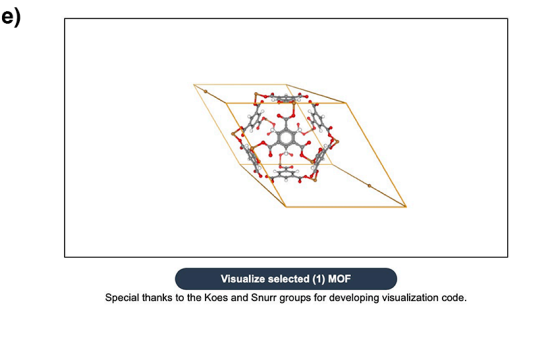
e)  Visualize selected (1) MOF
Special thanks to the Koes and Snurr groups for developing visualization code.

Figure 5-8. Sections of the MOFSimplify web interface. a) Interface for selecting a MOF for analysis and predicting properties of the selected MOF using ANNs trained on experimental data mined from the literature. The default MOF loaded upon selecting “Example MOF” is HKUST-1, a well-studied MOF.⁸⁵ b) The feedback interface for evaluating model predictions. c) The interface listing similar (i.e., LSNN) MOFs to the selected MOF as determined by the ANNs. d) Visualization of the selected MOF’s components. e) Visualization of the selected MOF’s unit cell.

5.3.3 Impact

The molSimplify Lite website was piloted on April 17, 2020 and monitored for 1 month (i.e., until May 14, 2020). Over that period, there were 290 unique visitors, 47 neural network predictions, and 730 total requests (Figure 5-9).

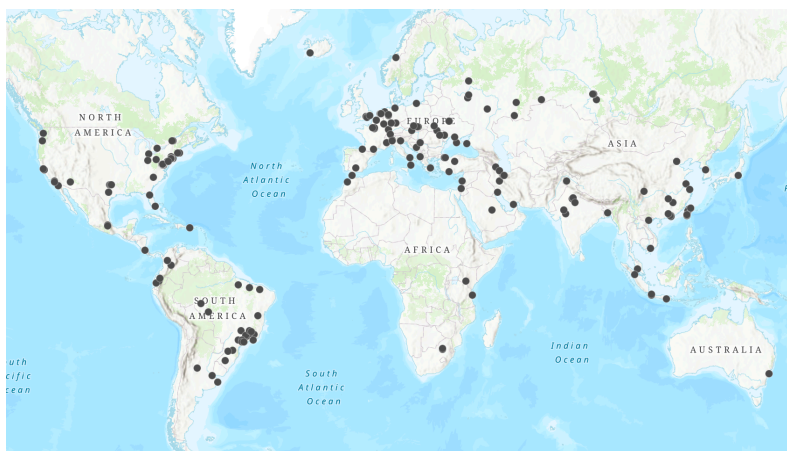


Figure 5-9. Locations of site accesses during first month of molSimplify Lite launch. Map was made with an online IP resolver⁴ and ArcGis.

Since then until March 2023, molSimplify Lite has received thousands of requests per month, amounting to approximately (for example, during the two weeks of March 3, 2023 to March 17, 2023, molSimplify Lite received on average 20-30 unique visitors per day, with about 100-150 total requests per day). Assuming an average of 50 requests per day and 20 unique visitors per day over a span of 3 years, the site has served approximately 22,000 visitors and 55,000 requests so far.

The MOFSimplify website was opened for public access upon publication on March 11, 2022. A year later, during the two weeks of March 3, 2023 to March 17, 2023, the website received 3,313 requests; assuming an average of 3,000 requests per month since launch, this amounts to about 36,000 requests so far. In addition, as of March 17, 2023, the website has received 18 feedback entries, indicating that users interacting with the site are able to provide valuable feedback on machine learning model predictions.

5.4 Conclusion

The development and implementation of design tools for transition metal complexes and metal-organic frameworks have significantly enhanced the ability of users to explore and provide feedback on machine learning models for property prediction. By offering an interactive platform for generating, analyzing, and comparing new structures, these tools bridge the gap between computational chemistry, machine learning, and experimental design, fostering a more collaborative and informed approach to transition metal complex and metal-organic framework discovery.

The molSimplify Lite and MOFSimplify web tools empower users to tailor TMC and MOF configurations and parameters, visualize their relationships with existing data,

and obtain valuable insights into their properties. Furthermore, the feedback component of MOFSimplify enables users to contribute valuable feedback, which can be used to refine and improve the underlying machine learning models in the future, leading to more accurate predictions and more efficient exploration of MOF design space.

Both tools have reached a high level of impact, serving hundreds of thousands of requests across the world. In the future, integration of these design tools into an autonomous high-throughput virtual screening workflow would not only facilitate the discovery of novel materials but also promotes a more comprehensive understanding of the landscape of transition metal complexes and metal-organic frameworks.

5.5 References

¹ <https://bootswatch.com/flatly/>. Accessed 2020.

² <https://3dmol.csb.pitt.edu/>

³ A. Nandy, G. Terrones, N. Arunachalam, C. Duan, D.W. Kastner, and H.J. Kulik, *Scientific Data* **9**, (2022).

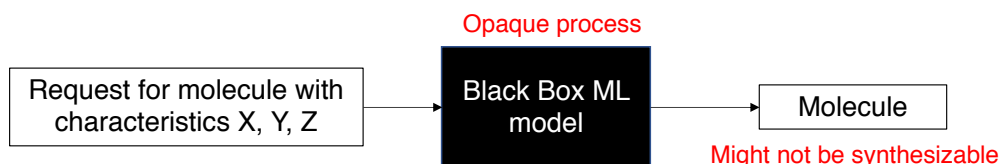
⁴ <https://www.bulkseotools.com/bulk-ip-to-location.php>

6 Ligand compatibility classification using stochastic negative addition: towards interactive investigation of new chemical structures

6.1 Introduction

Database availability for transition metal complexes has been growing rapidly; in 2019, the OHLDB (Octahedral Homoleptic Ligand Database) contained 1,250 structures and their properties. The next year, in 2020, the tmQM dataset contained 86,000 structures derived from the Cambridge Structural Database (CSD), with DFT-derived properties. As of 2021, our group initiated a database of >150,000 structures with DFT-derived properties. These efforts have enabled the development of machine learning models that can be used to design new chemicals; however, these models can be opaque to the synthetic chemist. Thus, a highly relevant question is how we can let experts interact with these models to perform inverse design.

The figure below demonstrates two paradigms for the use of machine learning models for chemical design. The top flow chart shows the workflow in which machine learning is used to perform 100% of the design tasks. The bottom flow chart shows a workflow in which user interaction and feedback can be incorporated into the black box ML model in order to increase transparency and integrate expert feedback. The goal of Ligandify is to introduce aspects of the latter paradigm in order to allow experts to participate in inverse design.



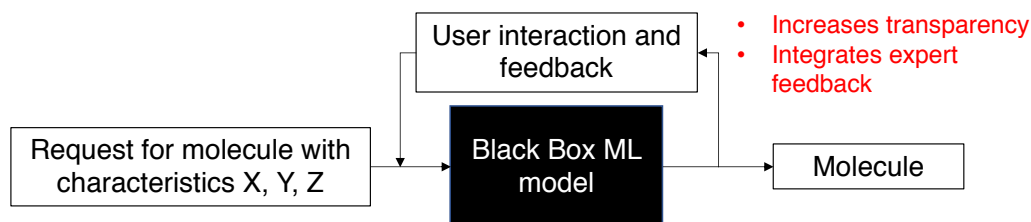


Figure 6-1. Paradigms for user interaction with machine learning models.

In Chapter 5, we developed web interfaces aligned with the latter paradigm, and in this chapter, we continue these efforts within the space of tetrahedral TMC catalysts.

6.1.1 Choice of design space

In previous work related to asymmetry, we found that the study of certain ligand combinations can be accelerated by forming homoleptic complexes out of them and deriving heteroleptic properties from homoleptic properties. However, there are some ligands that cannot be turned into homoleptic complexes, yet are relevant in real-world applications such as catalysis, as evidenced by their presence in the CSD. In the figure below, the pie chart on the left shows that approximately 13% of ligands found in octahedral complexes cannot form a homoleptic complex and need to be studied with coligands. The pie chart on the right shows that this figure is 10% for tetrahedral complexes.

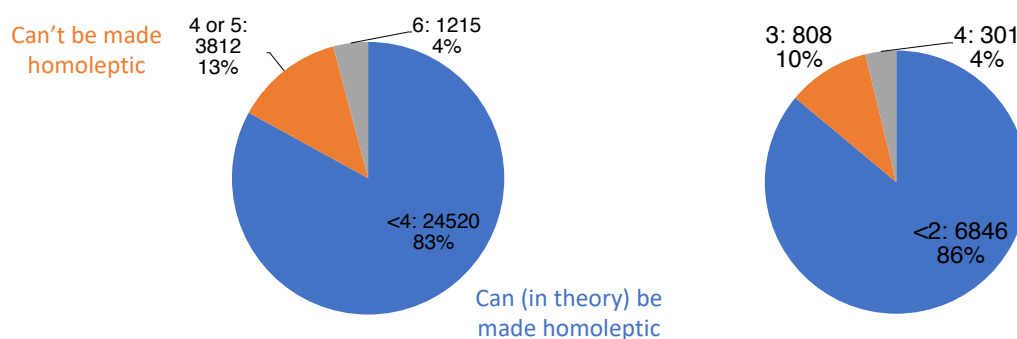


Figure 6-2. Distribution of ligand denticities found in the Cambridge Structural Database. The ligands considered are unique ligands found across all octahedral TMCs (left) and tetrahedral TMCs (right). The orange regions represent the body of ligands that require a co-ligand to be studied, but cannot be studied as part of a homoleptic complex.

In order for such ligands to be studied experimentally, one must be able to assess which co-ligands can be used to form a synthesizable complex. For this work, we chose to study a design space of transition metal catalysts from the Cambridge Structural Database (CSD).

As a first step towards identifying transition metal complexes that are likely to be catalysts, we did a preliminary study of which denticity combinations are likely to lead to catalytic behavior. In previous work, mononuclear complexes were extracted from the CSD. These complexes were counted, then recounted once filtered for the tag “catalysis” via the CrossRef API based on their DOIs. The table below shows how the distribution of ligand denticities changes when considering complexes with “catalysis” keywords.

Table 6-1. Percent changes in representation of octahedral complexes upon filtering for “catalysis” tag from CrossRef API

Complex Type (by number of unique ligands of a given type)	Percent change in representation
2+2+1+1	-38.3%
3+3	-22.7%
2+2+2	-17.2%
1+1+1+1+1+1	-8.9%
2+1+1+1+1	-8.9%
6	+16.7%
3+2+1	+33.4%
3+1+1+1	+35.4%
4+1+1	+40.0%
5+1	+57.2%

As seen in Table 6-1, 5+1 complexes increase in representation the most; this is likely because the monodentate ligand in a 5+1 complex can detach to form an active site. A similar analysis was carried out for tetrahedral complexes (Table 6-2); the results illustrate that 3+1 complexes experience some of the highest change in representation.

Table 6-2. Percent changes in representation of tetrahedral complexes upon filtering for “catalysis” tag from CrossRef API

Complex Type (by number of unique ligands of a given type)	Percent change in representation
2+2	-17.8%
1+1+1+1	-16.7%
2+1+1	+5.9%
3+1	+20.5%
4	+48.5%

The enrichment of 5+1 and 3+1 complexes is likely related to the concept of monodentate ligand dissociation; thus, (N-1)+1 complexes appear to be a good target for studying catalytic behavior. For this work, we chose to specifically focus on 3+1 complexes, although the methods used generalize to 5+1 complexes as well.

6.1.2 Classification via stochastic negative addition

For this work, we would like to provide a ligand to a recommendation engine and query what other ligands can be used to construct a complete, synthetically accessible complex.

In order to tackle the question of recommendation, it is important to actually define ligand co-occurrence, and to determine whether it is learnable. One main challenge faced in defining ligand co-occurrence is counting, since different complexes have different numbers of ligands. For the purpose of this work, co-occurrence was quantified as $p(\text{ligand } y \text{ is in TMC} \mid \text{ligand } x \text{ is in TMC})$, counted across all TMCs. However, it is important to note that the ideal distribution to learn is $p(\text{ligand } y \text{ can be added to vacancy in TMC} \mid \text{TMC with vacancy})$. The difference between the former and latter distributions is that the former overcounts TMCs with at least two ligand types,

one of which appears more than once. However, for TMCs containing only two distinct ligands, the distributions are identical.

Any machine learning model intending to learn these distributions for TMCs should work in a chemical fingerprint space (e.g. RAC-space), so that ligands outside the training set (i.e. ligands not found in the CSD) can be used as inputs to the model.

Since the CSD only contains positive examples of synthesizable TMCs, we can use stochastic negative addition to generate negative examples. To do this, we look at unique TMCs containing two distinct ligand types (e.g. a tridentate ligand and a monodentate ligand), and separate the ligand types into individual pools. Then, ligands from each pool are randomly paired with each other, and if the resulting complex is not found in the CSD, the example is marked as a negative. The number of possible combinations is vast compared to the number of complexes present in the CSD, and it is assumed that the majority of these combinations will not lead to a synthesizable complex. The quality of this assumption dictates the degree to which synthesizability is learnable; if the assumption does not hold, then there will be no learnable difference between the negative and positive examples.

In order to determine whether co-occurrence in the CSD is learnable based on RACs, one can use a ROC-AUC score for a simple classifier for whether a co-occurrence exists in the CSD based on RAC vectors; getting one >0.6 suggests that co-occurrence is a learnable quantity and that the assumptions of stochastic negative addition hold. A lower AUC score would suggest that the data is not sufficiently structured to yield learning.

6.2 Methods

6.2.1 Cambridge Structural Database search

We curated a set of mononuclear octahedral transition metal complexes (TMCs) from the Cambridge Structural Database¹ (CSD) to gain statistics on ligand conformations and symmetries observed in experimentally characterized complexes. We performed an initial search using the Conquest graphical interface to the CSD and Python application programming interface (API), applied to the November 2019 v5.41 data set with March 2020, May 2020, and September 2020 updates. We used the Conquest interface to query for X-ray crystallographic structures containing a transition metal atom from groups 3-12 and periods 4-6, excluding lanthanides. We used the data from an initial query required that the metal forms at least four bonds with either *p*-

block (i.e., the first five rows of groups 13 – 17), noble gas, or group 1 or 2 elements. This initial query produced 462,012 reftcodes for unique structures. In a later step, we then selected the subset that contain only four bonds between the metal and surrounding ligands. The CSD Python API was then used to iterate through every component molecule in each X-ray crystallographic structure to identify all non-polymeric molecules with distinct molecular weight (i.e., within the same crystal cell) and only one transition metal. For each of the resulting 240,117 mononuclear TMCs, the CSD mol2 files were saved for after adding any missing hydrogen atoms, preserving both the connectivity and bond orders present in the CSD. From this set, 75,383 four-coordinate metal complexes were evaluated for deviation of the angles² between the metal-connecting atoms from the valence shell electron pair repulsion (VSEPR) ideal angles for tetrahedral structures to eliminate structures that were square planar. Structures with lower angular deviation from the ideal tetrahedral molecular structure as compared to other ideal geometries were identified as tetrahedral, resulting in 21,881 tetrahedral complexes.

Across the set of CSD-mined tetrahedral complexes, ligands from each structure were isolated by removing the metal and identifying the ligands as disconnected sub-connectivity graphs. For each ligand, a dummy metal atom with identical connectivity to the removed metal with an atomic number of 0 was introduced to preserve the connectivity of the ligands to the metal without preserving metal identity. For each ligand with an added dummy atom, the previously reported atomic-number weighted connectivity matrix determinant was calculated.³ Additionally, the determinant of the atomic-number weighted bond-order weighted connectivity matrix was calculated for each ligand, where off-diagonal elements such as $Z_i Z_j$ ($i \neq j$) were set to the bond order reported in the CSD. Any ligand with a unique atomic-number weighted connectivity matrix determinant and bond-order weighted connectivity matrix determinant was tabulated as a distinct ligand in the set of all ligands across monometallic TMCs in the CSD. Using this procedure, 4,301 unique tridentate ligands and 1,046 unique monodentate ligands were identified. Additionally, for each distinct ligand, the list of CSD reftcodes for TMCs containing the ligand was tabulated.

6.2.2 Descriptors

For each ligand, we employed Moreau-Broto autocorrelations (ACs)⁴ after replacing the dummy metal atom with an Fe atom. Moreau-Broto ACs are representations formed by discrete operations on the heuristic properties of atoms in the

molecular graph in analogy to continuous ACs. An AC is evaluated over all the atoms in a molecule that are d bonds apart,

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

where P is the chosen property, and P_i and P_j are the specific values of the i^{th} and j^{th} atoms, respectively, that are d_{ij} bonds apart. The heuristic properties used were Pauling electronegativity, nuclear charge, identity (i.e., 1), topology, and size, which produced a $5d+5$ dimensional vector. For each ligand, a cutoff of $d=4$ and full scope (i.e., all atoms are used as starting atoms) were used to produce a length-25 feature vector.

6.2.3 Machine learning

To generate training data for a synthetic accessibility classifier, we utilized complexes sampled from the CSD as positive examples of synthetically accessible complexes and used stochastic negative addition⁵ to generate negative examples. Stochastic negative addition refers to the generation of false examples to aid in balancing the classes in the classification task. Positive examples were labeled as 1 and negative examples were labeled as 0. To choose complexes found in the CSD, a tridentate a tridentate ligand along with a compatible monodentate co-ligand corresponding that tridentate ligand were drawn at random. To generate negative examples, a tridentate ligand and a monodentate ligand were drawn at random, and if they were found to be paired together in the CSD, the pairing was rejected and the monodentate ligand was replaced with a randomly drawn monodentate ligand until the pairing could not be found in the CSD. Using this procedure, 5,000 positive and 5,000 negative examples were generated. For model training, we employed an 80-20 train-test split. For each example true or false ligand pair, the length-25 AC vectors of the tridentate and monodentate ligands were concatenated. This data was then z-score normalized and used as input to a fully-connected artificial neural network (ANN) with two hidden layers of 100 nodes each with rectified linear unit (ReLU) activation and an output layer of one node with sigmoid activation. The Adam optimizer⁶ with binary cross entropy loss was used to train the classifier for two epochs, resulting in a classifier that outputs a value between 0 and 1, signifying the similarity of a ligand pair to tetrahedral pairs found in the CSD.

6.3 Results and discussion

6.3.1 Indicators of ligand compatibility

To illustrate how steric properties differ between true and false pairs of ligands, the molecular weight distribution within 3+1 tetrahedral (non-square planar) complexes is shown below.

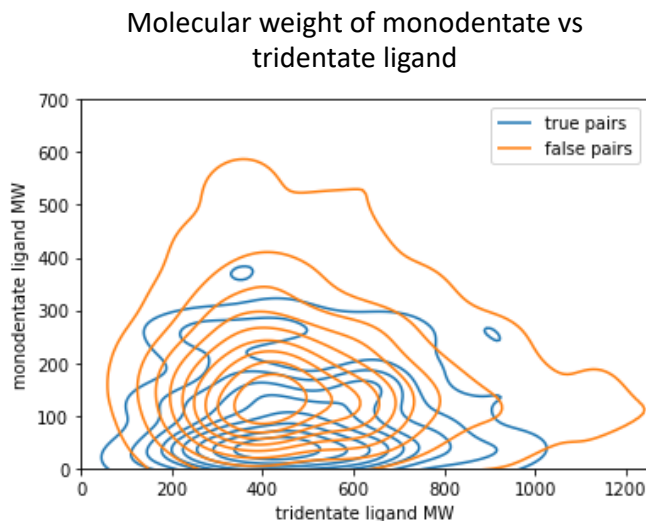


Figure 6-3. Molecular weight of true and false pairs of monodentate and tridentate ligands.

The figure above indicates that true pairs have smaller monodentate ligands than false pairs. Additionally, it indicates that true pairs contain a narrower range of tridentate and monodentate ligands. The conclusion that can be drawn from this figure is that ligand compatibility can be reliably correlated to molecular weight.

Below, more property correlations are illustrated for 3+1 tetrahedral complexes:

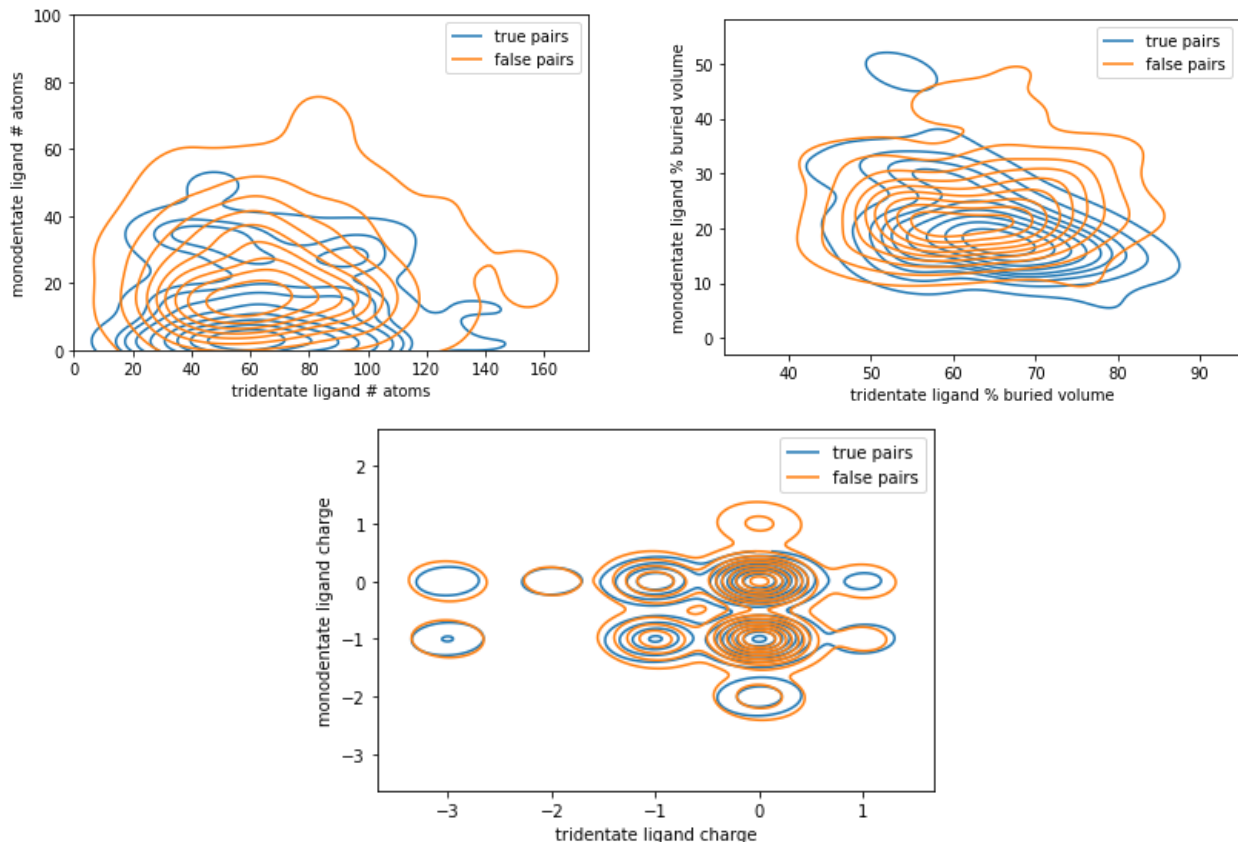


Figure 6-4. Comparison of property distributions (top left, number of atoms; top right, percent buried volume; bottom, ligand charge) for monodentate and tridentate ligands in true and false pairs.

These figures illustrate that the number of atoms and percent buried volume provide some information about ligand compatibility. However, it appears that chemical charge is not as informative of a factor for ligand compatibility as steric properties. (The false dataset had more (0,+1) charge pairs than the true dataset, but other than that, the distributions look similar.)

To test whether stochastic negative addition results in chemically distinguishable complexes, a subset of 50 true and 50 false pairs was taken from the training data set and assembled into tetrahedral complexes using molSimplify. Using previously reported methods, a spin-splitting energy was calculated using a metal center of Fe(II) for each pair. The results are shown in Figure 6-5.

The broader range of spin-splitting energies for false pairs indicates that false pairs are more likely to face convergence issues and/or have differing electronic properties from true pairs. This result demonstrates that true pairs have more tightly clustered electronic properties than false pairs, indicating that stochastic negative

addition can indeed result in negative training data with distinct chemical properties from positive training data.

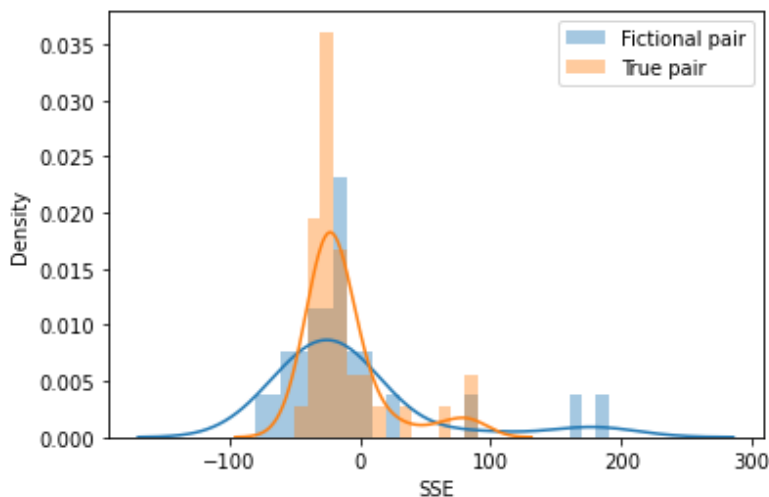


Figure 6-5. Spin splitting energy of true and fictional pairs of ligands

6.3.2 Model evaluation

We started from a set of 5,000 positive and negative examples of monodentate-tridentate ligand pairings that form known tetrahedral structures regardless of metal center. We evaluated the performance of the trained synthetic accessibility ANN classifier on the set-aside test data containing 1,000 positive and 1,000 negative examples, resulting in an area under the curve of the receiver operating characteristic (i.e., ROC-AUC score) of 0.87 and an overall accuracy of 78% at a classifier threshold of 0.5 (Figure 1). Adjusting the threshold to YY to avoid false positives would achieve an accuracy of XX. The high ROC-AUC score can be explained by strong distinguishability between positive and negative examples in feature space (Figure 6-6). An unsupervised Uniform Manifold Approximation and Projection (UMAP) of the 10,000 generated examples demonstrates that positive examples and negative examples tend to lie in distinct regions of feature space, leading to a clear decision boundary that can be learned by the ANN.

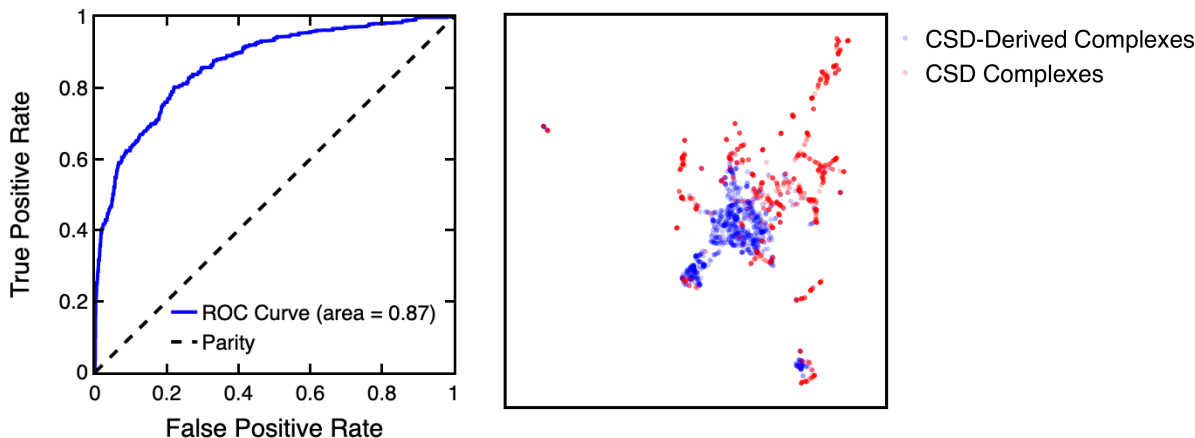


Figure 6-6. ROC curve for the trained ANN classifier (left) and the unsupervised UMAP representation of positive examples and negative examples in red and blue, respectively (right). The UMAP indicates a clear separation between the two groups, illustrating that stochastic negative addition creates negative examples that are distinguishable from positive examples in feature space. The high ROC-AUC of 0.87 illustrates that the ANN can learn the decision boundary between the positive and negative examples robustly and with high accuracy.

Based on Figure 6-6, it is clear that some “true” and “fictional” pairs are very obviously separable, while some are not. Based on the stochastic negative addition protocol used, the obviously separable regions indicate that many “fictional” pairs are obviously incompatible, leading them to be placed very far away in RAC space compared to true pairs. However, the regions that are not obviously separable indicate that some “fictional” pairs may indeed be valid pairs of ligands that would form a synthetically accessible complex, but the pair has not yet been reported in the CSD.

6.3.3 Usage notes

We introduce the Ligandify website, a tool for identifying ligands that are likely to form a synthetically accessible tetrahedral transition metal complex with a user-provided ligand. Ligandify requires that the user specify a monodentate ligand as a SMILES string that contains a placeholder metal atom representing the transition metal complex center and where it is connected to the ligand. Ligandify then provides a 3D visualization of the provided ligand. When the user requests Ligandify find candidate co-ligands, the 25-length AC feature vectors are generated for the provided ligand and

appended to each tridentate ligand AC feature vector, for all 4,301 tridentate ligands in the tetrahedral complex data set. The combined ACs are then normalized using the same scaling parameters used to z-score normalize the training data for the synthetic accessibility ANN. Ligandify then inputs the combined ACs to the synthetic accessibility ANN, which outputs a number between 0 and 1 for each tridentate ligand that indicates how similar a tetrahedral complex containing both the tridentate ligand and the provided monodentate ligand would be to complexes found in the CSD.

Once probabilities are generated for every tridentate ligand, the top five tridentate candidates for the provided monodentate ligand are selected and visualized on the website in 3D, enabling the user to compare the different options and select the most suitable candidate for their specific application. Metadata for each ligand, including connecting atom identities, a mol2 file, metal centers associated with the ligand, CSD reference codes containing the ligand, RACs, canonical SMILES string, Crippen LogP, and percent buried volume, is also generated and is downloadable in JSON format.

As an alternative to using the Ligandify web interface, users may also download the trained synthetic accessibility ANN model, ligand data, and Python backend code from the Ligandify GitHub repository. Our ANN model is provided in Pickle archive format and can be loaded directly by using scikit-learn version 0.23.2 and keras⁷ version 2.4.3 with a tensorflow⁸ backend in Python version 3.7.9. In addition to the ANN pickle, the repository contains a pickle archive of a Pandas dataframe containing the metadata for the monodentate and tridentate ligands found in the CSD. Users may also download the open-source Python code used by Ligandify for generating predictions with our model.

Ligandify (powered by molSimplify from the Kulik group at MIT)

Find compatible ligands using this tool

Enter a SMILES string for a ligand (including the metal as [X]), and click "Generate" to view the ligand. Then, press "Find Candidates" to find co-ligand candidates for the provided ligand.

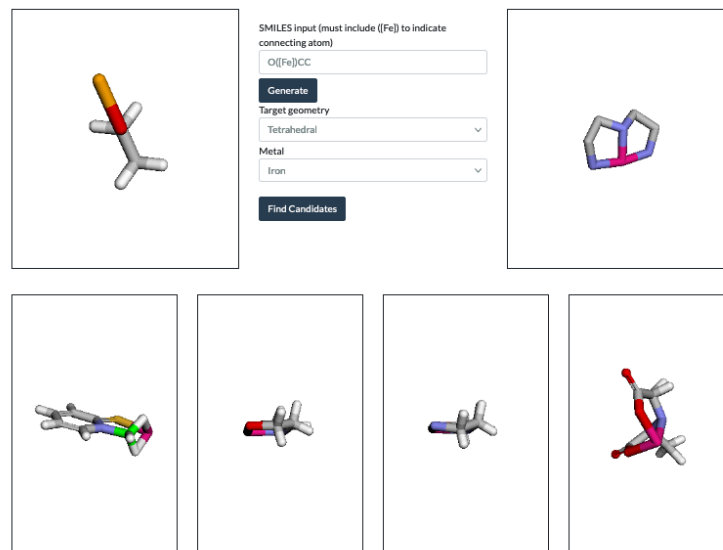


Figure 6-7. The Ligandify web interface. The user provides a SMILES string of a monodentate ligand including a placeholder metal atom, and Ligandify generates a 3D structure of the specified ligand. The “Find Candidates” button is used to find the top 5 tridentate co-ligand candidates for forming a tetrahedral complex with the provided monodentate ligand.

6.4 Conclusion

When distinguishing compatible ligands from incompatible ligands using individual physical and chemical properties rather than an ML classifier, physical properties such as ligand size and percent buried volume are key indicators of ligand compatibility. However, chemical properties such as charge appear to be similar between true and false pairs.

In this work, we demonstrate the use of an ML classifier to convert the physical properties of a molecule, as represented by a graph autocorrelation vector, into a prediction for likelihood of synthesizability.

One limitation of the above approach is that certain types of ligands are overrepresented. For example, functionalized porphyrins can be very similar in chemical space, yet be counted as distinct ligands. Hypothetically, a classifier model trained on a database containing a small number of unique ligands in a large number of

functionalized porphyrins would be excellent at determining compatibility among porphyrins but would exhibit poor performance when it comes to other types of ligands.

To allow users to interact with the ML classifier, we present a web interface where users can upload a ligand of interest and retrieve a recommendation of the top 5 co-ligand candidates for that ligand. This interface allows users to explore the output of the classifier as a starting point for investigating the properties of new complexes and ligands.

6.5 References

- ¹ Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallographica Section B Structural Science, Crystal Engineering and Materials* **2016**, *72* (2), 171-179. DOI: 10.1107/s2052520616003954.
- ² Duan, C.; Janet, J. P.; Liu, F.; Nandy, A.; Kulik, H. J. Learning from Failure: Predicting Electronic Structure Calculation Outcomes with Machine Learning Models. *Journal of Chemical Theory and Computation* **2019**, *15* (4), 2331-2345.
- ³ Taylor, M. G.; Yang, T.; Lin, S.; Nandy, A.; Janet, J. P.; Duan, C.; Kulik, H. J. Seeing Is Believing: Experimental Spin States from Machine Learning Model Structure Predictions. *The Journal of Physical Chemistry A* **2020**, *124* (16), 3286-3299. DOI: 10.1021/acs.jpca.0c01458.
- ⁴ Broto, P.; Moreau, G.; Vanduycke, 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; Vol. 19, pp 71-78.
- ⁵ Cáceres, E. L.; Mew, N. C.; Keiser, M. J. Adding Stochastic Negative Examples into Machine Learning Improves Molecular Bioactivity Prediction. *Journal of Chemical Information and Modeling* **2020**, *60* (12), 5957-5970. DOI: 10.1021/acs.jcim.0c00565.
- ⁶ Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. arXiv, 2014.
- ⁷ Keras; 2015. <https://keras.io> (accessed 3/12/2020).
- ⁸ TensorFlow: Large-scale machine learning on heterogeneous systems; 2015. Software available from tensorflow.org. (accessed 3/12/2020).

7 Recommendations for future work

Overall, this work can be extended in three main directions: (1) developing a fully autonomous workflow to make efficient use of computational resources, (2) incorporating feedback from web tools into machine learning models, and (3) developing expert-guided exploration of new chemical spaces. These recommendations point towards a vision of an autonomous high-throughput virtual screening workflow that can be steered by experts and practitioners to fruitfully explore chemical space.

7.1 Development and maintenance of fully autonomous workflows

One key limitation to machine learning models is the availability of high-quality training data. In this work, we have addressed this issue by relying on high-throughput DFT calculations to generate initial training data. Despite this approach's effectiveness, the high computational cost of DFT calculations places fundamental limits on the size of the training dataset and the regions of chemical space that can be explored.

To overcome these limitations, it is essential to develop fully autonomous workflows that enable the identification of high-uncertainty regions in chemical space and generate additional training data in those regions. One approach to identifying high-uncertainty regions is uncertainty quantification (UQ), which involves quantifying the uncertainty in the ML model's predictions. By leveraging UQ methods, it is possible to identify regions of chemical space where the model's predictions are least confident, indicating a lack of training data in those regions.

Once high-uncertainty regions are identified, a fully-autonomous workflow can use active learning strategies to generate new training data, which can then be used to retrain the ML model. These active learning strategies should select new data points that are most informative for reducing the model's uncertainty in high-uncertainty regions. Through this iterative process, the ML model can gradually reduce uncertainty in previously neglected regions of chemical space.

7.2 Incorporation of web-based feedback into machine learning models

In this work, we have developed web-based tools that have democratized access to our machine learning models for predicting the properties of transition metal complexes and MOFs. These tools allow users to interact with ML models and – in the case of MOFs, provide feedback on model predictions. Ideally, this feedback can then be used to improve our models’ accuracy and reliability.

One promising approach to leveraging web-based feedback is to incorporate it directly into the ML model's training process. Specifically, user feedback can be used to augment the training dataset with additional examples of challenging cases, thereby improving the model's ability to generalize to new data. For example, in the context of predicting the solvent-removal and thermal stability of MOFs, user feedback could be used to identify MOFs for which the model's predictions are particularly inaccurate, and additional data points could be generated for those MOFs to improve the model's accuracy in those regions of the chemical space.

Another approach to leveraging web-based feedback is to use it to refine the model's predictions directly and present tailored results to users. For example, in the context of predicting the density functional approximation for a given molecule, users could be prompted to provide feedback on the predicted functional's accuracy, and this feedback could be used to re-weight the predictions from different density functionals based on their relative performance for similar molecules.

7.3 Expert-informed chemical space exploration

Machine learning models such as the ones mentioned in this work can benefit from expert feedback. By leveraging the insights and knowledge of experts in the field, it is possible to guide the exploration of chemical space and focus on regions that are most relevant to specific applications.

One way to incorporate expert knowledge into ML models is through active learning. Specifically, experts can provide feedback on the model's predictions for specific molecules or applications, and this feedback can be used to guide the selection of new training data points. If this active learning workflow is implemented in a fully-autonomous way, experts can help steer the overall direction of DFT data generation by identifying regions of chemical space where the machine learning models are inaccurate

or unreliable. This way, an autonomous workflow can be biased to prioritize accuracy in areas of interest to experts, while also exploring under-explored regions of chemical space.

A. Appendix: Ligand Additivity

Text A-1. Description of mononuclear complex curation.

The Conquest interface was used to query for structures containing a transition metal atom (here, groups 3–12, periods 4–6, excluding La) that forms at least 4 bonds with either p-block (here, the first five rows of groups 13–17), noble gas, or group 1 or 2 elements. The only constraint imposed on the search was that 3D coordinates were determined. This initial query produced 462,012 Refcodes for unique structures. The CSD Python API was then used to iterate through every “component” molecule in each x-ray crystallographic structure to identify all non-polymeric molecules with distinct molecular weight (within the same crystal structure) and only one transition metal atom. For each of the resulting 240,117 mononuclear transition metal complexes the CSD mol2 files were saved after adding any missing hydrogen atoms, preserving both the connectivity and bond orders present in the CSD. From this diverse set, 6-coordinate metal centers without edge-bound or sandwich-type ligands were evaluated for deviation of the angles (as defined in Ref. 1) between the metal-connecting atoms from the VSEPR ideal angles for octahedral, pentagonal pyramidal, or trigonal prismatic structures. Structures with lowest angular deviation from ideal octahedral molecular structure over other geometries were identified as octahedral.

Text A-2. Description of complex symmetry identification.

Within each complex, a 3-hydrogen difference criterion was introduced to heuristically capture ligands that only differed in protonation state. If none of the equivalence conditions were met the ligands were identified as chemically identical. Ligand symmetry was identified by binning the octahedral structures by: (1) Ligand denticities (e.g., 5-denticity + 1-denticity) (2) maximum number of connecting atoms from a single ligand in the equatorial plane (e.g., 3-denticity + 3-denticity has at most 2 connecting atoms from the same ligand in the equatorial plane -> *fac* binding) (3) total number of unique ligands (e.g., for a monodentate-only complex with 5 unique ligands, two must be duplicates) and (4) for lower denticities, counts of pairs of *trans* ligands that are identical (e.g., for a monodentate-only complex with 3 pairs of identical ligands, if only 1 pair of *trans* ligands are identical -> other unique pairs must be *cis*). A nomenclature for ligand symmetry was created to describe the observed symmetries

following these rules (1) Denticities are denoted by numbers as the base of the ligand symmetry representation with higher denticities listed first. (2) If ligands are identical they are surrounded by "||" markers, with ligands with the highest number of identical copies of the same denticity listed first. (3) Tridentate ligands are identified as either *fac*- or *mer*- by f/m labels, and corresponding sets of three identical monodentate ligands are given equivalent f/m labels. (4) Pairs of identical monodentate ligands are identified as either *cis*- or *trans*- with c/t labels, pairs of non-identical monodentate ligands are given c/t labels only when there is one pair of non-identical monodentate ligands, and the ligand symmetry of identical pairs is ambiguous. Examples of ligand symmetry using this nomenclature are “|111111|” for a homoleptic monodentate-only complex, “4|11|t” for an equatorial tetradentate with *trans* identical monodentate ligands, and “2|11|c11” for a complex with a bidentate and 2 identical monodentate ligands in the equatorial plane with 2 chemically-distinct monodentate ligands in the axial positions.

Table A-1. Number of complexes grouped by number of unique ligands for all (all mononuclear octahedral transition metal complexes, all unique complexes with user-defined charges and nondisordered structures ("computation-ready"), and the subset of unique complexes with Fe(II) centers (Fe(II))). The percent of each category is also shown.

# ligands	all	percent	computationready	percent	Fe(II)	percent
1	19324	22.6%	4830	28.3%	661	55.0%
2	38780	45.3%	8191	47.9%	407	33.8%
3	22655	26.5%	3701	21.7%	129	10.7%
4	4530	5.3%	350	2.1%	5	0.4%
5	283	0.3%	13	0.1%	0	0.0%
6	3	0.0%	0	0.0%	0	0.0%
total	85575		17085		1202	

Text A-3. Poly enumeration discussion.

The cycle index for all possible octahedral complexes is

$$Z(O_h) = \frac{1}{48} (f_1^6 + 8f_3^2 + 7f_2^3 + 6f_1^2 f_4 + 9f_1^2 f_2^2 + 8f_6 + 3f_1^4 f_2 + 6f_4 f_2)$$

where the subscript, j , on f indicates the number of unique sites (e.g., 2 for $M(L_1)_5(L_2)_1$) up to a theoretical maximum of 6. Each individual term in the cycle index is computed as:

$$f_j = x_1^j + \dots + x_N^j$$

where N is the total number of unique ligands that can occupy the sites, and each x_i corresponds to one unique ligand set choice.

For a worked example of $N = 12$ unique ligands, we determine the theoretical number of complexes, the cardinality, as

$$\frac{N!}{(N-M) \prod_{k=1}^6 \left(\sum_{j=1}^M \mathbb{I}_{|L_j|=k} \right)!}$$

where its respective symmetry class $C = \{L_1, \dots, L_M\}$ is built from M distinguishable sites, j , and N are the ligand index and number of ligands as before. The indicator function, $\mathbb{I}_{|L_j|=k}$, where e.g., $|L_1|$ would be 6 for $(L_1)_6$, counts how many occurrences of each substituent combination exist.

Table A-2. Number of complexes grouped by symmetry class for all mononuclear octahedral transition metal complexes, all complexes with user-defined charges and non-disordered structures ("computation-ready"), and the subset of complexes with Fe(II) centers. The percent of each category is also shown.

Symmetry class	All	percent	computation-ready	percent	Fe(II)	percent
HO	19324	24%	4830	29%	661	55%
5+1	4881	6%	1216	7%	128	11%
CS	15138	19%	2447	15%	74	6%
TS	14861	18%	3745	22%	151	13%
FS	2184	3%	426	3%	36	3%
MS	1716	2%	357	2%	18	2%
CA	2571	3%	793	5%	26	2%
TA	3691	5%	760	5%	41	3%
FA	5828	7%	772	5%	21	2%

MAC	4706	6%	826	5%	17	1%
MAT	1821	2%	205	1%	14	1%
DCS	802	1%	113	1%	2	0%
DTS	1011	1%	96	1%	4	0%
EA	2225	3%	136	1%	4	0%
Total	80759		16722		1197	

Table A-3. Number of unique ligands of each denticity grouped by symmetry class for all complexes with user-defined charges and non-disordered structures.

	# complexes	# unique ligands by denticity						
		1	2	3	4	5	6	total
HO	4830	150	528	1719	0	0	813	3210
5+1	1216	242	0	0	0	507	0	749
TS	3745	677	786	0	700	0	0	2163
CS	2447	109	1342	0	295	0	0	1746
FS	426	48	0	275	0	0	0	323
MS	357	35	0	318	0	0	0	353
CA	793	302	119	0	206	0	0	627
DCS	113	27	105	0	0	0	0	132
TA	760	321	66	165	0	0	0	552
DTS	96	111	0	0	0	0	0	111
EA	136	66	118	0	0	0	0	184
FA	772	225	226	168	0	0	0	619
MAC	826	195	266	252	0	0	0	713
MAT	205	98	1	121	0	0	0	220

Table A-4. Number of unique ligands of each denticity grouped by symmetry class for all complexes with user-defined Fe(II) centers.

	# complexes	# unique ligands by denticity						
		1	2	3	4	5	6	total
HO	661	40	142	355	0	0	124	661
5+1	128	29	0	0	0	81	0	110

TS	151	53	71	0	49	0	0	173
CS	74	18	27	0	44	0	0	89
FS	36	11	0	24	0	0	0	35
MS	18	6	0	22	0	0	0	28
CA	26	18	6	0	17	0	0	41
DCS	2	3	3	0	0	0	0	6
TA	41	34	7	0	10	0	0	51
DTS	4	8	0	0	0	0	0	8
EA	4	5	3	0	0	0	0	8
FA	21	16	9	14	0	0	0	39
MAC	17	14	6	9	0	0	0	29
MAT	14	13	0	11	0	0	0	24

Table A-5. Number of monodentate ligands of each category from Fe(II) monodentate-only structures for the case where user defined charges and trustworthy structures were obtained ("All computation-ready") as well as the more stringent test where ligand charges were assigned as neutral. The bolded categories correspond to those used to study HO Fe(II) complexes as described in the main text.

Category	All computation-ready	neutral ligand charge
HO	40	36
2, or 3 type but not HO	48	21
Total 1, 2, or 3 types	88	57
only in 4-5-6-types	3	1

Table A-6. Number of complexes from monodentate-only configurations for each symmetry from the total user defined charges and trustworthy structures ("computation-ready") or for Fe(II) only. For Fe(II) complexes, theoretical spaces of $N = 56$ and $N = 88$ ligands are compared for each symmetry class along with the % of that space present in the CSD. The relevant symmetry class and notation is also provided.

					$N = 88$		$N = 56$		
	Computation-ready	Percent	Fe(II)	Percent	#	%	#	%	Symmetry class
HO	150	7%	40	39%	88	45.45%	56	71.43%	111111
5+1	330	16%	9	9%	7656	0.12%	3080	0.29%	111111 1
TS	944	46%	26	25%	7656	0.34%	3080	0.84%	1111 11 t

CS	92	5%	5	5%	7656	0.07%	3080	0.16%	1111 11 c
FS	48	2%	5	5%	3828	0.13%	1540	0.32%	111 f 111
MS	23	1%	0	0%	3828	0.00%	1540	0.00%	111 m 111
CA	46	2%	0	0%	329208	0.00%	83160	0.00%	1111 11c
DCS	0	0%	0	0%	109736	0.00%	27720	0.00%	11 c 11 c 11 c
TA	183	9%	3	3%	329208	0.00%	83160	0.00%	1111 11t
DTS	96	5%	4	4%	109736	0.00%	27720	0.01%	11 t 11 t 11 t
EA	17	1%	1	1%	329208	0.00%	83160	0.00%	11 c 11 c 11 t
FA	36	2%	4	4%	658416	0.00%	166,320	0.00%	111 f 11 1
MAC	35	2%	3	3%	658416	0.00%	166,320	0.00%	111 m 11 c1
MAT	39	2%	2	2%	658416	0.00%	166,320	0.00%	111 m 11 t1
Total	2039		102		3213056	0.00%	816,256	0.02%	

Table A-7. Homoleptic Fe(II) transition metal complex properties, DE_{H-L} (in kcal/mol) and singlet HOMO level (in eV) for the ligands NH_3 , CH_3CN , CO , and H_2O obtained with B3LYP/LACVP* calculations.

Ligand	DE_{H-L} (kcal/mol)	HOMO level (eV)
CH_3CN	43.67	-13.35
CO	30.34	-19.46
H_2O	-27.04	-15.16
NH_3	-9.64	-13.58

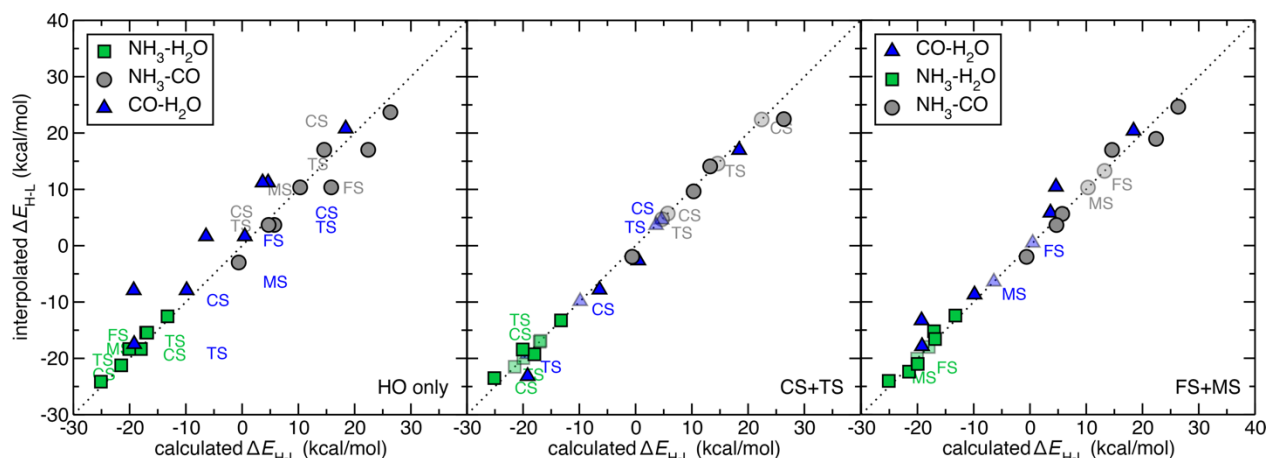


Figure A-1. Calculated vs. linearly interpolated ΔE_{H-L} (kcal/mol) for Fe(II) complexes with pairs of any of the three ligands: NH_3 , H_2O , and CO . From left to right: interpolation between homoleptic complexes (HO only), interpolation using homoleptic complexes as well as *cis* symmetric and *trans* symmetric complex energies (CS+TS), or interpolation using homoleptic complexes as well as *fac* and *mer* symmetric complex energies (FS+MS). Points are colored according to the pair of ligands they correspond

to: $\text{NH}_3\text{-H}_2\text{O}$ (green squares), $\text{NH}_3\text{-CO}$ (gray circles), and $\text{CO-H}_2\text{O}$ (blue triangles), as indicated in inset legend. Key isomers are annotated. Points provided for the fit are translucent, whereas the remaining points are opaque. In all panes, a black dotted parity line is shown.

Table A-8. ΔE_{H-L} values for pairs of two ligand types in Fe(II) complexes (in kcal/mol) along with interpolated values using only HO complexes, HO+CS/TS complexes, or HO+FS/MS complexes.

L ₁	L ₂	Symmetry	Number of ligands				splitting	interpolated		
			CO	H ₂ O	CH ₃ CN	NH ₃		HO	CS/TS	FS/MS
H ₂ O	CH ₃ CN	5+1	0	5	1	0	-18.4	-15.3	-19.9	-16.2
H ₂ O	CH ₃ CN	CS	0	4	2	0	-7.8	-3.5	-7.8	-5.4
H ₂ O	CH ₃ CN	TS	0	4	2	0	-12.8	-3.5	-12.8	-9.6
H ₂ O	CH ₃ CN	FS	0	3	3	0	5.4	8.3	2.8	5.4
H ₂ O	CH ₃ CN	MS	0	3	3	0	-0.9	8.3	-1.2	-0.9
CH ₃ CN	H ₂ O	CS	0	2	4	0	13.4	20.1	13.4	18.2
CH ₃ CN	H ₂ O	TS	0	2	4	0	10.5	20.1	10.5	14.0
CH ₃ CN	H ₂ O	5+1	0	1	5	0	26.3	31.9	27.1	30.9
CH ₃ CN	CO	5+1	1	0	5	0	42.6	41.4	42.5	41.9
CH ₃ CN	CO	CS	2	0	4	0	39.2	39.2	39.2	40.1
CH ₃ CN	CO	TS	2	0	4	0	41.2	39.2	41.2	40.0
CO	CH ₃ CN	FS	3	0	3	0	38.4	37.0	39.0	38.4
CO	CH ₃ CN	MS	3	0	3	0	38.2	37.0	40.2	38.2
CO	CH ₃ CN	CS	4	0	2	0	38.8	34.8	38.8	35.7
CO	CH ₃ CN	TS	4	0	2	0	39.1	34.8	39.1	35.6
CO	CH ₃ CN	5+1	5	0	1	0	35.0	32.6	34.7	33.0
H ₂ O	CO	5+1	1	5	0	0	-19.2	-17.5	-23.2	-17.9
H ₂ O	CO	CS	2	4	0	0	-9.9	-7.9	-9.9	-8.7
H ₂ O	CO	TS	2	4	0	0	-19.3	-7.9	-19.3	-13.3
H ₂ O	CO	FS	3	3	0	0	0.5	1.6	-2.6	0.5
H ₂ O	CO	MS	3	3	0	0	-6.4	1.6	-7.8	-6.4
CO	H ₂ O	CS	4	2	0	0	4.6	11.2	4.6	10.4
CO	H ₂ O	TS	4	2	0	0	3.6	11.2	3.6	5.8
CO	H ₂ O	5+1	5	1	0	0	18.4	20.8	17.0	20.4
NH ₃	H ₂ O	5+1	0	1	0	5	-13.3	-12.5	-13.3	-12.4
NH ₃	H ₂ O	CS	0	2	0	4	-17.1	-15.4	-17.1	-15.2
NH ₃	H ₂ O	TS	0	2	0	4	-16.9	-15.4	-16.9	-16.6
H ₂ O	NH ₃	FS	0	3	0	3	-18.0	-18.3	-19.3	-18.0
H ₂ O	NH ₃	MS	0	3	0	3	-20.1	-18.3	-18.4	-20.1
H ₂ O	NH ₃	CS	0	4	0	2	-21.5	-21.2	-21.5	-22.4
H ₂ O	NH ₃	TS	0	4	0	2	-20.0	-21.2	-20.0	-21.0
H ₂ O	NH ₃	5+1	0	5	0	1	-25.1	-24.1	-23.5	-24.0
NH ₃	CO	5+1	1	0	0	5	-0.6	-3.0	-2.5	-2.0

NH ₃	CO	CS	2	0	0	4	5.7	3.7	5.7	5.6
NH ₃	CO	TS	2	0	0	4	4.7	3.7	4.7	3.7
CO	NH ₃	FS	3	0	0	3	13.3	10.4	14.1	13.3
CO	NH ₃	MS	3	0	0	3	10.3	10.4	9.7	10.3
CO	NH ₃	CS	4	0	0	2	22.4	17.0	22.4	19.0
CO	NH ₃	TS	4	0	0	2	14.6	17.0	14.6	17.0
CO	NH ₃	5+1	5	0	0	1	26.4	23.7	22.5	24.6

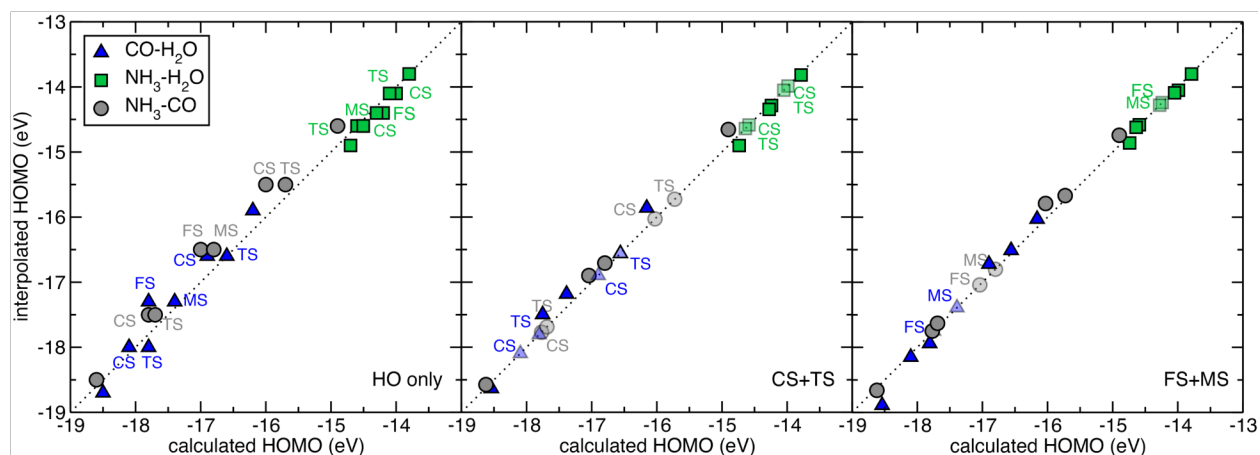


Figure A-2. Calculated vs. linearly interpolated HOMO level (eV) for singlet Fe(II) complexes with pairs of any of the three ligands: NH₃, H₂O, and CO. From left to right: interpolation between homoleptic complexes (HO only), interpolation using homoleptic complexes as well as *cis* symmetric and *trans* symmetric complex energies (CS+TS), or interpolation using homoleptic complexes as well as *fac* and *mer* symmetric complex energies (FS+MS). Points are colored according to the pair of ligands they correspond to: NH₃-H₂O (green squares), NH₃-CO (gray circles), and CO-H₂O (blue triangles), as indicated in inset legend. Key isomers are annotated. Points provided for the fit are translucent, whereas the remaining points are opaque. In all panes, a black dotted parity line is shown.

Table A-9. Mean absolute errors (MAE) in kcal/mol for Fe(II) complex ΔE_{H-L} estimates from HO-only averaging, including CS/TS in the averaging, and including FS/MS averaging for pairs of ligands, L₁ and L₂. The CS/TS and FS/MS errors are both averaged over only the points for which exact energies were not provided (interp.) as well as over all (all) points.

		ΔE_{H-L} MAE (kcal/mol)				
L ₁	L ₂	HO	CS+TS all	CS+TS interp.	FS+MS all	FS+MS interp.

H ₂ O	CH ₃ CN	6.3	0.7	1.4	2.6	3.5
CH ₃ CN	CO	2.1	0.4	0.8	1.4	1.9
H ₂ O	CO	5.1	1.2	2.4	2.3	3.1
NH ₃	H ₂ O	1.0	0.6	1.2	0.8	1.1
NH ₃	CO	2.5	1.1	2.2	1.3	1.7
Overall		3.4	0.8	1.6	1.7	2.3

Table A-10. Singlet HOMO level values for pairs of two ligand types in Fe(II) complexes (in eV) along with interpolated values using only HO complexes, HO+CS/TS complexes, or HO+FS/MS complexes.

L ₁	L ₂	Symmetry	Number of ligands				HOMO	Interpolated		
			CO	H ₂ O	CH ₃ CN	NH ₃		HO	CS/TS	FS/MS
CH ₃ CN	H ₂ O	5+1	0	1	5	0	-13.56	-13.65	-13.57	-13.63
CH ₃ CN	H ₂ O	CS	0	2	4	0	-13.82	-13.95	-13.82	-13.88
CH ₃ CN	H ₂ O	TS	0	2	4	0	-13.80	-13.95	-13.80	-13.85
H ₂ O	CH ₃ CN	FS	0	3	3	0	-14.19	-14.25	-14.15	-14.19
H ₂ O	CH ₃ CN	MS	0	3	3	0	-14.08	-14.25	-14.11	-14.08
H ₂ O	CH ₃ CN	CS	0	4	2	0	-14.48	-14.56	-14.48	-14.49
H ₂ O	CH ₃ CN	TS	0	4	2	0	-14.41	-14.56	-14.41	-14.46
H ₂ O	CH ₃ CN	5+1	0	5	1	0	-14.82	-14.86	-14.79	-14.84
CO	CH ₃ CN	5+1	5	0	1	0	-17.80	-18.44	-18.10	-18.32
CO	CH ₃ CN	CS	4	0	2	0	-16.82	-17.42	-16.82	-17.10
CO	CH ₃ CN	TS	4	0	2	0	-16.75	-17.42	-16.75	-17.02
CO	CH ₃ CN	FS	3	0	3	0	-16.02	-16.40	-15.92	-16.02
CO	CH ₃ CN	MS	3	0	3	0	-15.82	-16.40	-15.80	-15.82
CH ₃ CN	CO	CS	2	0	4	0	-15.02	-15.38	-15.02	-15.06
CH ₃ CN	CO	TS	2	0	4	0	-14.86	-15.38	-14.86	-14.98
CH ₃ CN	CO	5+1	1	0	5	0	-14.11	-14.37	-14.10	-14.24
CO	H ₂ O	5+1	5	1	0	0	-18.54	-18.75	-18.63	-18.89
CO	H ₂ O	CS	4	2	0	0	-18.10	-18.03	-18.10	-18.15
CO	H ₂ O	TS	4	2	0	0	-17.81	-18.03	-17.81	-17.94
H ₂ O	CO	FS	3	3	0	0	-17.75	-17.31	-17.50	-17.75
H ₂ O	CO	MS	3	3	0	0	-17.39	-17.31	-17.18	-17.39
H ₂ O	CO	CS	2	4	0	0	-16.90	-16.60	-16.90	-16.72
H ₂ O	CO	TS	2	4	0	0	-16.56	-16.60	-16.56	-16.51
H ₂ O	CO	5+1	1	5	0	0	-16.16	-15.88	-15.86	-16.03
H ₂ O	NH ₃	5+1	0	5	0	1	-14.74	-14.90	-14.90	-14.86
H ₂ O	NH ₃	CS	0	4	0	2	-14.59	-14.64	-14.59	-14.58

H ₂ O	NH ₃	TS	0	4	0	2	-14.64	-14.64	-14.64	-14.62
H ₂ O	NH ₃	FS	0	3	0	3	-14.24	-14.37	-14.29	-14.24
H ₂ O	NH ₃	MS	0	3	0	3	-14.28	-14.37	-14.35	-14.28
NH ₃	H ₂ O	CS	0	2	0	4	-13.99	-14.11	-13.99	-14.05
NH ₃	H ₂ O	TS	0	2	0	4	-14.05	-14.11	-14.05	-14.09
NH ₃	H ₂ O	5+1	0	1	0	5	-13.79	-13.85	-13.82	-13.80
CO	NH ₃	5+1	5	0	0	1	-18.62	-18.48	-18.58	-18.66
CO	NH ₃	CS	4	0	0	2	-17.77	-17.50	-17.77	-17.75
CO	NH ₃	TS	4	0	0	2	-17.69	-17.50	-17.69	-17.63
CO	NH ₃	FS	3	0	0	3	-17.04	-16.52	-16.90	-17.04
CO	NH ₃	MS	3	0	0	3	-16.80	-16.52	-16.71	-16.80
NH ₃	CO	CS	2	0	0	4	-16.03	-15.54	-16.03	-15.79
NH ₃	CO	TS	2	0	0	4	-15.73	-15.54	-15.73	-15.67
NH ₃	CO	5+1	1	0	0	5	-14.90	-14.56	-14.65	-14.74

Table A-11. Mean absolute errors (MAE) in eV for singlet complex HOMO levels from HO-only averaging, including CS/TS in the averaging, and including FS/MS averaging for pairs of ligands, L₁ and L₂. The CS/TS and FS/MS errors are both averaged over only the points for which exact energies were not provided (interp.) as well as over all (all) points.

		HOMO MAE (eV)				
L₁	L₂	HO	CS+TS all	CS+TS interp.	FS+MS all	FS+MS interp.
H ₂ O	CH ₃ CN	0.11	0.01	0.02	0.03	0.04
CH ₃ CN	CO	0.51	0.05	0.10	0.17	0.23
H ₂ O	CO	0.20	0.11	0.22	0.11	0.15
NH ₃	H ₂ O	0.09	0.04	0.08	0.03	0.04
NH ₃	CO	0.30	0.25	0.50	0.07	0.09
Overall		0.24	0.09	0.18	0.08	0.11

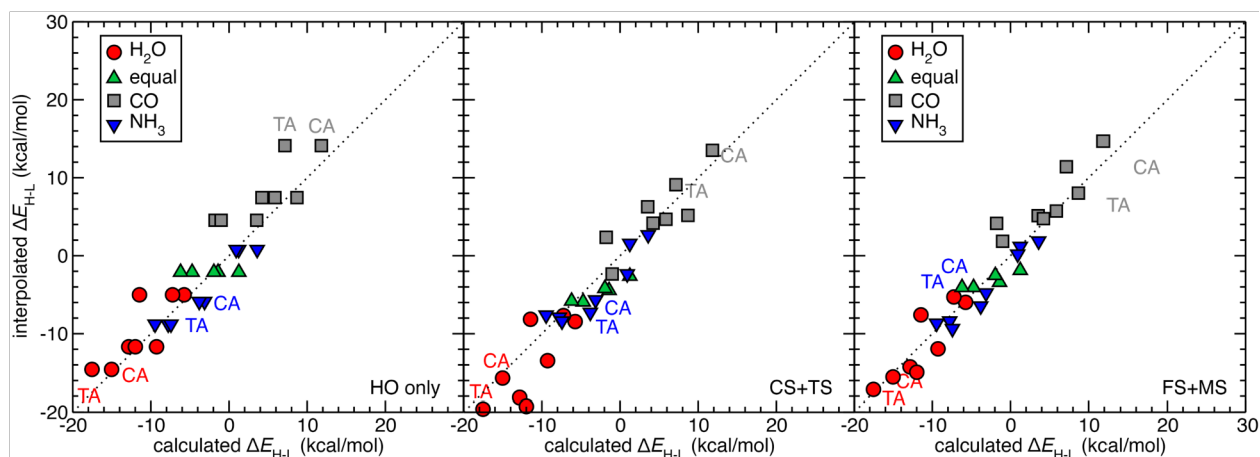


Figure A-3. Calculated vs. linearly interpolated ΔE_{H-L} (kcal/mol) for Fe(II) complexes with at least one each of three ligands: NH_3 , H_2O , and CO . From left to right: interpolation from homoleptic complexes (HO only), interpolation using homoleptic complexes as well as *cis* symmetric and *trans* symmetric complex energies derived from pairs of ligands (CS+TS), or interpolation using homoleptic complexes as well as *fac* and *mer* symmetric complex energies derived from pairs of ligands (FS+MS). Points are colored according to the ligand with the highest stoichiometric coefficient: H_2O (red circles), CO (gray squares), and NH_3 (blue down triangles), or equal weight of all ligands (green up triangles), as indicated in inset legend. Key isomers are annotated. In all panes, a black dotted parity line is shown.

Table A-12. ΔE_{H-L} values for Fe(II) complexes (in kcal/mol) with a combination of three ligand types, H_2O , CO , and CH_3CN , in Fe(II) complexes along with interpolated values using only HO complexes, HO+CS/TS complexes, or HO+FS/MS complexes. For the three EA isomers, L_3 is the ligand in the *trans* position as indicated in the main text.

L ₁	L ₂	L ₃	Symmetry	Number of ligands			splitting	Interpolated		
				CH ₃ CN	CO	H ₂ O		HO	CS/TS	FS/MS
H ₂ O	CO	CH ₃ CN	TA	1	1	4	-15.7	-5.7	-16.1	-11.5
H ₂ O	CO	CH ₃ CN	CA	1	1	4	-7.7	-5.7	-8.8	-7.0
CO	H ₂ O	CH ₃ CN	FA	1	2	3	3.2	3.9	-1.6	2.5
CO	H ₂ O	CH ₃ CN	MAC	1	2	3	-4.6	3.9	-4.1	0.4
CO	H ₂ O	CH ₃ CN	MAT	1	2	3	-5.5	3.9	-4.6	-1.9
CH ₃ CN	H ₂ O	CO	FA	2	1	3	5.2	6.1	1.8	4.7
CH ₃ CN	H ₂ O	CO	MAC	2	1	3	-2.3	6.1	-2.9	2.4
CH ₃ CN	H ₂ O	CO	MAT	2	1	3	-1.7	6.1	-4.4	0.3
CO	CH ₃ CN	H ₂ O	EA	2	2	2	6.1	15.7	10.0	12.1
H ₂ O	CO	CH ₃ CN	DTS	2	2	2	6.2	15.7	10.4	12.1
H ₂ O	CO	CH ₃ CN	DCS	2	2	2	10.9	15.7	13.1	15.1

H ₂ O	CH ₃ CN	CO	EA	2	2	2	12.0	15.7	11.7	13.5
CO	H ₂ O	CH ₃ CN	EA	2	2	2	12.9	15.7	12.1	13.6
H ₂ O	CO	CH ₃ CN	FA	1	3	2	9.0	13.4	14.5	13.5
H ₂ O	CO	CH ₃ CN	MAC	1	3	2	8.4	13.4	14.6	13.4
H ₂ O	CO	CH ₃ CN	MAT	1	3	2	3.5	13.4	9.9	11.1
CH ₃ CN	CO	H ₂ O	FA	2	3	1	22.9	25.2	21.9	25.3
CH ₃ CN	CO	H ₂ O	MAC	2	3	1	21.0	25.2	21.4	23.0
CH ₃ CN	CO	H ₂ O	MAT	2	3	1	23.1	25.2	22.4	22.9
CO	H ₂ O	CH ₃ CN	TA	1	4	1	15.7	23.0	21.4	20.7
CO	H ₂ O	CH ₃ CN	CA	1	4	1	19.4	23.0	21.7	23.1
H ₂ O	CH ₃ CN	CO	FA	3	1	2	12.1	17.9	15.7	17.4
H ₂ O	CH ₃ CN	CO	MAC	3	1	2	13.8	17.9	16.7	17.3
H ₂ O	CH ₃ CN	CO	MAT	3	1	2	9.7	17.9	14.2	15.2
CO	CH ₃ CN	H ₂ O	FA	3	2	1	23.6	27.4	26.1	26.9
CO	CH ₃ CN	H ₂ O	MAC	3	2	1	25.2	27.4	24.7	24.8
CO	CH ₃ CN	H ₂ O	MAT	3	2	1	23.8	27.4	24.8	24.8
CH ₃ CN	H ₂ O	CO	CA	4	1	1	25.2	29.7	26.3	29.2
CH ₃ CN	CO	H ₂ O	TA	4	1	1	26.9	29.7	25.9	27.0

Table A-13. $\Delta E_{\text{H-L}}$ values for Fe(II) complexes (in kcal/mol) with a combination of three ligand types, H₂O, CO, and NH₃, in Fe(II) complexes along with interpolated values using only HO complexes, HO+CS/TS complexes, or HO+FS/MS complexes. For the three EA isomers, L₃ is the ligand in the *trans* position as indicated in the main text.

				Number of ligands			Interpolated			
L ₁	L ₂	L ₃	symmetry	NH ₃	CO	H ₂ O	splitting	HO	CS/TS	FS/MS
H ₂ O	CO	NH ₃	CA	1	1	4	-15.0	-14.6	-15.7	-15.5
H ₂ O	CO	NH ₃	TA	1	1	4	-17.5	-14.6	-19.6	-17.2
CO	H ₂ O	NH ₃	FA	1	2	3	-5.7	-5.0	-8.4	-6.0
CO	H ₂ O	NH ₃	MAC	1	2	3	-7.2	-5.0	-7.7	-5.3
CO	H ₂ O	NH ₃	MAT	1	2	3	-11.5	-5.0	-8.2	-7.6
NH ₃	H ₂ O	CO	FA	2	1	3	-9.3	-11.7	-13.5	-12.0
NH ₃	H ₂ O	CO	MAC	2	1	3	-12.8	-11.7	-18.2	-14.3
NH ₃	H ₂ O	CO	MAT	2	1	3	-12.0	-11.7	-19.3	-14.9
H ₂ O	CO	NH ₃	DCS	2	2	2	1.3	-2.1	-2.6	-1.9
H ₂ O	CO	NH ₃	DTS	2	2	2	-4.7	-2.1	-5.9	-4.1
CO	NH ₃	H ₂ O	EA	2	2	2	-1.4	-2.1	-4.5	-3.4
H ₂ O	NH ₃	CO	EA	2	2	2	-6.2	-2.1	-5.8	-4.1
CO	H ₂ O	NH ₃	EA	2	2	2	-1.9	-2.1	-4.2	-2.5
H ₂ O	CO	NH ₃	FA	1	3	2	3.5	4.5	6.3	5.1

H2O	CO	NH3	MAC	1	3	2	-1.8	4.5	2.4	4.1
H2O	CO	NH3	MAT	1	3	2	-1.0	4.5	-2.3	1.8
NH3	CO	H2O	FA	2	3	1	8.7	7.4	5.2	8.0
NH3	CO	H2O	MAC	2	3	1	5.9	7.4	4.7	5.7
NH3	CO	H2O	MAT	2	3	1	4.2	7.4	4.2	4.7
CO	H2O	NH3	CA	1	4	1	11.8	14.1	13.5	14.7
CO	H2O	NH3	TA	1	4	1	7.1	14.1	9.1	11.4
H2O	NH3	CO	FA	3	1	2	-7.8	-8.8	-7.9	-8.4
H2O	NH3	CO	MAC	3	1	2	-7.4	-8.8	-8.4	-9.4
H2O	NH3	CO	MAT	3	1	2	-9.5	-8.8	-7.6	-8.7
CO	NH3	H2O	FA	3	2	1	3.6	0.8	2.7	1.9
CO	NH3	H2O	MAC	3	2	1	1.2	0.8	1.6	1.2
CO	NH3	H2O	MAT	3	2	1	0.9	0.8	-2.3	0.2
NH3	H2O	CO	CA	4	1	1	-3.1	-5.9	-5.7	-4.8
NH3	CO	H2O	TA	4	1	1	-3.8	-5.9	-7.3	-6.5

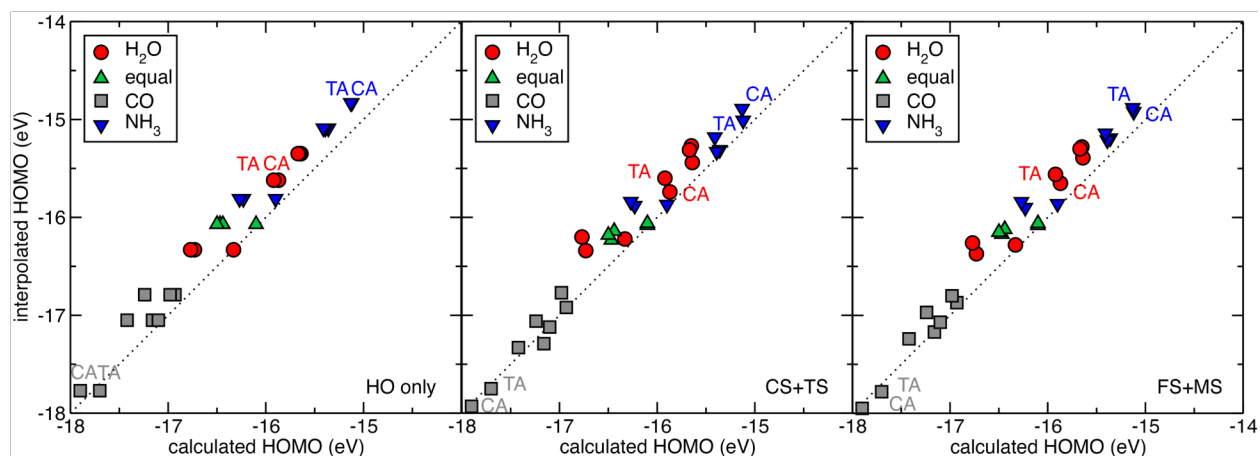


Figure A-4. Calculated vs. linearly interpolated HOMO level (eV) for singlet Fe(II) complexes with at least one each of three ligands: NH₃, H₂O, and CO. From left to right: interpolation from homoleptic complexes (HO only), interpolation using homoleptic complexes as well as *cis* symmetric and *trans* symmetric complex energies derived from pairs of ligands (CS+TS), or interpolation using homoleptic complexes as well as *fac* and *mer* symmetric complex energies derived from pairs of ligands (FS+MS). Points are colored according to the ligand with the highest stoichiometric coefficient: H₂O (red circles), CO (gray squares), and NH₃ (blue down triangles), or equal weight of all ligands (green up triangles), as indicated in inset legend. Key isomers are annotated. In all panes, a black dotted parity line is shown.

Table A-14. Singlet HOMO level values with a combination of three ligand types, H₂O, CO, and CH₃CN, in Fe(II) complexes (in eV) along with interpolated values using only HO complexes, HO+CS/TS complexes, or HO+FS/MS complexes. For the three EA isomers, L₃ is the ligand in the *trans* position as indicated in the main text.

L ₁	L ₂	L ₃	symmetry	Number of ligands			Interpolated			
				CH ₃ CN	CO	H ₂ O	HOMO	HO	CS/TS	FS/MS
H ₂ O	CO	CH ₃ CN	CA	1	1	4	-15.71	-15.58	-15.54	-15.60
H ₂ O	CO	CH ₃ CN	TA	1	1	4	-15.50	-15.58	-15.69	-15.48
CO	H ₂ O	CH ₃ CN	FA	1	2	3	-16.40	-16.29	-16.14	-16.32
CO	H ₂ O	CH ₃ CN	MAC	1	2	3	-16.34	-16.29	-16.46	-16.31
CO	H ₂ O	CH ₃ CN	MAT	1	2	3	-16.13	-16.29	-16.31	-16.20
CH ₃ CN	H ₂ O	CO	FA	2	1	3	-15.23	-15.27	-15.23	-15.30
CH ₃ CN	H ₂ O	CO	MAC	2	1	3	-15.13	-15.27	-15.06	-15.20
CH ₃ CN	H ₂ O	CO	MAT	2	1	3	-15.30	-15.27	-15.18	-15.18
H ₂ O	CO	CH ₃ CN	DCS	2	2	2	-15.89	-15.99	-15.76	-15.90
H ₂ O	CO	CH ₃ CN	DTS	2	2	2	-15.75	-15.99	-15.77	-15.80
CO	CH ₃ CN	H ₂ O	EA	2	2	2	-15.76	-15.99	-15.80	-15.82
H ₂ O	CH ₃ CN	CO	EA	2	2	2	-15.65	-15.99	-15.62	-15.81
CO	H ₂ O	CH ₃ CN	EA	2	2	2	-15.87	-15.99	-15.87	-15.87
H ₂ O	CO	CH ₃ CN	FA	1	3	2	-16.96	-17.01	-16.86	-16.91
H ₂ O	CO	CH ₃ CN	MAC	1	3	2	-16.77	-17.01	-16.82	-16.87
H ₂ O	CO	CH ₃ CN	MAT	1	3	2	-16.76	-17.01	-16.65	-16.77
CH ₃ CN	CO	H ₂ O	FA	2	3	1	-16.43	-16.71	-16.56	-16.61
CH ₃ CN	CO	H ₂ O	MAC	2	3	1	-16.29	-16.71	-16.41	-16.50
CH ₃ CN	CO	H ₂ O	MAT	2	3	1	-16.35	-16.71	-16.33	-16.46
CO	H ₂ O	CH ₃ CN	CA	1	4	1	-17.46	-17.73	-17.46	-17.63
CO	H ₂ O	CH ₃ CN	TA	1	4	1	-17.14	-17.73	-17.28	-17.48
H ₂ O	CH ₃ CN	CO	FA	3	1	2	-14.77	-14.97	-14.60	-14.77
H ₂ O	CH ₃ CN	CO	MAC	3	1	2	-14.85	-14.97	-14.92	-14.76
H ₂ O	CH ₃ CN	CO	MAT	3	1	2	-14.80	-14.97	-14.84	-14.72
CO	CH ₃ CN	H ₂ O	FA	3	2	1	-15.38	-15.69	-15.19	-15.49
CO	CH ₃ CN	H ₂ O	MAC	3	2	1	-15.38	-15.69	-15.15	-15.45
CO	CH ₃ CN	H ₂ O	MAT	3	2	1	-15.27	-15.69	-15.28	-15.44
CH ₃ CN	H ₂ O	CO	CA	4	1	1	-14.45	-14.67	-14.29	-14.47
CH ₃ CN	CO	H ₂ O	TA	4	1	1	-14.42	-14.67	-14.33	-14.42

Table A-15. Singlet HOMO level values with a combination of three ligand types, H₂O, CO, and NH₃, in Fe(II) complexes (in eV) along with interpolated values using only HO complexes, HO+CS/TS complexes, or HO+FS/MS complexes. For the three EA isomers, L₃ is the ligand in the *trans* position as indicated in the main text.

L ₁	L ₂	L ₃	Symmetry	Number of ligands			HOMO	Interpolated		
				NH ₃	CO	H ₂ O		HO	CS/TS	FS/MS
H ₂ O	CO	NH ₃	CA	1	1	4	-15.87	-15.62	-15.74	-15.65
H ₂ O	CO	NH ₃	TA	1	1	4	-15.92	-15.62	-15.60	-15.56
NH ₃	H ₂ O	CO	FA	2	1	3	-15.64	-15.35	-15.44	-15.39
NH ₃	H ₂ O	CO	MAC	2	1	3	-15.65	-15.35	-15.27	-15.28
NH ₃	H ₂ O	CO	MAT	2	1	3	-15.67	-15.35	-15.31	-15.30
CO	H ₂ O	NH ₃	FA	1	2	3	-16.73	-16.33	-16.34	-16.37
CO	H ₂ O	NH ₃	MAC	1	2	3	-16.77	-16.33	-16.20	-16.26
CO	H ₂ O	NH ₃	MAT	1	2	3	-16.33	-16.33	-16.22	-16.28
H ₂ O	CO	NH ₃	DCS	2	2	2	-16.47	-16.07	-16.23	-16.17
H ₂ O	CO	NH ₃	DTS	2	2	2	-16.10	-16.07	-16.08	-16.08
CO	NH ₃	H ₂ O	EA	2	2	2	-16.44	-16.07	-16.14	-16.12
H ₂ O	NH ₃	CO	EA	2	2	2	-16.10	-16.07	-16.06	-16.06
CO	H ₂ O	NH ₃	EA	2	2	2	-16.50	-16.07	-16.18	-16.15
H ₂ O	CO	NH ₃	FA	1	3	2	-17.42	-17.05	-17.33	-17.24
H ₂ O	CO	NH ₃	MAC	1	3	2	-17.16	-17.05	-17.29	-17.17
H ₂ O	CO	NH ₃	MAT	1	3	2	-17.10	-17.05	-17.12	-17.07
NH ₃	CO	H ₂ O	FA	2	3	1	-17.24	-16.79	-17.06	-16.97
NH ₃	CO	H ₂ O	MAC	2	3	1	-16.93	-16.79	-16.92	-16.87
NH ₃	CO	H ₂ O	MAT	2	3	1	-16.98	-16.79	-16.77	-16.80
CO	H ₂ O	NH ₃	CA	1	4	1	-17.90	-17.77	-17.93	-17.95
CO	H ₂ O	NH ₃	TA	1	4	1	-17.70	-17.77	-17.75	-17.78
H ₂ O	NH ₃	CO	FA	3	1	2	-15.36	-15.09	-15.31	-15.19
H ₂ O	NH ₃	CO	MAC	3	1	2	-15.39	-15.09	-15.33	-15.21
H ₂ O	NH ₃	CO	MAT	3	1	2	-15.41	-15.09	-15.18	-15.14
CO	NH ₃	H ₂ O	FA	3	2	1	-16.23	-15.81	-15.88	-15.90
CO	NH ₃	H ₂ O	MAC	3	2	1	-16.27	-15.81	-15.84	-15.84
CO	NH ₃	H ₂ O	MAT	3	2	1	-15.90	-15.81	-15.87	-15.86
NH ₃	H ₂ O	CO	CA	4	1	1	-15.12	-14.83	-15.01	-14.92
NH ₃	CO	H ₂ O	TA	4	1	1	-15.13	-14.83	-14.89	-14.88

Text A-3. Expressions for interpolating ternary complex energies from CS/TS binary complexes.

We indicate the stoichiometry in the complex to distinguish where the energies are derived from:

$$\begin{aligned}
E(\text{CA}-(\text{L}_1)_4(\text{L}_2)_1(\text{L}_3)_1) &= \frac{1}{2}(E(\text{CS}(\text{L}_1)_4(\text{L}_2)_2)+E(\text{CS}(\text{L}_1)_4(\text{L}_3)_2)) \\
E(\text{TA}-(\text{L}_1)_4(\text{L}_2)_1(\text{L}_3)_1) &= \frac{1}{2}(E(\text{TS}(\text{L}_1)_4(\text{L}_2)_2)+E(\text{TS}(\text{L}_1)_4(\text{L}_3)_2)) \\
E(\text{FA}-(\text{L}_1)_3(\text{L}_2)_2(\text{L}_3)_1) &= \frac{1}{2}(E(\text{CS}(\text{L}_1)_4(\text{L}_3)_2)+E(\text{CS}(\text{L}_2)_4(\text{L}_1)_2)) \\
E(\text{MAC}-(\text{L}_1)_3(\text{L}_2)_2(\text{L}_3)_1) &= \frac{1}{2}(E(\text{TS}(\text{L}_1)_4(\text{L}_3)_2)+E(\text{CS}(\text{L}_2)_4(\text{L}_1)_2)) \\
E(\text{MAT}-(\text{L}_1)_3(\text{L}_2)_2(\text{L}_3)_1) &= \frac{1}{2}(E(\text{TS}(\text{L}_1)_4(\text{L}_3)_2)+E(\text{TS}(\text{L}_2)_4(\text{L}_1)_2)) \\
E(\text{DCS}-(\text{L}_1)_2(\text{L}_2)_2(\text{L}_3)_2) &= \frac{1}{6}(E(\text{CS}(\text{L}_1)_4(\text{L}_2)_2)+E(\text{CS}(\text{L}_2)_4(\text{L}_1)_2)+E(\text{CS}(\text{L}_1)_4(\text{L}_3)_2) \\
&\quad +E(\text{CS}(\text{L}_3)_4(\text{L}_1)_2)+E(\text{CS}(\text{L}_2)_4(\text{L}_3)_2)+E(\text{CS}(\text{L}_3)_4(\text{L}_2)_2)) \\
E(\text{DCS}-(\text{L}_1)_2(\text{L}_2)_2(\text{L}_3)_2) &= \frac{1}{6}(E(\text{TS}(\text{L}_1)_4(\text{L}_2)_2)+E(\text{TS}(\text{L}_2)_4(\text{L}_1)_2)+E(\text{TS} \\
&\quad (\text{L}_1)_4(\text{L}_3)_2)+E(\text{TS}(\text{L}_3)_4(\text{L}_1)_2)+E(\text{TS}(\text{L}_2)_4(\text{L}_3)_2)+E(\text{TS}(\text{L}_3)_4(\text{L}_2)_2)) \\
E(\text{EA}-(\text{L}_1)_2(\text{L}_2)_2(\text{L}_3)_2) &= \frac{1}{6}(E(\text{CS}(\text{L}_1)_4(\text{L}_2)_2)+E(\text{CS}(\text{L}_2)_4(\text{L}_1)_2)+E(\text{TS}(\text{L}_1)_4(\text{L}_3)_2)+ \\
&\quad E(\text{TS}(\text{L}_3)_4(\text{L}_1)_2)+E(\text{TS}(\text{L}_2)_4(\text{L}_3)_2)+E(\text{TS}(\text{L}_3)_4(\text{L}_2)_2))
\end{aligned}$$

where the third ligand in the EA complex is the one that is *trans* to itself so energies involving that ligand are derived from the binary TS complexes, whereas the remainder are derived from CS complexes.

Table A-16. Mean absolute error (MAE) for singlet complex HOMO levels (in eV) and Fe(II) complex $\Delta E_{\text{H-L}}$ (in kcal/mol) for combinations of ligands, L₁, L₂, and L₃ evaluated with three averaging schemes: HO-only, CS/TS-derived averaging, and FS/MS-derived averaging.

			HOMO MAE (eV)			$\Delta E_{\text{H-L}}$ MAE (kcal/mol)		
L ₁	L ₂	L ₃	HO	CS+TS	FS+MS	HO	CS+TS	FS+MS
H ₂ O	CO	NH ₃	0.26	0.19	0.22	2.31	2.30	1.78
H ₂ O	CO	CH ₃ CN	0.21	0.10	0.09	5.23	2.45	3.22
Overall			0.24	0.15	0.16	3.77	2.38	2.50

Table A-17. Properties of 56 homoleptic Fe(II) complexes derived from either homoleptic examples in the CSD (36 "HO") with neutral ligands or from neutral ligands only present in binary and ternary Fe(II) complexes (20 "B or T") from DFT. Both the $\Delta E_{\text{H-L}}$ (in kcal/mol) and the HOMO level of the singlet complex (in eV) are shown. Complexes are distinguished by stoichiometry.

# atoms	Stoichiometry	Source	refcode	$\Delta E_{\text{H-L}}$ (kcal/mol)	HOMO level (eV)
37	Fe1C12N6H18	HO	ACEYOW01	-4.98	-12.97
19	Fe1O6H12	HO	AMAVOB	-26.78	-15.11

55	Fe1C18N12H24	HO	AXAKIT01	-9.06	-10.90
97	Fe1C30N6O6H54	HO	BIRSAZ	-30.77	-10.48
67	Fe1C30N6H30	HO	BUSVAO	-16.02	-11.82
79	Fe1C24O6H48	HO	BUSVES	-33.88	-12.51
25	Fe1N6H18	HO	CACDIW	-9.64	-13.58
73	Fe1C18N6O6H42	HO	CALMOS01	-21.32	-10.50
13	Fe1C6O6	HO	CEHHON	30.03	-19.46
61	Fe1C18N6O6H30	HO	CEMGUX	-23.44	-11.38
121	Fe1C42N6O6H66	HO	DECRAE	45.81	-10.93
67	Fe1C18N18H30	HO	DEDLAB	-8.03	-11.66
73	Fe1C18N6S6H42	HO	FAVDOV	-17.25	-10.27
85	Fe1C18N12S6H48	HO	FEGGAZ	-24.89	-10.13
73	Fe1C30N6O6H30	HO	FEHPYO	-26.10	-9.91
49	Fe1C18N6O6H18	HO	FEISXC01	-5.02	-12.21
73	Fe1C24N12H36	HO	FIWQUY	-8.06	-10.44
169	Fe1C84N24O12H48	HO	GOGSAZ01	-0.40	-9.19
61	Fe1C12N24H24	HO	HIPXOW	-5.43	-11.37
127	Fe1C36N6O6H78	HO	HUGVIQ	-30.16	-10.36
97	Fe1C24N24H48	HO	JANSAS01	-1.77	-11.43
61	Fe1C12O6S6H36	HO	JOHCOZ	-30.25	-10.56
61	Fe1C12N24H24	HO	JUVHEN	-1.95	-11.81
37	Fe1C12N6H18	HO	MICYFE10	43.67	-13.35
115	Fe1C30N24H60	HO	NIGXUY01	-1.83	-11.32
79	Fe1C18N24Cl6H30	HO	PEJQIF01	-1.93	-11.57
55	Fe1C6N30H18	HO	SIDMAW	-2.30	-12.01
55	Fe1C6N30H18	HO	SIDMEA	-7.85	-12.17
79	Fe1C18N24H36	HO	TUNBIN	-1.14	-11.55
37	Fe1C6O6H24	HO	USIMOA	-25.75	-13.70
97	Fe1C24N24Cl6H42	HO	VIFNAC01	-2.71	-11.82
115	Fe1C30N24H60	HO	VOJPAM	-1.60	-11.35
91	Fe1C18N42H30	HO	XOJCIL	-1.79	-11.14
79	Fe1C18N24Br6H30	HO	YAGYIP01	-1.49	-11.69
79	Fe1C18N24I6H30	HO	YAGYUB01	-1.66	-11.28
97	Fe1C24N24H48	HO	YEQKIR	-1.43	-11.32
97	Fe1C18O18P6H54	B or T	--	6.37	-12.08

115	Fe1C42N24H48	B or T	--	-19.67	-10.27
79	Fe1C18P6H54	B or T	--	-16.94	-12.36
61	Fe1C12N6O12H30	B or T	--	-38.56	-10.26
121	Fe1C60N12H48	B or T	--	-17.54	-10.69
145	Fe1C72N12H60	B or T	--	-16.08	-10.04
91	Fe1C36N12O6H36	B or T	--	-17.00	-11.21
151	Fe1C36O18P6H90	B or T	--	9.61	-11.51
169	Fe1C60O12P6H90	B or T	--	-10.82	-10.68
151	Fe1C72N24O6H48	B or T	--	-16.64	-10.11
61	Fe1C24N12Cl6H18	B or T	--	-16.56	-13.24
73	Fe1C24N18H30	B or T	--	-15.67	-11.23
97	Fe1C48N6H42	B or T	--	43.43	-10.79
139	Fe1C54N42H42	B or T	--	-4.66	-10.19
133	Fe1C60N12O12H48	B or T	--	-22.25	-9.33
73	Fe1C12N12H48	B or T	--	-9.36	-11.36
103	Fe1C42N24H36	B or T	--	-5.43	-10.10
19	Fe1C6N6H6	B or T	--	43.66	-15.08
127	Fe1C60N12O6H48	B or T	--	-25.06	-9.85
97	Fe1C24N24H48	B or T	--	-1.13	-11.32

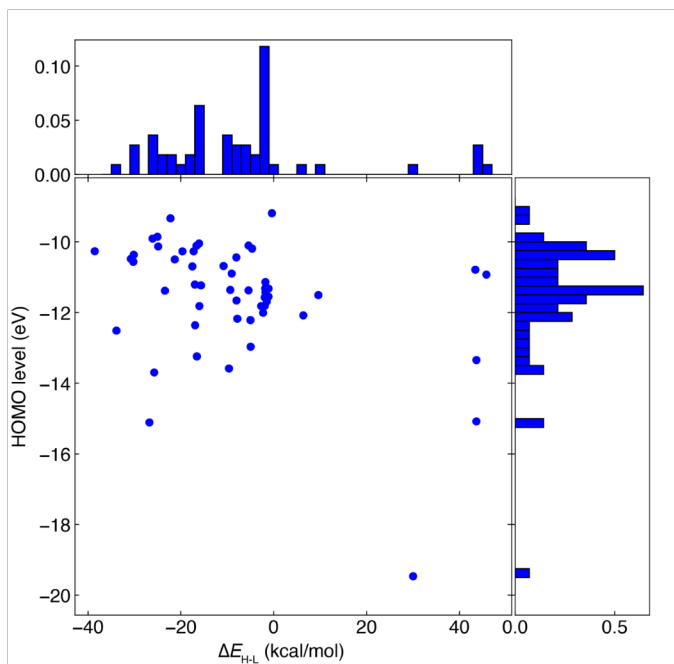


Figure A-5. The ΔE_{H-L} (in kcal/mol) vs. singlet HOMO level (in eV) for 56 homoleptic Fe(II) complexes shown as a scatter plot (middle) with normalized marginal 1D histograms at top and right, respectively computed with B3LYP.

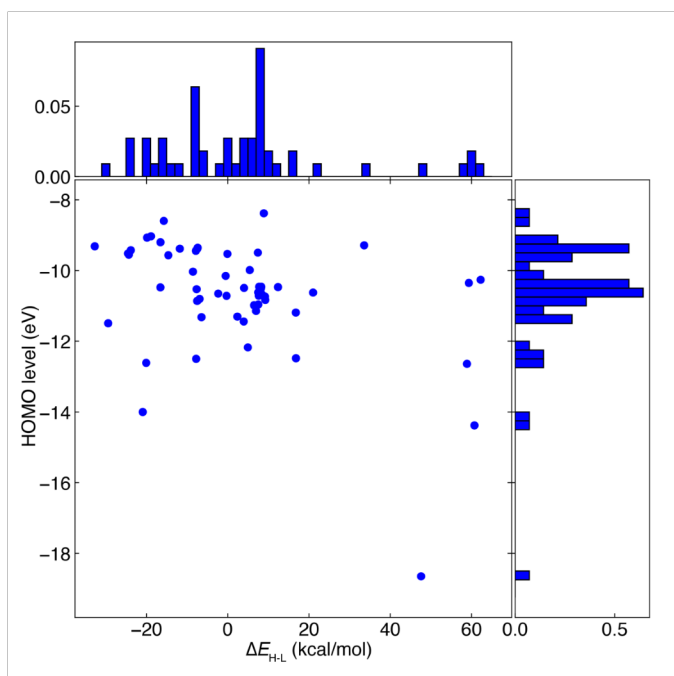


Figure A-6. The ΔE_{H-L} (in kcal/mol) vs. singlet HOMO level (in eV) for 56 homoleptic Fe(II) complexes shown as a scatter plot (middle) with normalized marginal 1D

histograms at top and right, respectively, computed with modified B3LYP with 10% HF exchange.

Table A-18. Twelve most common ligands in target zone (i.e., ΔE_{H-L} from -4 to +4 kcal/mol and singlet HOMO level from -14.0 to -13.0 eV) for Fe(II) complexes and the number of times the homoleptic-only averaging predicts that a complex in the zone will have that ligand in a binary complex or both binary and ternary complexes. Refcodes in italics refer to any complex that contains that ligand, whereas the remainder are the precise homoleptic Fe(II) complexes. The top 10 most frequent ligands were selected along with two high-frequency ligands (MeCN and ClPyz) for added diversity.

Refcode	Description	Short name	# binary	# binary or ternary	ΔE_{H-L} (kcal/mol)	HOMO (eV)
CEHHON	CO	CO	23	5071	30.0	-19.46
AMAVOB	H2O	H2O	4	671	-26.8	-15.11
RIFLEY	hydrogen isocyanide	CNH	6	1016	43.7	-15.08
<i>GOVKEK_comp_0</i>	4,4'-bipyridine	bpy20	2	364	-17.5	-10.69
USIMOA	methanol	MeOH	2	484	-25.7	-13.70
FAVDOV	dimethylthioformamide	DMTF	2	374	-17.3	-10.27
CACDIW	ammonia	NH3	4	532	-9.6	-13.58
<i>BOYWEU_comp_0</i>	dimethyltriazolopyrimidine	DMTP	2	348	-19.7	-10.27
<i>ARENUG_comp_0</i>	2-chloropyrazine	ClPyz	4	377	-16.6	-13.24
<i>MICYFE_10</i>	methyl isocyanide	CH ₃ CN (misc)	6	536	43.7	-13.35
ACEYOW01	NCCH3 (acetonitrile)	MeCN	2	439	-5.0	-12.97
<i>AMUZEP_comp_0</i>	4-(5-(pyridine-4-yl)-1,3,4-oxadiazol-2-yl)pyridine	bpy25	2	369	-16.6	-10.11

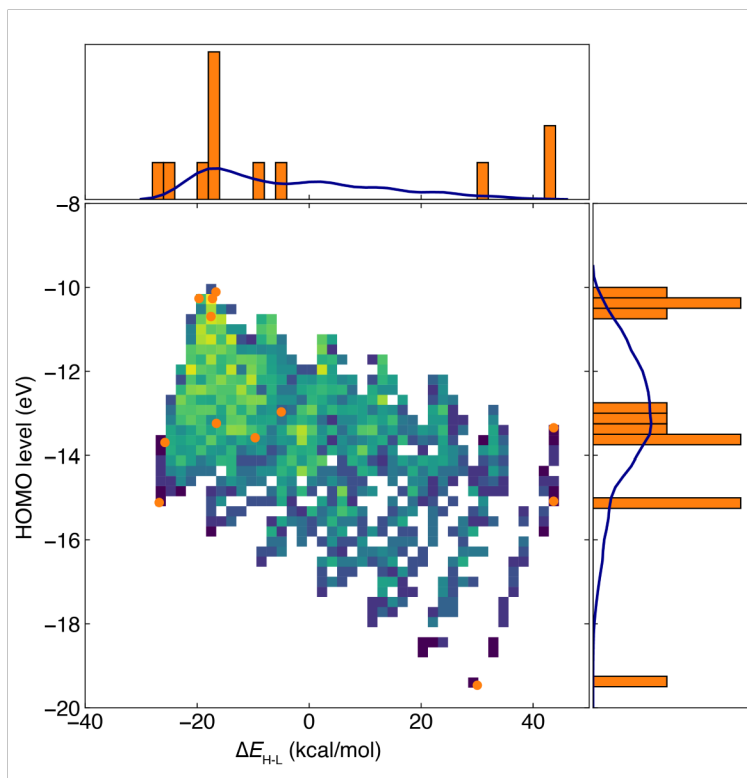


Figure A-7. Homoleptic-only interpolation for Fe(II) complexes: the ΔE_{H-L} (in kcal/mol) vs. singlet HOMO level (in eV) for 12 homoleptic Fe(II) complexes shown as both a scatter plot (middle, orange) and with normalized marginal 1D histograms at top and right (orange bars), respectively. The interpolated values are shown as a 2D histogram colored from low (purple) to high (yellow) density, and the same data is shown as a kernel density estimate on the histogram panes.

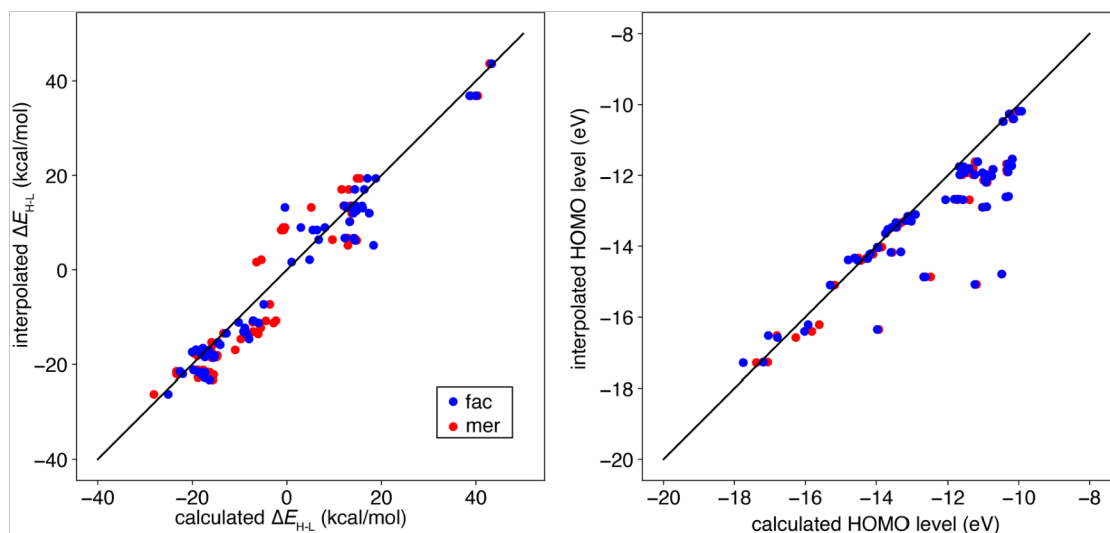


Figure A-8. Parity plot of calculated vs. HO-only interpolated ΔE_{H-L} (in kcal/mol, left) and HOMO level (in eV, right) for 66 *fac* (blue circles) and *mer* (red circles) derived from 12 homoleptic Fe(II) complexes. A black solid parity line is also shown.

Table A-19. MAEs of homoleptic-only interpolation over the FS/MS 132 complex subset with respect to calculated values.

	spin (kcal/mol)	HOMO (eV)
MS	3.88	0.78
FS	4.02	0.77

Table A-20. Binary complexes with their chemical formula and symmetry along with ΔE_{H-L} (in kcal/mol) and HOMO level (in eV) obtained through three approaches: homoleptic-only interpolation, FS/MS-augmented interpolation, and explicit calculation. For the cases that are FS or MS complexes, FS/MS-augmented interpolation refers to the explicitly calculated property. As a result, the reported MAEs are obtained over 5 complexes for FS/MS-interpolation but 8 complexes for HO-interpolation. The resulting calculated properties that are in the targeted zone are shown in bold.

Complex	Sym.	HO-interpolation		FS/MS-interpolation		Calculation	
		spin (kcal/mol)	HOMO (eV)	spin (kcal/mol)	HOMO (eV)	spin (kcal/mol)	HOMO (eV)
Fe(II)(MeCN) ₅ (CO)	5+1	0.9	-14.05	1.6	-13.95	1.5	-13.84
Fe(II)(MeCN) ₃ (NH ₃) ₃	MS	-7.3	-13.28	-3.6	-13.17	-3.6	-13.17
Fe(II)(CH ₃ CN) ₃ (MeOH) ₃	FS	9.0	-13.52	2.9	-13.67	2.9	-13.67

Fe(II)(CH ₃ CN) ₃ (MeOH) ₃	MS	9.0	-13.52	-0.8	-13.65	-0.8	-13.65
Fe(II)(ClPyz) ₄ (CO) ₂	CS	-1.0	-15.32	2.7	-13.73	4.9	-13.63
Fe(II)(ClPyz) ₄ (CO) ₂	TS	-1.0	-15.32	3.1	-13.70	5.7	-13.62
Fe(II)(MeOH) ₄ (CH ₃ CN) ₂	CS	-2.6	-13.58	-6.6	-13.56	-9.1	-13.80
Fe(II)(MeOH) ₄ (CH ₃ CN) ₂	TS	-2.6	-13.58	-9.1	-13.66	-13.9	-13.67
MAE		6.3	0.54	2.5	0.11		

Table A-21. Ternary complexes with their chemical formula and symmetry along with ΔE_{H-L} (in kcal/mol) and HOMO level (in eV) obtained through three approaches: homoleptic-only interpolation, FS/MS-augmented interpolation, and explicit calculation. The resulting calculated properties that are in the targeted zone are shown in bold.

Complex	Sym	HO-interpolation		FS/MS-interpolation		Calculation	
		spin (kcal/mol)	HOMO (eV)	spin (kcal/mol)	HOMO (eV)	spin (kcal/mol)	HOMO (eV)
Fe(II)(ClPyz) ₄ (CNH) ₁ (NH ₃) ₁	CA	-5.4	-13.60	-3.8	-13.32	-3.8	-13.37
Fe(II)(DMTF) ₃ (CO) ₂ (H ₂ O) ₁	FA	-3.1	-14.14	-3.0	-13.20	-8.2	-12.55
Fe(II)(NH ₃) ₄ (DMTF) ₁ (CO) ₁	TA	-4.3	-14.01	-3.2	-13.94	-0.2	-13.40
Fe(II)(ClPyz) ₄ (MeCN) ₁ (CO) ₁	CA	-6.9	-14.23	-3.8	-13.38	-0.5	-13.33
Fe(II)(MeCN) ₄ (ClPyz) ₁ (NH ₃) ₁	TA	-7.7	-13.12	-3.6	-13.03	-2.7	-12.83
Fe(II)(ClPyz) ₃ (CO) ₂ (DMTF) ₁	FA	-9.0	-13.29	-0.1	-13.96	-2.2	-13.20
MAE		4.8	0.62	2.5	0.37		

References

- ¹ C. Duan, J. P. Janet, F. Liu, A. Nandy and H. J. Kulik, J. Chem. Theory Comput. **15** (4), 2331-2345 (2019).

B. Appendix: Isovalent TMCs

Table B-1. Comparison of effects of equilibrium bond length differences from larger basis sets for B3LYP/LACVP* and B3LYP/def2-ZORA-TZVP (i.e., $a_{\text{HF}} = 0.2$) structures of Fe(II) or Ru(II)(CO)₆ in HS quintet and LS states.

The ratio of the resulting $S(\Delta E_{\text{H-L}})$ between the two basis sets evaluated with that basis set are shown as well along with their individual sensitivities in kcal/mol-HFX. The change in basis set for energetic evaluation is more significant than the change in structure (which is ca. 0.01-0.02 Å). The effect of exchange correlation functional choice (i.e., PBE vs. B3LYP) used in tuning with the LACVP* basis set is also compared on Fe(II)(CO)₆ and Ru(II)(CO)₆ HS and LS states. Average (avg) metal-ligand bond lengths (M-L dist.) are shown in Å along with the min-max range over all six individual bond lengths unless all bonds are the same in which case the range is indicated as '--'.

Geometric effects				
TMC	B3LYP/LACVP*		B3LYP/ZORA-def2-TZVP	
	M-L dist. (Å)		M-L dist. (Å)	
	avg	range	avg	range
LS Ru(II)(CO) ₆	2.057	--	2.052	--
HS Ru(II)(CO) ₆	2.479	2.429-2.505	2.457	2.409-2.540
LS Fe(II)(CO) ₆	1.942	--	1.937	--
HS Fe(II)(CO) ₆	2.330	2.314-2.361	2.311	2.289-2.322
3d:4d $S(\Delta E_{\text{H-L}})$ ratio (S in kcal/mol-HFX)	1.8 (-170 / -95)		1.6 (-244 / -156)	
GGA xc effects				
TMC	B3LYP/LACVP*		PBE/LACVP*	
Ru(II)(CO) ₆ $S(\Delta E_{\text{H-L}})$ kcal/mol-HFX	-95		-81	
Fe(II)(CO) ₆ $S(\Delta E_{\text{H-L}})$ kcal/mol-HFX	-170		-183	

Table B-2. Geometric and electronic criteria adapted from prior work¹ for excluding TMCs from data set with loose geometric cutoffs applied during resubmission indicated in parentheses.

The coordination number must be preserved as 6 for all mononuclear octahedral transition metal complexes (i.e., as judged through interatomic distances being within 1.37x the sum of the respective elements' covalent radii). First coordination shell metrics include the mean and maximum (max.) deviation in the angle ($\Delta\theta(\text{C}_i\text{-M-C}_j)$) with

coordinating atoms (C_i or C_j) and the metal from the expected values of 90° or 180° as well as the maximum overall difference between metal-coordinating atom bond lengths over all ligands and specifically in the equatorial (eq.) plane. Ligand distortion metrics include the maximum root mean square deviation (RMSD) of any atom from the starting structure. For ligands that are expected to be linear, additional checks are applied on the deviation of the angle formed by the metal and the first two atoms of the ligand (A, B) from 180° . Tightened checks for homoleptic TMCs are indicated for $\max(\Delta d)$ with a special threshold for singlet states due to our expectation of higher symmetry for these structures.

Geometric criteria			
Coordination number			
6 (6)			
First coordination shell			
mean($\Delta\theta(C_i-M-C_j)$) 12° (16°)	max($\Delta\theta(C_i-M-C_j)$) 22.5° (27.0°)	max(Δd) 1.00 Å (1.25 Å) homoleptic: 0.4 Å (0.2 Å for singlets)	max(Δd_{eq}) 0.35 Å (0.45 Å)
Ligand distortion metrics			
max(RMSD) 0.30 Å (0.40 Å)		mean($\Delta\theta(M-A-B)$) 20° (30°)	max($\Delta\theta(M-A-B)$) 28° (40°)
Electronic criteria			
Metal spin		1.0 μ_B	
Deviation of $\langle S^2 \rangle$		1.0 μ_B	

Table B-3. Number of $3d$ or $4d$ TMCs eliminated at each sequential filtering step along with how many are retained, and the cumulative total is listed as overall. The filtering steps were: i) convergence (completion, passing loose thresholds upon 5 24 hour job resubmissions), ii) geometry metrics, iii) deviations of $\langle S^2 \rangle$ from the expected value by more than $1\mu_B$, and iv) deviations of metal spin of more than $1\mu_B$.

All calculations were started from converged $a_{\text{HF}} = 0.20$ (i.e., B3LYP results), and then subsequent calculations were attempted at up to a total of six either sequentially increasing (i.e., $a_{\text{HF}} = 0.25$ and 0.30) or decreasing (i.e., $a_{\text{HF}} = 0.15, 0.10, 0.05,$ and 0.00) exchange fractions. If the initial B3LYP result did not converge, none of the subsequent calculations were attempted. If one of the subsequent calculations was attempted but failed in either direction, it was also abandoned. This procedure refers only to convergence failures of the geometry optimization or failing loose geometry checks during resubmissions (i.e., step i). All points were checked for steps ii-iv at the end, and failure of one data point to pass the criteria in ii, iii, or iv does not mean that other HF exchange fractions would necessarily be affected.

Check Performed	3d TMCs		4d TMCs	
	Retained	Eliminated at step	Retained	Eliminated at step
Start	26,877	--	19,100	--
Converged	26,326	531	18,743	357
Geometry	23,813	2,513	16,516	2,227
$\langle S^2 \rangle$	22,881	932	16,196	320
Metal spin	21,699	1,182	15,330	866
Overall	21,699	5,178	15,330	3,770

Table B-4. Summary of computed sensitivities: all $S(3d)$ or $S(4d)$ results that were obtained (“results”) along with those that were eliminated for having fewer than 4 points or due to the inability to eliminate slope changes between two points. All lines were retained regardless of R^2 values, but any points where LOOCV errors for a point were 5 kcal/mol or higher were eliminated.

The cases for which R^2 was below 0.99 are indicated in the $R^2 < 0.99$ subset. The retained results after these filtering steps for both $3d$ and $4d$ TMCs are shown along with the number of matched-ligand pairs that can be generated by identifying metals in the same oxidation/spin states but different principal quantum numbers. These numbers are generally limited by the smaller size of the $4d$ TMC set. Specifics of the filtering procedure: we required that at least four of seven points converged and passed data fidelity checks. We then carried out leave one out cross validation (LOOCV) of the linear fit relationship of a_{HF} with the relevant property (e.g., $\Delta E_{\text{H-L}}$). Individual points

with LOOCV errors greater than 5 kcal/mol were removed along with any points that, upon removal, would increase the R^2 value of the remaining points to above 0.99. If at least four points remained after this first step, the slope was evaluated between each adjacent pair of points, and these slopes were compared for changes in sign as an indication of discontinuous points that had passed the LOOCV check. If the pair of points that changed the sign of the slope included one on the two extrema (i.e., lowest a_{HF} or highest a_{HF} value), they were removed. For any cases that had at least four points after these two filtering steps, a best-fit line and R^2 value was computed on the remaining points. The majority ($> 90\%$) of points after these steps had R^2 values of 0.99 or higher. To ensure fair comparison of properties of $3d$ and $4d$ TMCs, sets formed from matching pairs of ligand fields, oxidation states, and electron configurations across $3d$ and $4d$ metals were obtained. The size of this final data set was primarily limited by the smaller size of the valid $4d$ TMC dataset.

Property	3d TMCs					4d TMCs					# 3d-4d pairs
	results	< 4 pts.	slope sign	retain	$R^2 < 0.99$ subset	results	< 4 pts.	slope sign	retain	$R^2 < 0.99$ subset	
$\Delta E_{\text{H-L}}$	703	101	2	600	15	399	113	2	284	10	239
$\Delta E_{\text{I-L}}$	1,072	80	35	957	152	850	180	25	645	91	488
$\Delta E_{\text{H-I}}$	676	98	4	574	37	361	101	5	255	10	209

Table B-5. The ratio of S for $\Delta E_{\text{I-L}}$ to $\Delta E_{\text{H-L}}$ or $\Delta E_{\text{H-I}}$ to $\Delta E_{\text{H-L}}$ for $3d$ and $4d$ TMCs for all 154 TMCs for which all 3 quantities are available, and all sensitivities are negative. The average, standard deviation, minimum, maximum, and range for each ratio is reported.

	Pairs	3d IS-LS/HS-LS					4d IS-LS/HS-LS				
		avg.	std.	min.	max.	range	avg.	std.	min.	max.	range
Cr(II)	11	0.24	0.03	0.20	0.33	0.13	0.36	0.06	0.29	0.48	0.19
Fe(II)	35	0.38	0.07	0.21	0.50	0.29	0.33	0.09	0.12	0.48	0.36
Fe(III)	30	0.42	0.06	0.27	0.51	0.24	0.31	0.11	0.10	0.56	0.47
Mn(II)	36	0.41	0.06	0.28	0.52	0.24	0.38	0.08	0.21	0.72	0.51
Mn(III)	42	0.37	0.05	0.25	0.58	0.33	0.48	0.09	0.31	0.69	0.38
Overall	154	0.38	0.07	0.20	0.58	0.39	0.38	0.11	0.10	0.72	0.62
		3d HS-IS/HS-LS					4d HS-IS/HS-LS				
		avg.	std.	min.	max.	range	avg.	std.	min.	max.	range
Cr(II)	11	0.76	0.04	0.67	0.83	0.17	0.65	0.07	0.52	0.76	0.23
Fe(II)	35	0.64	0.07	0.52	0.79	0.27	0.67	0.09	0.52	0.86	0.34
Fe(III)	30	0.59	0.07	0.50	0.82	0.32	0.69	0.09	0.53	0.90	0.38
Mn(II)	36	0.59	0.10	0.47	1.09	0.62	0.63	0.08	0.46	0.87	0.42
Mn(III)	42	0.63	0.05	0.45	0.74	0.29	0.53	0.10	0.35	0.81	0.45
Overall	154	0.62	0.09	0.45	1.09	0.63	0.62	0.11	0.35	0.90	0.55

Table B-6. Summary of sensitivities (in kcal/mol·HFX) for pairs of 3*d* and 4*d* TMCs grouped by metal and oxidation state. The total number of 3*d*/4*d* pairs is indicated along with the average, standard deviation, minimum, maximum, and range for the 3*d* or 4*d* TMC subsets. A best-fit line for the relationship between the sensitivities and the associated R^2 value are also shown.

$S(\Delta E_{H-L})$														
	3 <i>d</i>						4 <i>d</i>					$S(4d) = mS(3d) + C$		
	Pairs	avg.	std.	min.	max.	range	avg.	std.	min.	max.	range	m	C	R^2
Cr(II)	16	-72.2	22.4	-120.1	-33.4	86.7	-45.2	21.9	-74.6	23.2	97.8	0.55	-5.11	0.32
Mn(III)	51	-73.1	18.1	-134.6	-37.2	97.4	-48.5	13.3	-89.3	-27.5	61.9	0.55	-8.24	0.56
Mn(II)	72	-118.9	31.3	-187.1	-62.2	124.8	-76.6	20.9	-137.9	-40.1	97.8	0.53	-13.96	0.63
Fe(III)	36	-80.6	17.0	-120.5	-53.7	66.9	-42.2	19.5	-75.2	45.3	120.5	0.88	28.67	0.59
Fe(II)	64	-110.4	31.6	-175.2	-37.4	137.7	-62.8	16.2	-95.6	-33.5	62.0	0.41	-17.17	0.65
Overall	239	-97.9	33.2	-187.1	-33.4	153.7	-59.6	22.5	-137.9	45.3	183.2	0.55	-5.63	0.67
$S(\Delta E_{H+})$														
	3 <i>d</i>						4 <i>d</i>					$S(4d) = mS(3d) + C$		
	Pairs	avg.	std.	min.	max.	range	avg.	std.	min.	max.	range	m	C	R^2
Cr(II)	12	-55.9	13.7	-77.0	-21.5	55.5	-26.7	18.8	-41.4	32.6	74.0	1.13	36.71	0.68
Mn(III)	84	-52.4	15.8	-99.4	-24.2	75.2	-28.7	12.5	-71.1	-12.3	58.8	0.59	2.16	0.55
Mn(II)	38	-68.1	15.1	-110.1	-41.6	68.5	-51.5	15.7	-97.1	-27.6	69.5	0.55	-14.37	0.27
Fe(III)	31	-46.1	7.6	-64.2	-34.4	29.8	-29.4	8.3	-50.0	-15.5	34.5	0.78	6.28	0.50
Fe(II)	44	-73.6	23.1	-157.6	-39.7	117.9	-44.1	15.2	-97.5	-24.6	72.9	0.54	-4.48	0.67
Overall	209	-59.0	19.4	-157.6	-21.5	136.1	-36.1	16.6	-97.5	32.6	130.1	0.65	2.36	0.58
$S(\Delta E_{L-})$														
	3 <i>d</i>						4 <i>d</i>					$S(4d) = mS(3d) + C$		
	Pairs	avg.	std.	min.	max.	range	avg.	std.	min.	max.	range	m	C	R^2
Cr(III)	46	-25.2	11.9	-59.8	-6.7	53.1	-20.8	7.4	-41.5	-9.1	32.4	0.42	-10.14	0.46
Cr(II)	88	-18.9	11.8	-63.7	13.7	77.4	-15.5	10.1	-38.7	17.3	56.0	0.16	-12.46	0.04
Mn(III)	77	-31.1	13.3	-81.8	-14.7	67.2	-26.1	17.0	-142.9	-13.6	129.3	0.21	-19.56	0.03
Mn(II)	43	-47.7	20.4	-88.0	-18.2	69.8	-25.5	17.0	-85.1	-3.8	81.3	0.53	-0.14	0.40
Fe(III)	80	-43.5	19.8	-140.7	-18.6	122.1	-15.2	12.6	-73.3	49.7	123.0	0.27	-3.24	0.18
Fe(II)	52	-39.7	23.0	-98.4	17.3	115.6	-18.4	13.3	-42.6	29.5	72.1	0.18	-11.21	0.10
Co(III)	55	-29.5	16.4	-115.5	-4.5	111.0	7.1	8.8	-16.8	36.8	53.6	0.06	8.86	0.01
Co(II)	47	-63.3	24.8	-100.0	-13.2	86.7	-25.1	20.9	-51.4	46.4	97.8	0.64	15.13	0.57
Overall	488	-36.3	22.0	-140.7	17.3	158.0	-18.6	16.3	-142.9	49.7	192.6	0.28	-8.46	0.14

Table B-7. Outliers where the $S(4d)$ value exceeds the $S(3d)$ value, both shown in kcal/mol-HFX. The R^2 value for each sensitivity determination and the difference between the first-row and second-row sensitivity are shown grouped by the type of spin splitting.

4e HS-LS									
$3d$	$4d$	L_1	L_2	symmetry	$S(3d)$	R^2	$S(4d)$	R^2	$S(3d)-S(4d)$
Fe(II)	Ru(II)	NH ₃	H ₂ O	trans	-37.4	1.00	-38.0	1.00	0.6
Mn(II)	Tc(II)	NCCH ₃	CO	5+1	-136.7	1.00	-137.9	0.99	1.2
Mn(II)	Tc(II)	NCCH ₃	CNCH ₃	trans	-114.5	0.99	-118.4	0.97	4.0
Mn(III)	Tc(III)	F ⁻	--	homoleptic	-37.2	1.00	-50.6	0.99	13.5
Cr(II)	Mo(II)	F ⁻	--	homoleptic	-33.4	0.91	-68.6	0.97	35.3
2e HS-IS									
$3d$	$4d$	L_1	L_2	symmetry	$S(3d)$	R^2	$S(4d)$	R^2	$S(3d)-S(4d)$
Mn(II)	Tc(II)	CO	H ₂ O	trans	-57.2	0.97	-76.2	0.99	19.0
Mn(II)	Tc(II)	NCCH ₃	H ₂ O	trans	-51.7	1.00	-97.1	0.96	45.5
2e IS-LS									
$3d$	$4d$	L_1	L_2	symmetry	$S(3d)$	R^2	$S(4d)$	R^2	$S(3d)-S(4d)$
Mn(III)	Tc(III)	H ₂ S	F ⁻	5+1	-20.2	1.00	-35.7	0.98	15.5
Fe(II)	Ru(II)	NH ₃	CN ⁻	trans	3.9	0.98	-13.0	1.00	16.9
Cr(II)	Mo(II)	NH ₃	H ₂ S	cis	-2.8	1.00	-20.3	1.00	17.5
Cr(II)	Mo(II)	NH ₃	CNCH ₃	cis	-12.7	0.98	-32.9	0.90	20.2
Cr(II)	Mo(II)	NH ₃	CO	trans	13.7	1.00	-13.8	1.00	27.5
Cr(II)	Mo(II)	NH ₃	--	homoleptic	-6.4	1.00	-34.1	0.83	27.6
Mn(II)	Tc(II)	NCCH ₃	CNCH ₃	trans	-49.3	0.97	-85.1	0.93	35.8
Mn(III)	Tc(III)	Cl ⁻	NCCH ₃	5+1	-24.7	0.99	-63.3	0.93	38.6
Mn(III)	Tc(III)	Cl ⁻	PH ₃	5+1	-24.4	1.00	-67.3	0.92	42.9
Mn(III)	Tc(III)	Cl ⁻	CNCH ₃	5+1	-31.3	0.99	-77.5	0.95	46.2
Fe(II)	Ru(II)	F ⁻	CN ⁻	trans	17.3	0.96	-37.6	1.00	54.9
Mn(III)	Tc(III)	Cl ⁻	NCCH ₃	trans	-34.6	0.99	-142.9	0.82	108.3

Table B-8. Outliers where the $S(4d)$ or $S(3d)$ value is positive (both shown in kcal/mol-HFX) for 4-electron ΔE_{H-L} and 2-electron ΔE_{H-I} . The R^2 value for each sensitivity determination and the difference between the first-row and second-row exchange sensitivity are shown grouped by the type of spin splitting.

4e HS-LS									
$3d$	$4d$	L_1	L_2	symmetry	$S(3d)$	R^2	$S(4d)$	R^2	$S(3d)-S(4d)$
Cr(II)	Mo(II)	NH ₃	Cl ⁻	cis	-34.0	1.00	23.2	1.00	-57.2

Fe(III)	Ru(III)	F ⁻	--	homoleptic	-56.9	1.00	45.3	1.00	-102.2
2e HS-IS									
3d	4d	L ₁	L ₂	symmetry	S(3d)	R ²	S(4d)	R ²	S(3d)-S(4d)
Cr(II)	Mo(II)	NH ₃	Cl ⁻	cis	-21.5	1.00	32.6	1.00	-54.1

Table B-9. Outliers where the $S(4d)$ or $S(3d)$ value is positive (both shown in kcal/mol-HFX) for 2-electron ΔE_{L-L} . The R^2 value for each sensitivity determination and the sensitivity difference are shown. Only the bottom three cases are $S(3d) > 0$, the rest are $S(4d) > 0$.

3d	4d	L ₁	L ₂	symmetry	S(3d)	R ²	S(4d)	R ²	S(3d)-S(4d)
Co(III)	Rh(III)	NCCH ₃	NH ₃	5+1	-21.1	1.00	19.2	0.84	-40.3
Co(III)	Rh(III)	NCCH ₃	CO	5+1	-37.9	0.93	14.0	0.95	-51.9
Co(III)	Rh(III)	NCCH ₃	CN ⁻	5+1	-35.1	1.00	3.1	0.97	-38.2
Co(III)	Rh(III)	NCCH ₃	H ₂ O	5+1	-18.3	0.99	7.7	0.91	-26.1
Co(III)	Rh(III)	NH ₃	F ⁻	5+1	-18.1	1.00	3.3	0.97	-21.4
Co(III)	Rh(III)	CNCH ₃	CN ⁻	5+1	-36.7	1.00	4.6	0.87	-41.3
Co(III)	Rh(III)	CNCH ₃	F ⁻	5+1	-32.4	1.00	5.3	0.99	-37.7
Co(III)	Rh(III)	H ₂ O	NCCH ₃	5+1	-13.8	0.99	6.4	0.97	-20.1
Co(III)	Rh(III)	H ₂ O	NH ₃	5+1	-20.1	0.99	8.9	0.96	-28.9
Co(III)	Rh(III)	H ₂ O	CO	5+1	-7.9	0.97	11.7	0.98	-19.6
Co(III)	Rh(III)	H ₂ O	CN ⁻	5+1	-34.1	1.00	6.7	0.99	-40.8
Co(III)	Rh(III)	H ₂ O	F ⁻	5+1	-12.2	0.99	6.2	1.00	-18.5
Co(III)	Rh(III)	NH ₃	NCCH ₃	cis	-24.2	1.00	6.6	0.99	-30.8
Co(III)	Rh(III)	CNCH ₃	NCCH ₃	cis	-36.5	1.00	11.8	0.96	-48.3
Co(III)	Rh(III)	H ₂ O	NCCH ₃	cis	-14.5	0.99	7.5	0.94	-21.9
Co(III)	Rh(III)	H ₂ O	NH ₃	cis	-23.6	1.00	7.4	0.93	-31.0
Co(III)	Rh(III)	H ₂ S	Cl ⁻	cis	-115.5	1.00	6.2	1.00	-121.7
Co(III)	Rh(III)	NH ₃	CN ⁻	cis	-33.2	1.00	7.7	0.91	-40.9
Co(III)	Rh(III)	NH ₃	F ⁻	cis	-19.1	0.99	6.1	0.98	-25.2
Co(III)	Rh(III)	H ₂ O	F ⁻	cis	-20.5	1.00	4.9	0.99	-25.4
Co(III)	Rh(III)	NCCH ₃	CNCH ₃	cis	-27.4	0.99	12.5	0.96	-39.9
Co(III)	Rh(III)	H ₂ O	CNCH ₃	cis	-26.2	1.00	8.2	0.90	-34.4
Co(II)	Rh(II)	F ⁻	--	homoleptic	-59.0	1.00	46.4	1.00	-105.4
Co(III)	Rh(III)	H ₂ O	--	homoleptic	-4.5	0.92	5.9	0.89	-10.3
Fe(II)	Ru(II)	CN ⁻	--	homoleptic	13.5	0.91	17.3	1.00	-3.8
Co(III)	Rh(III)	NCCH ₃	Cl ⁻	trans	-19.1	0.89	5.8	1.00	-24.9
Co(III)	Rh(III)	NCCH ₃	CN ⁻	trans	-28.2	0.96	9.8	1.00	-38.1
Co(III)	Rh(III)	NCCH ₃	CNCH ₃	trans	-26.9	1.00	36.8	0.80	-63.7
Co(III)	Rh(III)	NH ₃	Cl ⁻	trans	-13.1	0.97	9.3	1.00	-22.4
Co(III)	Rh(III)	NH ₃	CN ⁻	trans	-27.0	1.00	15.1	0.99	-42.1
Co(III)	Rh(III)	NH ₃	F ⁻	trans	-9.3	0.99	3.7	0.99	-13.0
Co(III)	Rh(III)	NH ₃	H ₂ O	trans	-26.8	1.00	6.4	0.96	-33.2
Co(III)	Rh(III)	CO	CN ⁻	trans	-31.9	1.00	5.5	1.00	-37.4
Co(III)	Rh(III)	CO	CNCH ₃	trans	-22.6	0.93	22.5	0.99	-45.1
Co(III)	Rh(III)	Cl ⁻	CNCH ₃	trans	-41.9	0.94	16.2	0.98	-58.1
Co(III)	Rh(III)	Cl ⁻	PH ₃	trans	-38.8	0.98	12.1	0.98	-50.9
Co(III)	Rh(III)	CN ⁻	NH ₃	trans	-37.9	0.96	9.5	0.94	-47.4
Co(III)	Rh(III)	CN ⁻	F ⁻	trans	-28.9	0.87	13.4	0.97	-42.3
Co(III)	Rh(III)	CN ⁻	PH ₃	trans	-44.2	1.00	19.5	0.98	-63.6
Co(III)	Rh(III)	H ₂ S	Cl ⁻	trans	-25.4	0.96	10.5	1.00	-35.9
Co(III)	Rh(III)	H ₂ S	CN ⁻	trans	-32.2	0.99	12.1	1.00	-44.3
Co(III)	Rh(III)	CNCH ₃	CO	trans	-54.2	1.00	14.4	0.98	-68.7
Co(III)	Rh(III)	CNCH ₃	CN ⁻	trans	-37.8	0.99	8.1	0.98	-45.9

Co(III)	Rh(III)	CNCH ₃	H ₂ O	trans	-52.9	0.91	9.8	0.90	-62.7
Co(III)	Rh(III)	PH ₃	Cl ⁻	trans	-25.0	0.98	15.9	0.99	-40.9
Co(III)	Rh(III)	H ₂ O	CO	trans	-9.3	0.96	17.3	0.99	-26.6
Co(III)	Rh(III)	H ₂ O	CNCH ₃	trans	-21.4	1.00	3.9	0.89	-25.3
Fe(II)	Ru(II)	CN ⁻	NH ₃	trans	-44.4	0.96	29.5	1.00	-74.0
Fe(II)	Ru(II)	F ⁻	CO	trans	-15.6	1.00	9.5	1.00	-25.1
Fe(II)	Ru(II)	CNCH ₃	Cl ⁻	trans	-44.3	1.00	10.9	0.98	-55.1
Fe(III)	Ru(III)	Cl ⁻	CN ⁻	trans	-28.6	1.00	49.7	1.00	-78.3
Fe(II)	Ru(II)	NH ₃	CN ⁻	trans	3.9	0.98	-13.0	1.00	16.9
Cr(II)	Mo(II)	NH ₃	CO	trans	13.7	1.00	-13.8	1.00	27.5
Fe(II)	Ru(II)	F ⁻	CN ⁻	trans	17.3	0.96	-37.6	1.00	54.9

Table B-10. Adiabatic spin-splitting energies, $\Delta E_{\text{H-L}}$, for the 4-electron difference HS-to-LS energies in kcal/mol for 3d TMCs containing CO and H₂O ligands with varied a_{HF} values as indicated in the table alongside the computed sensitivity and the R^2 value from a linear fit. Values in italics have been averaged between neighboring points or extrapolated using the linear fit if not between neighboring points. The symmetry and composition of the complex is indicated in the table.

M	CO	H ₂ O	sym.	a_{HF}							R^2	S
				0.00	0.05	0.10	0.15	0.20	0.25	0.30		
Cr(II)	6	0	homoleptic	2.0	-5.5	-11.7	-19.9	-28.2	-36.4	-41.2	1.00	-148.7
Cr(II)	5	1	5+1	-0.8	-7.4	-13.6	-19.4	-24.7	-29.6	-34.0	1.00	-110.5
Cr(II)	4	2	trans	-1.2	-7.8	-14.0	-18.1	-23.4	-28.2	-34.0	1.00	-106.0
Cr(II)	2	4	cis	-14.0	-19.9	-25.4	-30.5	-35.1	-39.2	-42.9	0.99	-96.3
Cr(II)	2	4	trans	-12.0	-18.8	-25.1	-31.0	-36.2	-40.8	-44.7	0.99	-109.5
Cr(II)	1	5	5+1	-22.9	-28.5	-33.6	-38.2	-42.2	-45.8	-48.9	0.99	-86.4
Cr(II)	0	6	homoleptic	-43.7	-45.6	-47.4	-49.1	-50.8	-52.5	-54.1	1.00	-34.4
Mn(III)	6	0	homoleptic	-16.1	-19.9	-24.4	-28.5	-32.3	-35.8	-39.1	1.00	-76.6
Mn(III)	5	1	5+1	-9.2	-13.5	-18.5	-23.1	-27.4	-31.3	-35.0	1.00	-86.1
Mn(III)	4	2	trans	-6.6	-11.6	-16.3	-20.8	-24.9	-28.8	-32.5	1.00	-86.1
Mn(III)	2	4	cis	-19.8	-24.6	-28.9	-32.9	-36.6	-40.0	-43.2	0.99	-77.7
Mn(III)	2	4	trans	-18.0	-23.1	-27.6	-31.7	-35.5	-39.0	-42.3	0.99	-80.3
Mn(III)	1	5	5+1	-26.2	-30.3	-34.0	-37.4	-40.7	-43.7	-46.6	1.00	-67.6
Mn(III)	0	6	homoleptic	-32.9	-35.3	-38.0	-40.7	-43.3	-45.8	-48.2	1.00	-51.5
Mn(II)	6	0	homoleptic	29.5	19.7	10.5	1.7	-6.6	-14.3	-21.4	1.00	-169.9
Mn(II)	5	1	5+1	20.7	11.1	2.0	-6.5	-14.3	-21.5	-28.1	1.00	-163.0
Mn(II)	4	2	trans	10.8	2.2	-5.9	-13.4	-20.9	-27.5	-34.6	1.00	-150.3
Mn(II)	2	4	cis	-4.0	-12.2	-19.7	-26.7	-33.0	-38.7	-43.7	0.99	-132.4
Mn(II)	2	4	trans	-7.5	-15.3	-22.6	-29.3	-35.4	-40.6	-45.3	0.99	-126.3
Mn(II)	1	5	5+1	-16.1	-23.1	-29.5	-35.4	-40.7	-45.4	-49.5	0.99	-111.4
Mn(II)	0	6	homoleptic	-36.1	-39.7	-43.0	-46.3	-49.1	-51.8	-54.9	1.00	-62.7
Fe(III)	6	0	homoleptic	43.4	37.2	29.5	23.7	18.1	12.6	7.3	1.00	-120.5
Fe(III)	5	1	5+1	33.8	28.5	22.8	17.4	12.2	7.1	2.0	1.00	-105.7
Fe(III)	4	2	trans	24.0	19.5	14.7	10.1	5.6	1.2	-3.2	1.00	-90.5
Fe(III)	2	4	cis	3.0	-0.8	-4.9	-8.8	-12.6	-16.4	-20.2	1.00	-77.1
Fe(III)	2	4	trans	1.0	-3.0	-6.8	-10.3	-13.9	-17.4	-20.9	1.00	-72.3
Fe(III)	1	5	5+1	-6.9	-10.4	-13.8	-17.0	-20.3	-23.5	-26.8	1.00	-66.1
Fe(III)	0	6	homoleptic	-15.5	-18.0	-20.6	-23.2	-25.9	-28.7	-31.6	1.00	-53.7
Fe(II)	6	0	homoleptic	64.8	57.1	47.9	38.9	30.3	22.2	14.6	1.00	-170.0

Fe(II)	5	1	5+1	49.2	39.8	31.0	22.6	14.7	7.3	0.5	1.00	-162.5
Fe(II)	4	2	trans	30.1	26.4	18.5	12.8	2.7	-3.3	-8.0	0.99	-135.2
Fe(II)	2	4	cis	17.2	9.5	2.4	-4.3	-10.4	-15.8	-20.7	0.99	-126.5
Fe(II)	2	4	trans	2.9	-3.3	-9.0	-14.3	-19.1	-23.5	-27.4	0.99	-101.3
Fe(II)	1	5	5+1	2.3	-3.7	-9.4	-14.6	-19.4	-23.7	-27.6	0.99	-99.5
Fe(II)	0	6	homoleptic	-15.1	-18.1	-21.1	-23.9	-26.6	-28.6	-31.1	1.00	-53.4
Co(III)	0	6	homoleptic	5.2	4.3	3.3	2.3	1.1	0.9	-0.9	0.92	-19.6

Table B-11. Adiabatic spin-splitting energies, ΔE_{H-L} , for the 4-electron difference HS-to-LS energies in kcal/mol for 4*d* TMCs containing CO and H₂O ligands with varied a_{HF} values as indicated in the table alongside the computed sensitivity and the R^2 value from a linear fit. Values in italics have been averaged between neighboring points or extrapolated using the linear fit if not between neighboring points. The symmetry and composition of the complex is indicated in the table.

M	CO	H ₂ O	sym.	a_{HF}							R^2	S
				0.00	0.05	0.10	0.15	0.20	0.25	0.30		
Mo(II)	4	2	trans	50.0	47.6	44.8	42.0	39.2	36.5	33.8	1.00	-54.5
Tc(III)	4	2	trans	44.1	40.7	37.3	34.0	30.9	27.8	24.9	1.00	-64.0
Tc(III)	2	4	cis	34.7	31.3	28.1	24.9	19.4	16.4	12.7	0.99	-74.5
Tc(III)	2	4	trans	41.3	37.5	33.8	30.2	26.8	23.6	20.4	1.00	-69.6
Tc(III)	1	5	5+1	29.3	26.5	23.7	20.9	18.1	15.4	12.7	1.00	-55.4
Tc(III)	0	6	homoleptic	4.4	2.7	1.0	-0.7	-2.5	-4.2	-5.1	1.00	-32.8
Tc(II)	6	0	homoleptic	106.6	100.8	95.1	88.4	83.1	78.0	73.0	1.00	-113.3
Tc(II)	4	2	trans	88.1	82.7	77.4	72.3	68.5	62.6	58.0	1.00	-99.7
Tc(II)	2	4	cis	70.4	65.3	60.2	55.2	50.5	45.8	41.4	1.00	-97.0
Tc(II)	2	4	trans	67.3	62.4	57.7	53.1	48.7	44.5	40.4	1.00	-89.6
Tc(II)	1	5	5+1	52.9	48.5	44.2	40.0	36.0	32.0	28.5	1.00	-81.7
Tc(II)	0	6	homoleptic	11.9	9.7	7.5	5.3	3.2	1.2	-0.8	1.00	-42.4
Ru(III)	6	0	homoleptic	108.2	104.4	100.7	96.2	91.8	90.1	86.7	0.99	-72.7
Ru(III)	4	2	trans	88.2	85.5	82.7	79.9	77.1	74.4	71.6	1.00	-55.3
Ru(III)	2	4	cis	64.5	62.1	59.6	57.2	54.7	52.3	49.8	1.00	-48.8
Ru(III)	2	4	trans	66.0	63.7	61.5	59.2	57.0	54.7	52.4	1.00	-45.5
Ru(III)	1	5	5+1	52.3	50.3	48.3	46.4	44.4	42.3	40.2	1.00	-40.3
Ru(III)	0	6	homoleptic	32.6	31.5	30.5	29.6	28.7	27.5	26.3	0.99	-21.2
Ru(II)	6	0	homoleptic	135.3	130.5	125.7	120.9	116.2	111.6	107.0	1.00	-94.5
Ru(II)	4	2	trans	94.1	90.1	86.3	82.3	78.2	74.2	70.4	1.00	-79.5
Ru(II)	2	4	cis	81.7	77.8	73.6	69.3	65.1	60.9	56.8	1.00	-83.7
Ru(II)	2	4	trans	61.2	58.0	54.8	51.4	48.0	44.7	41.5	1.00	-65.9
Ru(II)	1	5	5+1	59.2	56.1	52.8	49.4	46.0	42.6	39.3	1.00	-67.0
Ru(II)	0	6	homoleptic	30.0	28.2	26.2	24.3	22.4	20.4	18.5	1.00	-38.4

Table B-12. Sensitivities, S , in kcal/mol·HFX, for $\Delta E_{\text{H-L}}$ 4-electron difference adiabatic spin-splitting energies computed among pairs of $3d$ and $4d$ TMCs containing CO and H₂O ligands. The symmetry and composition of the complex is indicated in the table.

M(3d)/M(4d)	n CO	6-n H ₂ O	sym.	$S(3d)$	$S(4d)$
Cr(II)/Mo(II)	4	2	trans	-106.0	-54.5
Mn(III)/Tc(III)	4	2	trans	-86.1	-64.0
Mn(III)/Tc(III)	2	4	cis	-77.7	-74.5
Mn(III)/Tc(III)	2	4	trans	-80.3	-69.6
Mn(III)/Tc(III)	1	5	5+1	-67.6	-55.4
Mn(III)/Tc(III)	0	6	homoleptic	-51.5	-32.8
Mn(II)/Tc(II)	6	0	homoleptic	-169.9	-113.3
Mn(II)/Tc(II)	4	2	trans	-150.3	-99.7
Mn(II)/Tc(II)	2	4	cis	-132.4	-97.0
Mn(II)/Tc(II)	2	4	trans	-126.3	-89.6
Mn(II)/Tc(II)	1	5	5+1	-111.4	-81.7
Mn(II)/Tc(II)	0	6	homoleptic	-62.7	-42.4
Fe(III)/Ru(III)	6	0	homoleptic	-120.5	-72.7
Fe(III)/Ru(III)	4	2	trans	-90.5	-55.3
Fe(III)/Ru(III)	2	4	cis	-77.1	-48.8
Fe(III)/Ru(III)	2	4	trans	-72.3	-45.5
Fe(III)/Ru(III)	1	5	5+1	-66.1	-40.3
Fe(III)/Ru(III)	0	6	homoleptic	-53.7	-21.2
Fe(II)/Ru(II)	6	0	homoleptic	-170.0	-94.5
Fe(II)/Ru(II)	4	2	trans	-135.2	-79.5
Fe(II)/Ru(II)	2	4	cis	-126.5	-83.7
Fe(II)/Ru(II)	2	4	trans	-101.3	-65.9
Fe(II)/Ru(II)	1	5	5+1	-99.5	-67.0
Fe(II)/Ru(II)	0	6	homoleptic	-53.4	-38.4

Table B-13. Adiabatic spin-splitting energies, $\Delta E_{\text{H-I}}$, for the 2-electron difference HS-to-IS energies in kcal/mol for $3d$ TMCs containing CO and H₂O ligands with varied a_{HF} values as indicated in the table alongside the computed sensitivity and the R^2 value from a linear fit. Values in italics have been averaged between neighboring points or extrapolated using the linear fit if not between neighboring points. The symmetry and composition of the complex is indicated in the table.

M	CO	H ₂ O	sym.	a_{HF}							R^2	S
				0.00	0.05	0.10	0.15	0.20	0.25	0.30		
Cr(II)	6	0	homoleptic	<i>21.8</i>	17.8	13.1	8.8	4.7	0.9	-2.7	1.00	-81.7
Cr(II)	5	1	5+1	18.3	13.7	9.2	5.1	1.1	-2.5	-5.9	1.00	-80.9
Cr(II)	4	2	trans	17.2	12.5	8.1	5.5	1.6	-2.1	-6.8	1.00	-77.0
Cr(II)	2	4	cis	7.5	3.2	-0.8	-4.5	-8.0	-11.1	-14.0	1.00	-71.8
Cr(II)	2	4	trans	3.1	-1.0	-4.7	-8.1	-11.2	-14.0	-16.5	0.99	-65.1
Cr(II)	1	5	5+1	-0.9	-4.6	-8.0	-11.2	-14.1	-16.7	-19.1	0.99	-60.8
Cr(II)	0	6	homoleptic	-15.1	-16.5	-17.8	-19.1	-20.4	-21.7	-22.9	1.00	-26.1

Mn(III)	6	0	homoleptic	17.0	13.5	10.1	7.0	4.1	1.4	-1.2	1.00	-60.6
Mn(III)	5	1	5+1	13.1	9.7	6.6	3.7	0.9	-1.7	-4.2	1.00	-57.5
Mn(III)	4	2	trans	11.5	8.4	5.4	2.6	0.0	-2.6	-4.9	1.00	-54.8
Mn(III)	2	4	cis	0.8	-1.9	-4.4	-6.7	-8.8	-10.8	-12.7	1.00	-44.6
Mn(III)	2	4	trans	3.6	1.1	-1.3	-3.5	-5.7	-7.8	-9.8	1.00	-44.7
Mn(III)	1	5	5+1	-1.3	-3.6	-5.8	-7.8	-9.7	-11.5	-13.3	1.00	-39.7
Mn(III)	0	6	homoleptic	-7.6	-9.1	-10.9	-12.7	-14.5	-16.3	-18.0	1.00	-35.1
Mn(II)	5	1	5+1	-2.0	-7.6	-12.8	-17.6	-22.0	-26.1	-29.8	1.00	-92.6
Mn(II)	4	2	trans	-16.6	-20.8	-24.6	-28.1	-28.3	-31.3	-35.1	0.97	-57.2
Mn(II)	2	4	cis	-20.1	-24.2	-27.8	-31.1	-34.1	-36.8	-39.3	0.99	-63.5
Mn(II)	2	4	trans	-17.1	-21.6	-25.7	-29.4	-32.7	-35.7	-38.4	0.99	-70.7
Mn(II)	1	5	5+1	-21.0	-25.2	-28.9	-32.3	-35.3	-38.0	-40.5	0.99	-64.6
Mn(II)	0	6	homoleptic	-31.4	-33.8	-36.1	-38.3	-40.4	-42.4	-44.3	1.00	-42.9
Fe(III)	6	0	homoleptic	13.3	10.1	7.0	4.0	1.1	-1.7	-4.5	1.00	-59.0
Fe(III)	5	1	5+1	<i>5.1</i>	2.4	-0.3	-3.0	-5.6	-8.2	-11.0	1.00	-53.3
Fe(III)	4	2	trans	<i>-1.8</i>	<i>-4.3</i>	-6.8	-9.1	-13.6	-15.8	-16.0	0.92	-50.1
Fe(III)	2	4	cis	<i>-6.8</i>	-9.2	-11.7	-14.0	-16.4	-18.7	-21.1	1.00	-47.4
Fe(III)	2	4	trans	-3.3	-6.2	-8.9	-11.6	-14.1	-16.7	-19.2	1.00	-52.7
Fe(III)	1	5	5+1	-11.1	-13.5	-15.8	-18.0	-20.1	-22.3	-24.5	1.00	-44.4
Fe(III)	0	6	homoleptic	-19.4	-21.1	-22.8	-24.6	-26.3	-28.2	-30.0	1.00	-35.3
Fe(II)	6	0	homoleptic	15.0	11.5	6.6	1.8	-2.7	-6.9	-10.7	1.00	-87.9
Fe(II)	5	1	5+1	13.9	8.2	2.9	-2.1	-6.7	-10.9	-14.7	1.00	-95.4
Fe(II)	4	2	trans	<i>15.2</i>	10.4	4.7	1.2	-3.6	-8.1	-14.1	1.00	-94.9
Fe(II)	2	4	cis	<i>-3.4</i>	-7.1	-11.8	-16.0	-19.6	-22.8	-25.6	0.99	-73.7
Fe(II)	2	4	trans	-0.9	-5.6	-10.1	-14.3	-18.1	-21.6	-24.7	0.99	-79.6
Fe(II)	1	5	5+1	-3.3	-7.8	-12.1	-15.9	-19.5	-22.6	-25.5	0.99	-73.9
Fe(II)	0	6	homoleptic	-17.8	-19.7	-21.5	-23.1	-24.6	-26.0	-27.2	0.99	-31.3
Co(III)	0	6	homoleptic	-7.4	-8.3	-9.1	-10.0	-10.8	-11.7	-12.5	1.00	-16.8

Table B-14. Adiabatic spin-splitting energies, ΔE_{H-I} , for the 2-electron difference HS-to-IS energies in kcal/mol for 4d TMCs containing CO and H₂O ligands with varied a_{HF} values as indicated in the table alongside the computed sensitivity and the R^2 value from a linear fit. Values in italics have been averaged between neighboring points or extrapolated using the linear fit if not between neighboring points. The symmetry and composition of the complex is indicated in the table.

M	CO	H ₂ O	sym.	a_{HF}							R^2	S
				0.00	0.05	0.10	0.15	0.20	0.25	0.30		
Mo(II)	4	2	trans	53.9	52.8	51.1	49.4	47.8	46.1	44.5	1.00	-32.1
Tc(III)	6	0	homoleptic	<i>49.8</i>	<i>48.1</i>	46.4	44.5	42.8	41.1	39.4	1.00	-34.7
Tc(III)	4	2	trans	48.6	46.4	44.3	42.3	40.4	38.5	36.8	1.00	-39.4
Tc(III)	2	4	cis	35.1	33.1	31.2	29.4	27.6	26.0	24.4	1.00	-35.5
Tc(III)	1	5	5+1	30.2	29.0	27.8	26.6	25.4	24.2	23.0	1.00	-24.0
Tc(III)	0	6	homoleptic	19.1	18.6	18.0	17.4	16.8	16.1	15.4	1.00	-12.3
Tc(II)	4	2	trans	43.2	38.9	34.9	31.1	27.2	22.7	20.9	0.99	-76.2
Tc(II)	2	4	trans	25.7	22.2	18.9	15.8	12.8	10.1	7.5	1.00	-60.8
Ru(III)	4	2	trans	<i>43.8</i>	<i>41.3</i>	<i>38.8</i>	36.3	33.7	31.2	28.8	1.00	-50.0
Ru(III)	2	4	cis	<i>22.1</i>	<i>20.6</i>	<i>19.2</i>	17.7	16.2	14.7	13.3	1.00	-29.3
Ru(III)	2	4	trans	<i>30.7</i>	<i>28.6</i>	26.6	24.4	22.3	20.3	18.3	1.00	-41.2

Ru(III)	1	5	5+1	20.4	18.8	17.3	15.6	14.0	12.4	10.9	1.00	-31.6
Ru(III)	0	6	homoleptic	8.8	8.0	7.4	6.7	6.0	5.1	4.2	0.99	-15.5
Ru(II)	6	0	homoleptic	51.5	48.7	46.0	43.2	40.3	37.5	34.8	1.00	-55.6
Ru(II)	4	2	trans	52.2	48.8	45.3	41.7	38.2	34.9	31.7	1.00	-68.4
Ru(II)	2	4	trans	37.9	35.3	32.5	29.7	26.9	24.2	21.5	1.00	-55.0

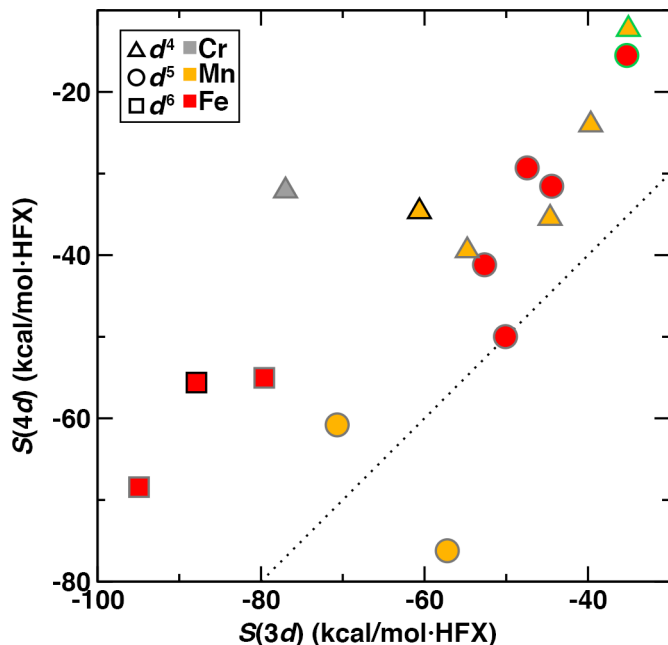


Figure B-1. $S(3d)$ vs $S(4d)$ for ΔE_{H-L} (in kcal/mol·HFX) of all TMCs with CO or H₂O ligands in both oxidation states, colored by element (Cr/Mo in gray, Mn/Tc in orange, or Fe/Ru in red) and with symbols corresponding to formal electron configuration (d^4 in triangles, d^5 in circles, and d^6 in squares). All hexa-aqua complexes are outlined in green, all hexa-carbonyl complexes are outlined in black, and the remaining symbols are outlined in dark gray. A dotted parity line is shown for reference.

Table B-15. Sensitivities, S , in kcal/mol·HFX for ΔE_{L-L} and ΔE_{H-L} 2-electron difference adiabatic spin-splitting energies computed among pairs of 3d and 4d TMCs containing CO and H₂O ligands. The symmetry and composition of the complex is indicated in the table.

		$S(\Delta E_{L-L})$			
M(3d)/M(4d)	n CO	6-n H ₂ O	sym.	$S(3d)$	$S(4d)$
Cr(III)/Mo(III)	4	2	trans	-24.7	-24.7
Cr(III)/Mo(III)	2	4	trans	-22.1	-33.9
Cr(II)/Mo(II)	6	0	homoleptic	-63.7	-12.0
Cr(II)/Mo(II)	4	2	trans	-29.0	-22.4
Cr(II)/Mo(II)	2	4	cis	-26.2	-19.6
Cr(II)/Mo(II)	2	4	trans	-44.6	-29.1
Cr(II)/Mo(II)	1	5	5+1	-25.6	-22.4

Cr(II)/Mo(II)	0	6	homoleptic	-8.3	-9.1
Mn(III)/Tc(III)	4	2	trans	-31.3	-24.7
Mn(III)/Tc(III)	2	4	cis	-33.1	-27.0
Mn(III)/Tc(III)	1	5	5+1	-27.9	-31.3
Mn(III)/Tc(III)	0	6	homoleptic	-16.5	-22.5
Mn(II)/Tc(II)	2	4	trans	-55.6	-28.8
Fe(III)/Ru(III)	2	4	cis	-30.1	-18.5
Fe(III)/Ru(III)	4	2	trans	-19.6	-4.3
Fe(III)/Ru(III)	1	5	5+1	-21.6	-6.5
Fe(III)/Ru(III)	0	6	homoleptic	-18.4	-4.2
Fe(II)/Ru(II)	6	0	homoleptic	-82.1	-37.2
Fe(II)/Ru(II)	4	2	trans	-43.6	-9.7
Fe(II)/Ru(II)	2	4	trans	-21.7	-10.9
Co(III)/Rh(III)	6	0	homoleptic	-48.6	-16.8
Co(III)/Rh(III)	2	4	trans	-9.3	17.3
Co(III)/Rh(III)	1	5	5+1	-7.9	11.7
Co(III)/Rh(III)	0	6	homoleptic	-4.5	5.9
Co(II)/Rh(II)	4	2	trans	-95.4	-50.6
Co(II)/Rh(II)	2	4	trans	-69.9	-36.6
$S(\Delta E_{H-I})$					
Cr(II)/Mo(II)	4	2	trans	-77.0	-32.1
Mn(III)/Tc(III)	6	0	homoleptic	-60.6	-34.7
Mn(III)/Tc(III)	4	2	trans	-54.8	-39.4
Mn(III)/Tc(III)	2	4	cis	-44.6	-35.5
Mn(III)/Tc(III)	1	5	5+1	-39.7	-24.0
Mn(III)/Tc(III)	0	6	homoleptic	-35.1	-12.3
Mn(II)/Tc(II)	4	2	trans	-57.2	-76.2
Mn(II)/Tc(II)	2	4	trans	-70.7	-60.8
Fe(III)/Ru(III)	4	2	trans	-50.1	-50.0
Fe(III)/Ru(III)	2	4	cis	-47.4	-29.3
Fe(III)/Ru(III)	2	4	trans	-52.7	-41.2
Fe(III)/Ru(III)	1	5	5+1	-44.4	-31.6
Fe(III)/Ru(III)	0	6	homoleptic	-35.3	-15.5
Fe(II)/Ru(II)	6	0	homoleptic	-87.9	-55.6
Fe(II)/Ru(II)	4	2	trans	-94.9	-68.4
Fe(II)/Ru(II)	2	4	trans	-79.6	-55.0

Table B-16. Adiabatic spin-splitting energies, ΔE_{I-L} , for the 2-electron difference IS-to-LS energies in kcal/mol for 3d TMCs containing CO and H₂O ligands with varied a_{HF} values as indicated in the table alongside the computed sensitivity, S , and the R^2 value from a linear fit. Values in italics have been averaged or extrapolated using the linear fit if not between neighboring points. The symmetry and composition of the complex is indicated in the table.

M	CO	H ₂ O	sym.	a_{HF}							R^2	S
				0.00	0.05	0.10	0.15	0.20	0.25	0.30		
Cr(III)	2	4	trans	-19.5	-20.9	-22.2	-23.3	-24.3	-25.3	-26.2	0.99	-22.1

Cr(III)	4	2	trans	-18.1	-19.7	-21.1	-22.4	-23.5	-24.6	-25.5	0.99	-24.7
Cr(III)	5	1	5+1	-19.4	-20.8	-22.0	-23.1	-24.2	-25.2	-26.1	0.99	-22.2
Cr(II)	0	6	homoleptic	-28.7	-29.1	-29.5	-30.0	-30.4	-30.8	-31.2	1.00	-8.3
Cr(II)	1	5	5+1	-22.1	-23.9	-25.5	-26.9	-28.1	-29.0	-29.8	0.98	-25.6
Cr(II)	2	4	cis	-21.5	-23.2	-24.6	-26.0	-27.1	-28.1	-29.4	0.99	-26.2
Cr(II)	2	4	trans	-15.1	-17.8	-20.3	-22.8	-25.0	-26.8	-28.2	0.99	-44.6
Cr(II)	4	2	trans	-18.4	-20.3	-22.1	-23.5	-25.0	-26.1	-27.1	0.99	-29.0
Cr(II)	5	1	5+1	-19.7	-21.1	-22.9	-24.5	-25.8	-27.0	-28.1	0.99	-27.8
Cr(II)	6	0	homoleptic	-21.9	-23.2	-24.8	-34.6	-36.0	-37.3	-38.5	0.90	-63.7
Mn(III)	0	6	homoleptic	-25.3	-26.2	-27.1	-28.0	-28.8	-29.5	-30.2	1.00	-16.5
Mn(III)	1	5	5+1	-24.9	-26.7	-28.3	-29.7	-31.0	-32.2	-33.3	0.99	-27.9
Mn(III)	2	4	cis	-20.6	-22.7	-24.5	-26.2	-27.8	-29.2	-30.5	0.99	-33.1
Mn(III)	2	4	trans	-22.6	-24.2	-26.4	-28.2	-29.8	-31.2	-32.5	0.99	-32.7
Mn(III)	4	2	trans	-18.1	-20.0	-21.8	-23.4	-24.9	-26.3	-27.5	1.00	-31.3
Mn(III)	5	1	5+1	-21.7	-23.2	-25.1	-26.7	-28.2	-29.6	-30.8	0.99	-30.4
Mn(III)	6	0	homoleptic	-32.5	-33.4	-34.5	-35.6	-36.4	-37.2	-37.9	0.99	-17.8
Mn(II)	0	6	homoleptic	-4.7	-5.9	-6.9	-7.8	-8.7	-9.4	-10.3	0.99	-18.6
Mn(II)	1	5	5+1	4.9	2.1	-0.6	-3.1	-5.4	-7.4	-9.0	0.99	-46.8
Mn(II)	2	4	cis	16.1	12.0	8.1	4.4	1.1	-1.9	-4.5	0.99	-68.9
Mn(II)	2	4	trans	9.6	6.3	3.1	0.1	-2.7	-4.9	-6.9	0.99	-55.6
Mn(II)	4	2	trans	27.5	23.0	18.7	14.6	7.5	3.9	0.5	0.99	-93.1
Mn(II)	5	1	5+1	22.8	18.7	14.8	11.1	7.7	4.6	1.7	1.00	-70.4
Fe(III)	0	6	homoleptic	4.0	3.1	2.3	1.4	0.4	-0.6	-1.6	1.00	-18.4
Fe(III)	1	5	5+1	4.3	3.1	2.0	1.0	-0.1	-1.2	-2.3	1.00	-21.6
Fe(III)	2	4	cis	10.0	8.3	6.8	5.2	3.7	2.3	0.9	1.00	-30.1
Fe(III)	2	4	trans	4.3	3.2	2.2	1.2	0.3	-0.7	-1.7	1.00	-19.6
Fe(III)	5	1	5+1	29.3	26.1	23.1	20.4	17.8	15.3	13.0	1.00	-54.1
Fe(III)	6	0	homoleptic	30.1	27.1	22.5	19.7	16.9	14.3	11.8	0.99	-61.5
Fe(II)	0	6	homoleptic	2.8	1.6	0.4	-0.8	-2.0	-2.7	-3.9	1.00	-22.2
Fe(II)	1	5	5+1	5.6	4.1	2.7	1.3	0.1	-1.1	-2.1	1.00	-25.7
Fe(II)	2	4	cis	19.1	16.6	14.2	11.7	9.3	7.0	4.9	1.00	-47.8
Fe(II)	2	4	trans	3.8	2.4	1.1	0.0	-1.0	-1.9	-2.7	0.99	-21.7
Fe(II)	4	2	trans	18.1	15.9	13.7	11.6	9.4	7.2	5.0	1.00	-43.6
Fe(II)	5	1	5+1	35.3	31.6	28.1	24.7	21.3	18.2	15.1	1.00	-67.1
Fe(II)	6	0	homoleptic	49.9	45.6	41.3	37.1	33.0	29.1	25.3	1.00	-82.1
Co(III)	0	6	homoleptic	12.8	12.7	12.5	12.3	11.9	11.7	11.4	0.92	-4.5
Co(III)	1	5	5+1	12.0	11.8	11.5	11.2	10.7	10.2	9.7	0.97	-7.9
Co(III)	2	4	trans	11.8	11.6	11.4	10.9	10.4	9.8	9.1	0.96	-9.3
Co(III)	5	1	5+1	39.6	37.5	35.3	33.1	30.9	28.8	26.7	1.00	-43.1
Co(III)	6	0	homoleptic	47.8	45.1	42.6	40.1	37.7	35.4	33.2	1.00	-48.6
Co(II)	0	6	homoleptic	-13.2	-15.0	-17.0	-18.5	-19.8	-20.7	-21.9	0.98	-28.8
Co(II)	1	5	5+1	-0.7	-5.2	-9.4	-13.5	-16.5	-18.9	-20.8	0.98	-67.8
Co(II)	2	4	trans	1.0	-3.3	-7.4	-11.0	-14.6	-17.4	-19.8	0.99	-69.9
Co(II)	4	2	trans	14.9	9.8	4.8	-0.2	-4.9	-9.4	-13.5	1.00	-95.4
Co(II)	5	1	5+1	22.2	16.5	11.0	5.8	0.8	-3.8	-8.2	1.00	-101.5

Table B-17. Adiabatic spin-splitting energies, ΔE_{I-L} , for the 2-electron difference IS-to-LS energies in kcal/mol for 4d TMCs containing CO and H₂O ligands with varied a_{HF} values as indicated in the table alongside the computed sensitivity and the R^2 value from a linear fit. Values in italics have been averaged between neighboring points or extrapolated using the linear fit if not between neighboring points. The symmetry and composition of the complex is indicated in the table.

M	CO	H ₂ O	sym.	a_{HF}							R^2	S
				0.00	0.05	0.10	0.15	0.20	0.25	0.30		
Mo(III)	4	2	trans	-3.2	-4.6	-6.0	-7.2	-8.4	-9.6	-10.6	1.00	-24.7
Mo(III)	2	4	trans	-1.9	-4.0	-5.9	-7.8	-9.4	-10.8	-12.0	0.99	-33.9
Mo(III)	1	5	5+1	-7.2	-9.3	-11.2	-12.8	-14.3	-15.7	-16.9	0.99	-32.0
Mo(III)	0	6	homoleptic	<i>-18.8</i>	-19.4	-20.1	-20.7	-21.3	<i>-21.9</i>	<i>-22.6</i>	1.00	-12.6
Mo(II)	6	0	homoleptic	-15.4	-16.1	-16.7	-17.3	-17.9	<i>-18.5</i>	-19.0	1.00	-12.0
Mo(II)	4	2	trans	-3.9	-5.1	-6.3	-7.5	-8.6	-9.6	-10.6	1.00	-22.4
Mo(II)	2	4	cis	-6.8	-7.9	-8.9	-9.9	-10.8	-11.8	-12.7	1.00	-19.6
Mo(II)	2	4	trans	11.8	10.3	8.8	7.3	5.9	4.5	3.1	1.00	-29.1
Mo(II)	1	5	5+1	1.8	0.6	-0.5	-1.6	-2.7	-3.8	-4.9	1.00	-22.4
Mo(II)	0	6	homoleptic	-17.6	-18.2	-18.7	-19.1	-19.6	-20.0	-20.3	0.99	-9.1
Tc(III)	4	2	trans	-4.5	-5.8	-7.0	-8.3	-9.5	-10.7	-11.9	1.00	-24.7
Tc(III)	2	4	cis	-0.4	-1.8	-3.1	-4.5	-5.8	-7.2	-8.5	1.00	-27.0
Tc(III)	1	5	5+1	-0.9	-2.5	-4.1	-5.7	-7.3	-8.8	-10.3	1.00	-31.3
Tc(III)	0	6	homoleptic	-14.7	-15.9	-17.0	-18.2	-19.2	-20.3	-21.4	1.00	-22.5
Tc(II)	2	4	trans	41.6	40.2	38.8	37.4	35.9	34.4	33.0	1.00	-28.8
Ru(III)	2	4	cis	42.1	41.2	40.4	39.5	38.5	37.5	36.5	1.00	-18.5
Ru(III)	2	4	trans	<i>35.5</i>	<i>35.3</i>	<i>35.1</i>	34.9	34.7	34.4	34.1	0.90	-4.3
Ru(III)	1	5	5+1	31.2	31.2	31.0	30.8	30.4	29.9	29.3	0.91	-6.5
Ru(III)	0	6	homoleptic	<i>23.3</i>	23.1	23.1	22.9	22.7	22.4	22.1	0.96	-4.2
Ru(II)	6	0	homoleptic	<i>83.4</i>	<i>81.5</i>	<i>79.6</i>	77.8	75.9	74.1	72.2	1.00	-37.2
Ru(II)	4	2	trans	41.4	41.3	41.0	40.6	40.0	39.3	38.6	0.96	-9.7
Ru(II)	2	4	trans	23.3	22.8	22.2	21.7	21.1	20.6	20.0	1.00	-10.9
Rh(III)	6	0	homoleptic	<i>79.6</i>	78.8	78.0	77.2	76.3	75.5	74.6	1.00	-16.8
Rh(III)	2	4	cis	34.4	35.8	37.0	38.1	39.1	39.8	<i>40.9</i>	0.99	21.8
Rh(III)	2	4	trans	25.7	26.8	27.8	28.7	29.5	30.3	30.9	0.99	17.3
Rh(III)	1	5	5+1	29.5	30.4	31.1	31.7	32.2	32.7	33.0	0.98	11.7
Rh(III)	0	6	homoleptic	29.3	29.9	30.3	30.7	30.9	31.0	31.0	0.89	5.9
Rh(II)	4	2	trans	55.0	52.4	49.9	47.4	44.8	42.3	39.8	1.00	-50.6
Rh(II)	2	4	trans	32.3	30.9	29.1	27.3	25.4	23.4	21.5	1.00	-36.6

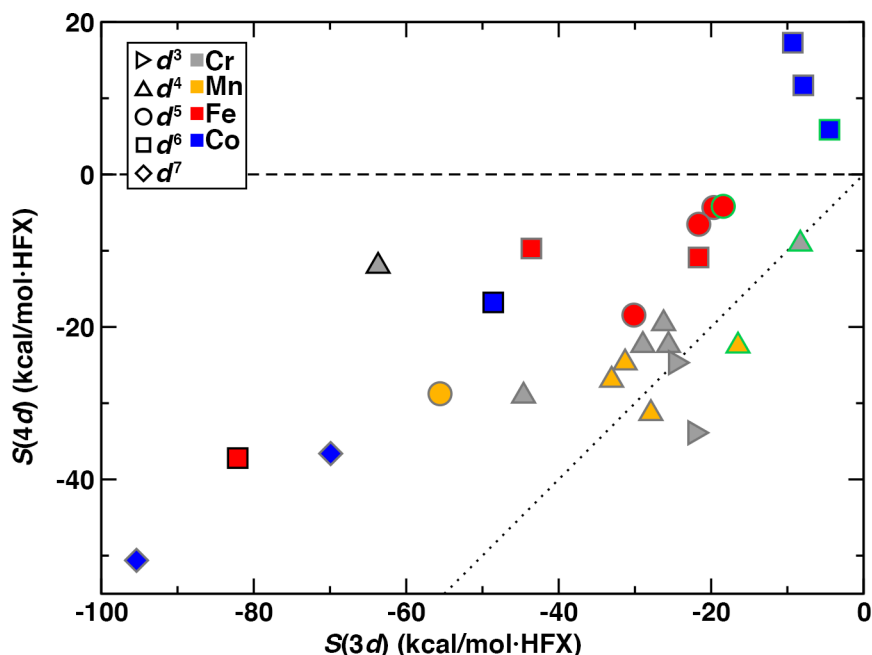


Figure B-2. $S(3d)$ vs $S(4d)$ for ΔE_{L-L} (in kcal/mol·HFX) of all TMCs with CO or H₂O ligands in both oxidation states, colored by element (Cr/Mo in gray, Mn/Tc in orange, Fe/Ru in red, and Co/Rh in blue) and with symbols corresponding to formal electron configuration (d^3 in right-pointing triangles, d^4 in up-pointing triangles, d^5 in circles, d^6 in squares, and d^7 in diamonds). All hexa-aqua complexes are outlined in green, all hexa-carbonyl complexes are outlined in black, and the remaining symbols are outlined in dark gray. A dotted parity line is shown for reference.

Table B-18. Dependence of HS/LS preference based on sign of ΔE_{H-L} for all 239 pairs of 3d and 4d complexes for which ΔE_{H-L} was evaluated. This analysis is completed for different a_{HF} fractions labeled at the top of the table. The $a_{HF} = 0.4$ value is extrapolated from the sensitivity. Each preference is grouped by metal and oxidation state and then summarized at the bottom as overall.

	a_{HF} value						a_{HF} value				
	0.0	0.1	0.2	0.3	0.4		0.0	0.1	0.2	0.3	0.4
3d						4d					
Cr(II) LS	0	0	0	0	0	Mo(II) LS	13	13	13	11	10
Cr(II) HS	16	16	16	16	16	Mo(II) HS	3	3	3	5	6
Mn(III) LS	2	0	0	0	0	Tc(III) LS	44	40	34	29	18
Mn(III) HS	49	51	51	51	51	Tc(III) HS	7	11	17	22	33
Mn(II) LS	29	17	4	0	0	Tc(II) LS	72	72	72	71	70
Mn(II) HS	43	55	68	72	72	Tc(II) HS	0	0	0	1	2
Fe(III) LS	28	22	16	10	6	Ru(III) LS	36	36	36	36	36
Fe(III) HS	8	14	20	26	30	Ru(III) HS	0	0	0	0	0
Fe(II) LS	52	39	27	15	5	Ru(II) LS	64	64	64	64	64
Fe(II) HS	12	25	37	49	59	Ru(II) HS	0	0	0	0	0

Overall LS	111	78	47	25	11	Overall LS	229	225	219	211	198
Overall HS	128	161	192	214	228	Overall HS	10	14	20	28	41

Table B-19. Dependence of GS preference (HS, LS, or IS) for the 247 pairs for which all compatible spin states have been converged. This corresponds to the 155 pairs of $3d$ and $4d$ complexes for which HS/IS/LS states are all converged in addition to the 92 cases of d^3 Cr(III)/Mo(III) or d^7 Co(II)/Rh(II) pairs for which IS/LS states were converged. The variation is shown with a_{HF} fractions, as labeled at the top of the table. The $a_{\text{HF}} = 0.4$ value is extrapolated from the sensitivity. The overall LS/IS/HS count is shown at the bottom of the table along with the LS/IS count for the d^3/d^7 cases.

	a_{HF} value						a_{HF} value				
	0.0	0.1	0.2	0.3	0.4		0.0	0.1	0.2	0.3	0.4
$3d$						$4d$					
Cr(III) LS	2	1	0	0	0	Mo(III) LS	11	10	9	6	5
Cr(III) IS	43	44	45	45	45	Mo(III) IS	34	35	36	39	40
Cr(II) LS	0	0	0	0	0	Mo(II) LS	0	0	0	0	0
Cr(II) IS	9	8	3	1	0	Mo(II) IS	12	12	12	12	12
Cr(II) HS	3	4	9	11	12	Mo(II) HS	0	0	0	0	0
Mn(III) LS	0	0	0	0	0	Tc(III) LS	0	0	0	0	0
Mn(III) IS	22	10	2	2	1	Tc(III) IS	42	42	42	42	42
Mn(III) HS	20	32	40	40	41	Tc(III) HS	0	0	0	0	0
Mn(II) LS	18	12	2	0	0	Tc(II) LS	36	36	36	36	36
Mn(II) IS	0	0	0	0	0	Tc(II) IS	0	0	0	0	0
Mn(II) HS	18	24	34	36	36	Tc(II) HS	0	0	0	0	0
Fe(III) LS	23	18	12	7	5	Ru(III) LS	30	30	30	30	30
Fe(III) IS	0	0	0	0	0	Ru(III) IS	0	0	0	0	0
Fe(III) HS	7	12	18	23	25	Ru(III) HS	0	0	0	0	0
Fe(II) LS	30	21	16	8	3	Ru(II) LS	35	35	35	35	35
Fe(II) IS	1	0	0	0	0	Ru(II) IS	0	0	0	0	0
Fe(II) HS	4	14	19	27	32	Ru(II) HS	0	0	0	0	0
Co(II) LS	38	27	14	10	1	Rh(II) LS	47	47	47	47	47
Co(II) IS	9	20	33	37	46	Rh(II) IS	0	0	0	0	0
d^3/d^7 LS	11	21	33	37	46	d^3/d^7 LS	11	10	9	6	5
d^3/d^7 IS	54	65	78	82	91	d^3/d^7 IS	45	45	45	45	45
Overall LS	111	79	44	25	9	Overall LS	159	158	157	154	153
Overall IS	84	82	83	85	92	Overall IS	88	89	90	93	94
Overall HS	52	86	120	137	146	Overall HS	0	0	0	0	0

Table B-20. Dependence of HS/LS preference based on sign of $\Delta E_{\text{H-L}}$ for the 155 pairs of 3d and 4d complexes for which HS/IS/LS states are all converged with a_{HF} fractions labeled at the top of the table. The $a_{\text{HF}} = 0.4$ value is extrapolated from the sensitivity. This analysis excludes the existence of the IS state and only compares the HS and LS energies. Each preference is grouped by metal and oxidation state and then summarized at the bottom as overall.

	a_{HF} value						a_{HF} value				
	0.0	0.1	0.2	0.3	0.4		0.0	0.1	0.2	0.3	0.4
3d						4d					
Cr(II) LS	0	0	0	0	0	Mo(II) LS	11	11	11	10	9
Cr(II) HS	12	12	12	12	12	Mo(II) HS	1	1	1	2	3
Mn(III) LS	2	0	0	0	0	Tc(III) LS	37	34	29	24	15
Mn(III) HS	40	42	42	42	42	Tc(III) HS	5	8	13	18	27
Mn(II) LS	18	12	2	0	0	Tc(II) LS	36	36	36	36	36
Mn(II) HS	18	24	34	36	36	Tc(II) HS	0	0	0	0	0
Fe(III) LS	23	18	12	7	5	Ru(III) LS	30	30	30	30	30
Fe(III) HS	7	12	18	23	25	Ru(III) HS	0	0	0	0	0
Fe(II) LS	31	21	16	8	3	Ru(II) LS	35	35	35	35	35
Fe(II) HS	4	14	19	27	32	Ru(II) HS	0	0	0	0	0
Overall LS	74	51	30	15	8	Overall LS	149	146	141	135	125
Overall HS	81	104	125	140	147	Overall HS	6	9	14	20	30

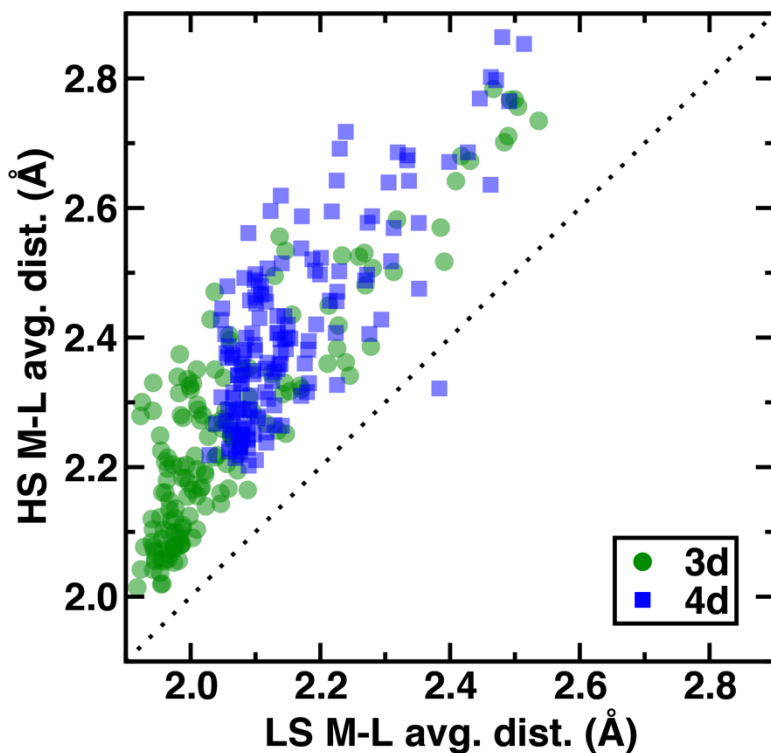


Figure B-3. Absolute LS M-L averaged distance (in Å) vs HS M-L averaged distance (in Å) for 155 pairs of equilibrium structures at $a_{\text{HF}} = 0.2$ in the 4-electron LS-to-HS spin

state comparison for $3d$ (translucent green circles) and $4d$ (translucent blue squares) TMCs. A black dotted parity line is also shown. All $4d$ TMCs generally have longer bond lengths than $3d$ TMCs due to the larger metal covalent radius.

Table B-21. Covalent radii used for $3d$ and $4d$ metals (left) and ligand coordinating elements (right) to obtain relative distances in this work. Recommended covalent radii were obtained from Ref. 2. Where LS and HS covalent radii were provided (e.g., Fe), they were averaged, and where multiple radii based on hybridization (i.e., C) were provided, they were also averaged.

M	r_{cov} (Å)	M	r_{cov} (Å)	L	r_{cov} (Å)	L	r_{cov} (Å)
Cr	1.39	Mo	1.54	C	0.73	He	0.28
Mn	1.50	Tc	1.47	N	0.71	P	1.07
Fe	1.42	Ru	1.46	O	0.66	S	1.05
Co	1.38	Rh	1.42	F	0.57	Cl	1.02

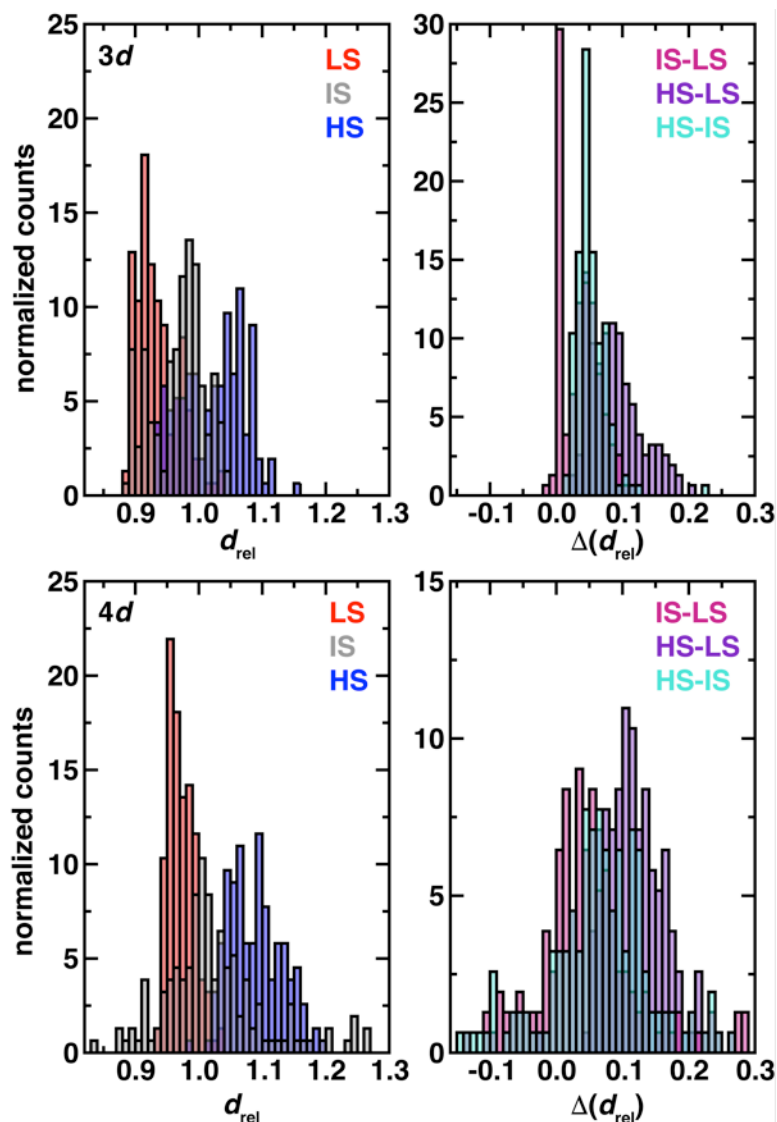


Figure B-4. Normalized histograms for 3d TMCs (top) and 4d TMCs (bottom) with the same x-axis ranges and values for the d_{rel} (left) and difference in d_{rel} (right) plots. The left plots show the averaged (over all 6 metal–ligand bonds) d_{rel} values grouped in translucent histograms by LS (red), IS (gray), and HS (blue) states for 155 (or 247) pairs for which HS, IS, and LS states are all converged (247 is for IS-LS). The right plots show the difference in average d_{rel} values by state of the complexes in translucent normalized histograms: 2-electron IS-LS in magenta, 4-electron HS-LS in purple, and 2-electron HS-IS in cyan.

Table B-22. Average, minimum, maximum, and standard deviation (std.) of relative distances for all HS, IS, and LS states as well as the statistics for the HS-LS, IS-LS, and HS-IS differences compared for 155 pairs of 3d and 4d TMCs for which the LS, IS, and HS states are both defined and successfully converged. The relative distances were

computed with respect to the covalent radii of the metal and coordinating atoms and they were then averaged over all equatorial and axial bond lengths.

3d TMCs						
	LS	HS	HS-LS	IS	IS-LS	HS-IS
avg	0.93	1.02	0.09	0.97	0.04	0.05
min	0.89	0.93	0.03	0.89	-0.02	0.02
max	1.04	1.15	0.20	1.04	0.11	0.22
std.	0.03	0.05	0.04	0.04	0.03	0.02
4d TMCs						
	LS	HS	HS-LS	IS	IS-LS	HS-IS
avg	0.97	1.09	0.12	1.02	0.05	0.06
min	0.94	0.99	-0.03	0.83	-0.14	-0.14
max	1.03	1.20	0.22	1.27	0.29	0.27
std.	0.02	0.04	0.04	0.08	0.08	0.08

Table B-23. Average, minimum, maximum, and standard deviation (std.) of relative distances for IS and LS states as well as the statistics for the IS-LS differences compared for 92 pairs of 3d and 4d TMCs that have d^3 and d^7 electron configurations and therefore only LS and IS states are defined. The relative distances were computed with respect to the covalent radii of the metal and coordinating atoms and they were then averaged over all equatorial and axial bond lengths. The trends in the averages of these LS and IS sets are roughly comparable to those obtained for LS or IS states in the cases where HS states were also valid.

3d TMCs			
	LS	IS	IS-LS
avg	0.99	1.03	0.04
min	0.94	0.96	0.00
max	1.06	1.14	0.18
std.	0.02	0.04	0.03
4d TMCs			
	LS	IS	IS-LS
avg	1.00	1.04	0.04
min	0.93	0.94	-0.05
max	1.10	1.13	0.08
std.	0.04	0.06	0.03

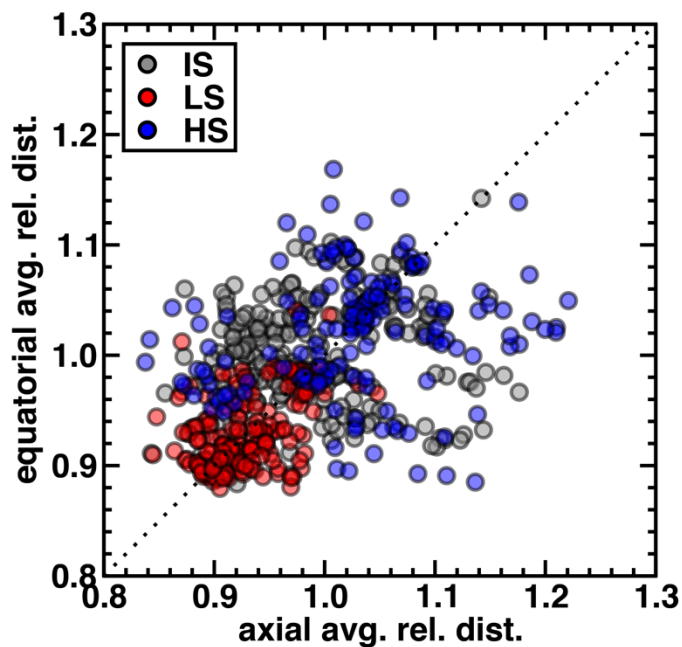


Figure B-5. Relative distance averaged over two axial metal–ligand bonds vs. over four equatorial metal–ligand bonds for $3d$ TMCs: 155 pairs of LS (red translucent circles) or HS (blue translucent circles) as well as 247 pairs of IS (gray translucent circles). A parity line is shown in dotted black, and the plot area is square.

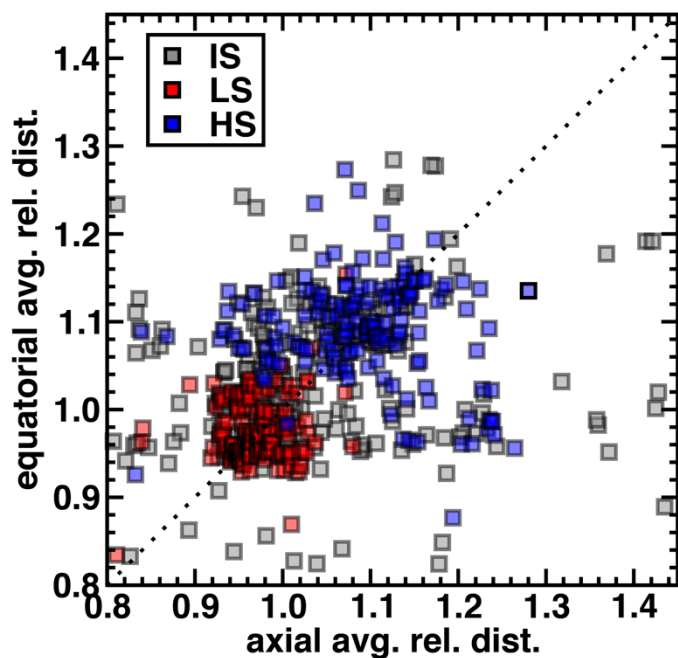


Figure B-6. Relative distance averaged over two axial metal–ligand bonds vs. over four equatorial metal–ligand bonds for $4d$ TMCs: 155 pairs of LS (red translucent squares) or

HS (blue translucent squares) as well as 247 pairs of IS (gray translucent squares). A parity line is shown in dotted black, and the plot area is square.

Table B-24. Best-fit lines ($y = mx + b$) of the difference in spin state relative bond lengths, Δd_{rel} , vs. sensitivity, $S(\Delta E)$ (in kcal/mol·HFX), for $3d$ TMCs and $4d$ TMCs grouped by relative spin states: 155 cases of 4-electron H-L, 155 cases of 2-electron H-I, and 247 cases of 2-electron I-L. The best-fit line for the overall $3d$ or $4d$ TMC set containing 557 points is also shown. The correlation (R^2) for each line is reported. We also report at the bottom of the table the trendlines obtained for the $3d$ or $4d$ $S(\Delta E_{\text{H-L}})$ vs. the LS d_{rel} as well as the line obtained by correlating both data sets together.

	m	b	R^2
3d H-L	-766.22	-25.518	0.86
3d H-I	-657.19	-23.219	0.61
3d I-L	-575.25	-21.925	0.55
3d overall	-772.62	-18.455	0.81
4d H-L	-444.54	-5.305	0.71
4d H-I	10.55	-36.818	0.00
4d I-L	-13.48	-23.038	0.01
4d overall	-126.92	-27.557	0.18
3d H-L vs. LS d_{rel}	276.07	-351.44	0.08
4d H-L vs. LS d_{rel}	422.87	-468.37	0.16
All H-L vs. LS d_{rel}	532.2	-582.6	0.29

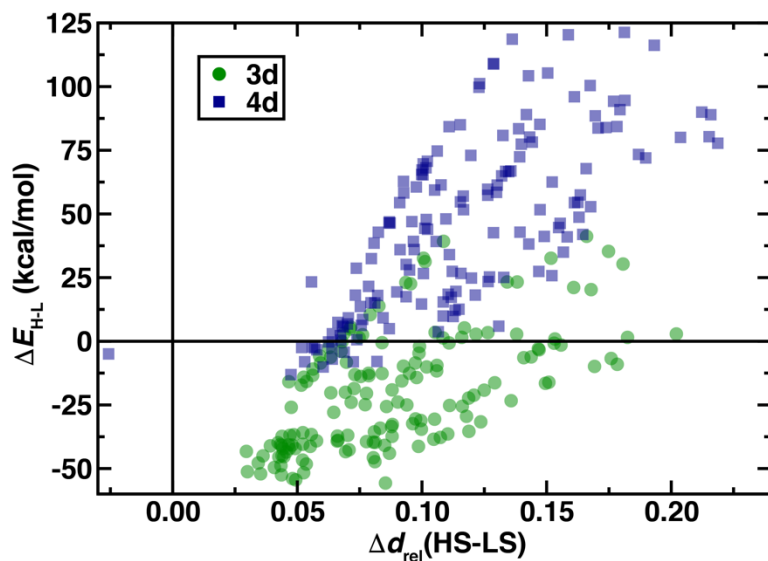


Figure B-7. The difference in average relative metal–ligand distance, Δd_{rel} , for the 4-electron HS-LS state comparison vs. $\Delta E_{\text{H-L}}$ (in kcal/mol) for 155 pairs of HS and LS

states at $a_{\text{HF}} = 0.2$ for $3d$ (translucent green circles) and $4d$ (translucent blue squares) TMCs. A zero axis shows when spin state ordering changes or the sign of the relative metal–ligand bond length difference changes.

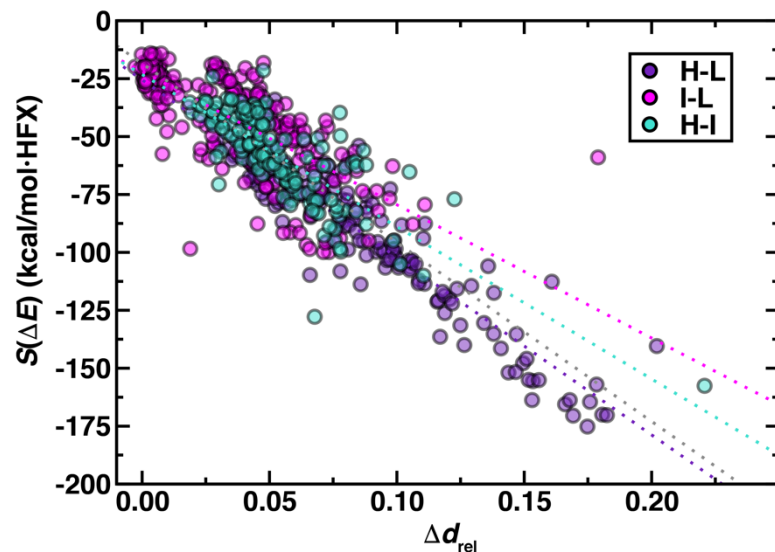


Figure B-8. Difference for $3d$ TMCs in average relative distance between relevant spin states, Δd_{rel} , vs the sensitivity of spin state splitting, $S(\Delta E)$, for the same spin states in kcal/mol·HFX grouped by 155 pairs of 4-electron HS-LS (H-L, purple translucent circles), 247 pairs of 2-electron IS-LS (I-L, magenta translucent circles), and 155 pairs of 2-electron HS-IS (H-I, cyan translucent circles). Best-fit lines have been obtained through each spin state definition and shown as dotted solid lines in the same colors. A gray dotted line has been fit through all data.

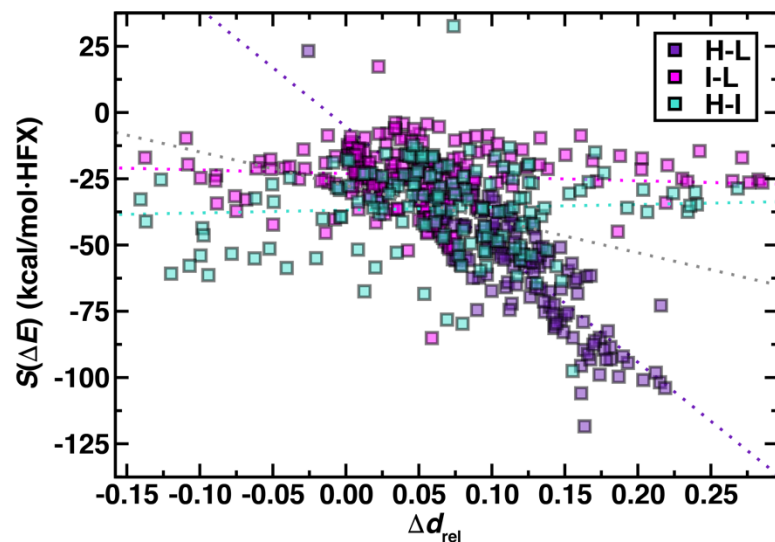


Figure B-9. Difference for $4d$ TMCs in average relative distance between relevant spin states, Δd_{rel} , vs the sensitivity of spin state splitting, $S(\Delta E)$, for the same spin states in kcal/mol-HFX grouped by 155 pairs of 4-electron HS-LS (H-L, purple translucent squares), 247 pairs of 2-electron IS-LS (I-L, magenta translucent squares), and 155 pairs of 2-electron HS-IS (H-I, cyan translucent squares). Best-fit lines have been obtained through each spin state definition and shown as dotted solid lines in the same colors. A gray dotted line has been fit through all data.

Table B-25. Properties of Fe(II)(He)_6 and Ru(II)(He)_6 LS and HS complexes: the M-He bond length (in Å) for the LS singlet and HS quintet states, and the relative bond lengths with respect to covalent radii.

The sensitivities of the total energies $S(E_{\text{LS}})$ or $S(E_{\text{HS}})$ of the LS and HS states, respectively, are also shown in units of kcal/mol-HFX for both the energies and spin splitting evaluated at the LS geometry (LS vertical) and HS geometry (HS vertical). The adiabatic $S(\Delta E_{\text{H-L}})$ is also shown evaluated from the equilibrium geometry of both spin states. The $\Delta E_{\text{H-L}}$ for B3LYP (20% exchange) is shown as well in kcal/mol for each of the three geometries.

		Fe(II)(He)_6				
Property	M-He (Å)	d_{rel}	$S(E_{\text{LS}})$	$S(E_{\text{HS}})$	$\Delta E_{\text{H-L}}$	$S(\Delta E_{\text{H-L}})$
LS vertical	1.84	1.08	25.64	-21.47	-48.3	-47.12
HS vertical	2.11	1.24	16.30	-21.59	-72.8	-37.89
Adiabatic	--	--	25.64	-21.59	-63.8	-47.25
		Ru(II)(He)_6				
Property	M-He (Å)	d_{rel}	$S(E_{\text{LS}})$	$S(E_{\text{HS}})$	$\Delta E_{\text{H-L}}$	$S(\Delta E_{\text{H-L}})$
LS vertical	1.93	1.11	-23.12	-63.00	3.4	-39.88

HS vertical	2.35	1.35	-37.96	-66.24	-50.7	-28.28
Adiabatic	--	--	-23.12	-66.24	-31.2	-43.28

Table B-26. Properties of Fe(II)(CO)₆ and Ru(II)(CO)₆ LS and HS complexes: the M–He bond length (in Å) for the LS singlet and HS quintet states, and the relative bond lengths with respect to covalent radii.

The sensitivities of the total energies $S(E_{LS})$ or $S(E_{HS})$ of the LS and HS states, respectively, are also shown in units of kcal/mol·HFX for both the energies and spin splitting evaluated at the LS geometry (LS vertical) and HS geometry (HS vertical). The adiabatic $S(\Delta E_{H-L})$ is also shown evaluated from the equilibrium geometry of both spin states. The ΔE_{H-L} for B3LYP (20% exchange) is shown as well in kcal/mol for each of the three geometries.

		Fe(II)(CO) ₆				
Property	M-He (Å)	d_{rel}	$S(E_{LS})$	$S(E_{HS})$	ΔE_{H-L}	$S(\Delta E_{H-L})$
LS vertical	1.94	0.90	82.1	-31.9	106.6	-114.1
HS vertical	2.34	1.09	-5.6	-91.6	-19.7	-86.0
Adiabatic	--	--	86.2	-83.8	30.3	-170.0
		Ru(II)(CO) ₆				
Property	M-He (Å)	d_{rel}	$S(E_{LS})$	$S(E_{HS})$	ΔE_{H-L}	$S(\Delta E_{H-L})$
LS vertical	2.06	0.94	-34.0	-84.9	213.3	-50.9
HS vertical	2.48	1.13	-64.1	-129.1	37.5	-64.9
Adiabatic	--	--	-29.8	-124.3	116.2	-94.5

Table B-27. Characteristics of total energy sensitivities for the HS and LS states, $S(E_{HS})$ and $S(E_{LS})$ in kcal/mol·HFX of 3d and 4d TMCs: minimum, maximum, average, and the number that are positive.

Property	3d TMC $S(E_{LS})$	3d TMC $S(E_{HS})$	4d TMC $S(E_{LS})$	4d TMC $S(E_{HS})$
minimum	-684.3	-776.4	-764.8	-809.4
maximum	148.8	38.5	44.6	-24.0
average	-222.4	-323.1	-290.9	-354.6
# > 0	26	1	2	0

References

- ¹ 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*, 13973-13986.
- ² Cordero, B.; Gómez, V.; Platero-Prats, A. E.; Revés, M.; Echeverría, J.; Cremades, E.; Barragán, F.; Alvarez, S., Covalent Radii Revisited. *Dalton Trans.* **2008**, 2832-2838.