

MIT Open Access Articles

Ligand additivity relationships enable efficient exploration of transition metal chemical space

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Arunachalam, Naveen, Gugler, Stefan, Taylor, Michael G, Duan, Chenru, Nandy, Aditya et al. 2022. "Ligand additivity relationships enable efficient exploration of transition metal chemical space." *The Journal of Chemical Physics*, 157 (18).

As Published: 10.1063/5.0125700

Publisher: AIP Publishing

Persistent URL: <https://hdl.handle.net/1721.1/146976>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of use: Creative Commons Attribution 4.0 International license



Ligand additivity relationships enable efficient exploration of transition metal chemical space

Cite as: J. Chem. Phys. **157**, 184112 (2022); <https://doi.org/10.1063/5.0125700>

Submitted: 13 September 2022 • Accepted: 24 October 2022 • Accepted Manuscript Online: 24 October 2022 • Published Online: 14 November 2022

Published open access through an agreement with Massachusetts Institute of Technology

 Naveen Arunachalam,  Stefan Gugler,  Michael G. Taylor, et al.



View Online



Export Citation



CrossMark

ARTICLES YOU MAY BE INTERESTED IN

[Many recent density functionals are numerically ill-behaved](#)

The Journal of Chemical Physics **157**, 174114 (2022); <https://doi.org/10.1063/5.0121187>

[Computing chemical potentials of solutions from structure factors](#)

The Journal of Chemical Physics **157**, 121101 (2022); <https://doi.org/10.1063/5.0107059>

[Perspective on integrating machine learning into computational chemistry and materials science](#)

The Journal of Chemical Physics **154**, 230903 (2021); <https://doi.org/10.1063/5.0047760>

 **The Journal of Chemical Physics** **Special Topics** Open for Submissions [Learn More](#)

Ligand additivity relationships enable efficient exploration of transition metal chemical space

Cite as: J. Chem. Phys. 157, 184112 (2022); doi: 10.1063/5.0125700

Submitted: 13 September 2022 • Accepted: 24 October 2022 •

Published Online: 14 November 2022



View Online



Export Citation



CrossMark

Naveen Arunachalam,¹ Stefan Gugler,¹ Michael G. Taylor,¹ Chenru Duan,^{1,2} Aditya Nandy,^{1,2} Jon Paul Janet,¹ Ralf Meyer,¹ Jonas Oldenstaedt,¹ Daniel B. K. Chu,¹ and Heather J. Kulik^{1,a)}

AFFILIATIONS

¹Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

²Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

^{a)}Author to whom correspondence should be addressed: hjkulik@mit.edu

ABSTRACT

To accelerate the exploration of chemical space, it is necessary to identify the compounds that will provide the most additional information or value. A large-scale analysis of mononuclear octahedral transition metal complexes deposited in an experimental database confirms an under-representation of lower-symmetry complexes. From a set of around 1000 previously studied Fe(II) complexes, we show that the theoretical space of synthetically accessible complexes formed from the relatively small number of unique ligands is significantly (~816k) larger. For the properties of these complexes, we validate the concept of ligand additivity by inferring heteroleptic properties from a stoichiometric combination of homoleptic complexes. An improved interpolation scheme that incorporates information about *cis* and *trans* isomer effects predicts the adiabatic spin-splitting energy to around 2 kcal/mol and the HOMO level to less than 0.2 eV. We demonstrate a multi-stage strategy to discover leads from the 816k Fe(II) complexes within a targeted property region. We carry out a coarse interpolation from homoleptic complexes that we refine over a subspace of ligands based on the likelihood of generating complexes with targeted properties. We validate our approach on nine new binary and ternary complexes predicted to be in a targeted zone of discovery, suggesting opportunities for efficient transition metal complex discovery.

© 2022 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0125700>

I. INTRODUCTION

In recent years, virtual high-throughput screening (VHTS)^{1–8} with first-principles density functional theory (DFT) and machine learning (ML) models^{9–17} has greatly accelerated the discovery of new molecules and materials.^{18–23} 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^{24–29} 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.^{32–34} 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.^{35–38} 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 multi-reference character.^{42–44} 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.^{45–47}

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.^{56–59} 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.

II. COMPUTATIONAL DETAILS

A. 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 (supplementary material, Text S1). 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 (supplementary material, Text S2). 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

(supplementary material, Text S2). This led to a nomenclature for 66 ligand symmetry classes (supplementary material, Text S2).

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. III. 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, denticities, and complex symmetries in the CSD.

B. Electronic structure calculations

DFT geometry optimizations were performed using a development version of TeraChem v1.9.^{67,68} The B3LYP^{69–71} 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

III. RESULTS AND DISCUSSION

A. 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. II). 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 (Fig. 1 and supplementary material, Table S1). In fact, 28%

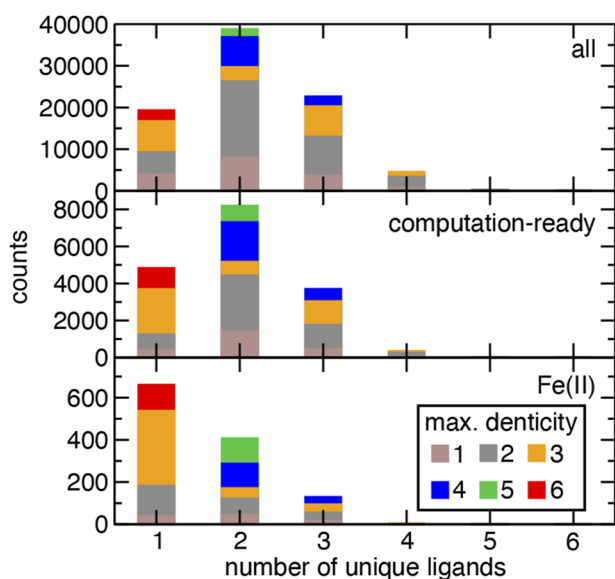


FIG. 1. Stacked, unnormalized histogram of the number of complexes in the CSD grouped by the number of unique ligand types and by the highest denticity of ligands in the complex (monodentate in brown, bidentate in gray, tridentate in orange, tetradentate in blue, pentadentate in green, and hexadentate 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).

of all unique computation-ready complexes are homoleptic, and a majority (76%) contain no more than two ligand types (Fig. 1 and supplementary material, Table S1). 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 (Fig. 1 and supplementary material, Table S1). 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 (Fig. 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.

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 (supplementary material, Text S3).^{81–83} 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 (supplementary material, Text S3). For example, in (L₁)₄(L₂)₂, there is one occurrence of a four-substitution site and one occurrence of a two-substitution site, whereas in (L₁)₂(L₂)₂(L₃)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 I). For up to two unique ligand types, the five unique symmetry classes consist of monoheteroleptic (5+1) M(L₁)₅(L₂)₁ complexes, *trans* symmetric (TS) or *cis* symmetric (CS) M(L₁)₄(L₂)₂ complexes, and *fac* symmetric (FS) or *mer* symmetric (MS) M(L₁)₃(L₂)₃ complexes (Fig. 2 and Table I). Both 5+1 and CS/TS complexes each form $N(N-1)$ complexes for N ligands (i.e., 132 each for $N = 12$, Table I). 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.

Expanding to up to three unique ligand types introduces eight additional symmetry classes (Fig. 3). These include M(L₁)₄L₂L₃ *cis* asymmetric (CA) and *trans* asymmetric (TA) complexes; the three types of M(L₁)₂(L₂)₂(L₃)₂ configurations in equatorial asymmetric (EA), double *cis* symmetric (DCS), or double *trans* symmetric (DTS) symmetries; and three M(L₁)₂(L₂)₃L₃ 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 I). 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 I). Similarly, FA/MAT/MAC complexes can each

TABLE I. 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	$x^4x^1x^1$	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

form $N(N-1)(N-2)/2!$ complexes in two isomers each (i.e., 3960 for $N = 12$, Table I).

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.

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

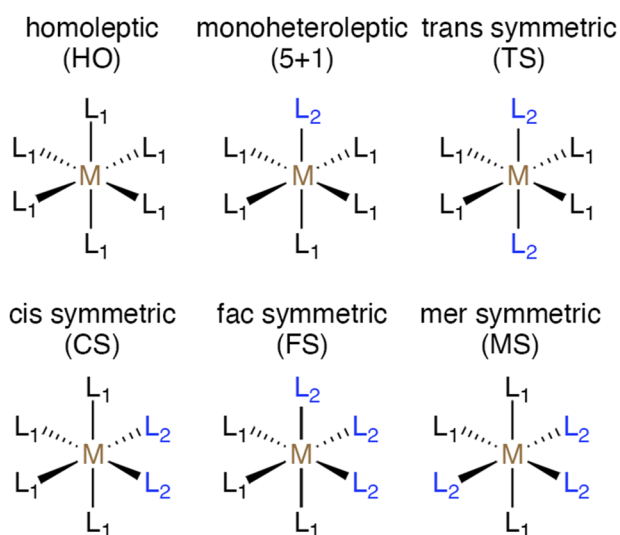


FIG. 2. Symmetry classes for transition metal complexes with up to two unique ligands, L_1 and L_2 , from left to right and top to bottom: homoleptic (HO) $M(L_1)_6$, monoheteroleptic (5+1) $M(L_1)_5L_2$, *trans* symmetric (TS) $M(L_1)_4(L_2)_2$, *cis* symmetric (CS) $M(L_1)_4(L_2)_2$, *fac* symmetric (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 L_1 and L_2 are swapped.

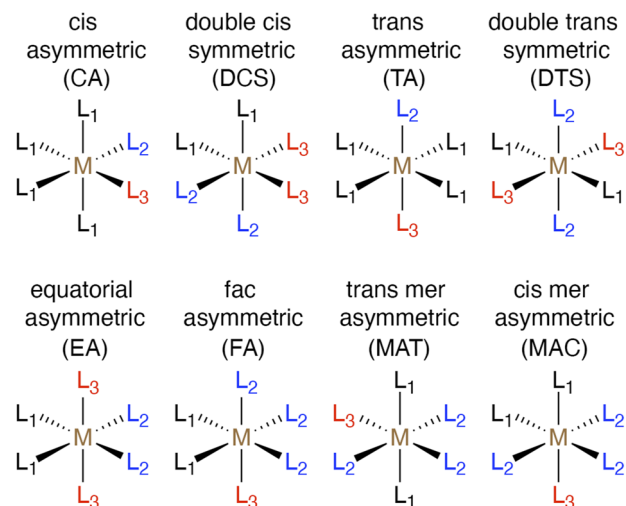


FIG. 3. Symmetry classes for transition metal complexes with three unique ligands, L_1 , L_2 , and L_3 , from left to right and top to bottom: *cis* asymmetric (CA) $M(L_1)_4L_2L_3$, double *cis* symmetric (DCS) $M(L_1)_2(L_2)_2(L_3)_2$, *trans* asymmetric (TA) $M(L_1)_4L_2L_3$, double *trans* symmetric (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) $M(L_1)_2(L_2)_2(L_3)_2$, *trans mer* asymmetric (MAT) $M(L_1)_2(L_2)_2(L_3)_2$, and *cis mer* asymmetric (MAC) $M(L_1)_2(L_2)_2(L_3)_2$. 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.

in comparison to homoleptic complexes. Nevertheless, there are far fewer binary and especially ternary complexes in the CSD (supplementary material, Table S1). 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 (Fig. 4 and supplementary material, Table S2). All ternary complexes are underrepresented, but those with equal stoichiometry (i.e., EA, DCS, or DTS) have very few examples in the CSD (Fig. 4 and supplementary material, Table S2). 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 (Fig. 4 and supplementary material, Table S2).

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 (Fig. 1 and supplementary material, Tables S3–S4). 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 (supplementary material, Table S4). Similar trends hold for all unique computation-ready complexes (supplementary material, Table S3). 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 (supplementary material, Tables S3–S4).

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 (supplementary material, Table S5). 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 (supplementary material, Table S5). While the majority of HO complex monodentate ligands were neutral, over half of the additional binary or ternary ligands have a non-zero charge (supplementary material, Table S6). 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 supplementary material, Table S6). 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 (supplementary material, Table S6). 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

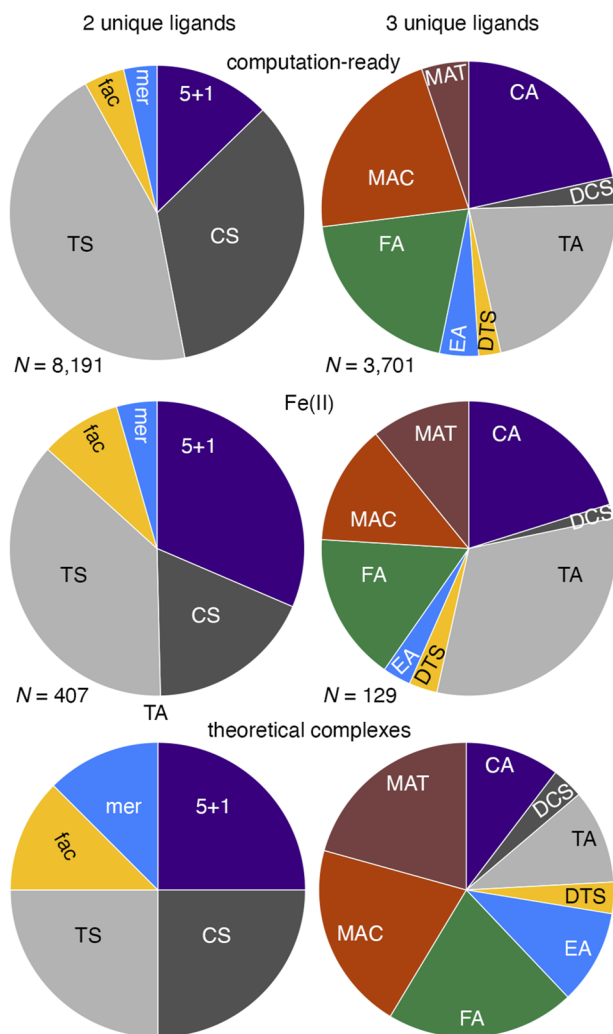


FIG. 4. Percent of all unique mononuclear octahedral transition metal complexes in the CSD with user-defined charges (top, computation-ready) and the Fe(II) subset [middle, Fe(II)], grouped by a symmetry class for cases with two unique ligands (left: 5+1, *cis* symmetric, CS, *trans* symmetric, TS, *fac*, 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.

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.

B. 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 strong-field 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, ΔE_{H-L} , it can be expected that the weak-field 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,

$$E(M(L_1)_x(L_2)_y(L_3)_{6-x-y}) = \frac{x}{6}E(M(L_1)_6) + \frac{y}{6}E(M(L_2)_6) + \frac{6-x-y}{6}E(M(L_3)_6). \quad (1)$$

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 (Fig. 5 and supplementary material, Table S7). Similar observations can be made on combinations with H₂O, CO, and NH₃ (supplementary material, Fig. S1). Nevertheless, there are significant outliers in the interpolated vs actual ΔE_{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 (Fig. 5). For example, the CS complex Fe(II)(H₂O)₄(CO)₂ ΔE_{H-L} is predicted accurately (predicted: -7.9 kcal/mol vs calculated: -9.9 kcal/mol) from homoleptic interpolation (Fig. 5 and supplementary material, Table S8). 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 work³⁴ (Fig. 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 (Fig. 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 (supplementary material, Fig. S2 and Tables S8–S9).

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 (Fig. 6 and supplementary material, Tables S7 and S10–S11). As with spin 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) ΔE_{H-L} and the TS complex was more like the equivalent 5+1 complex, we find

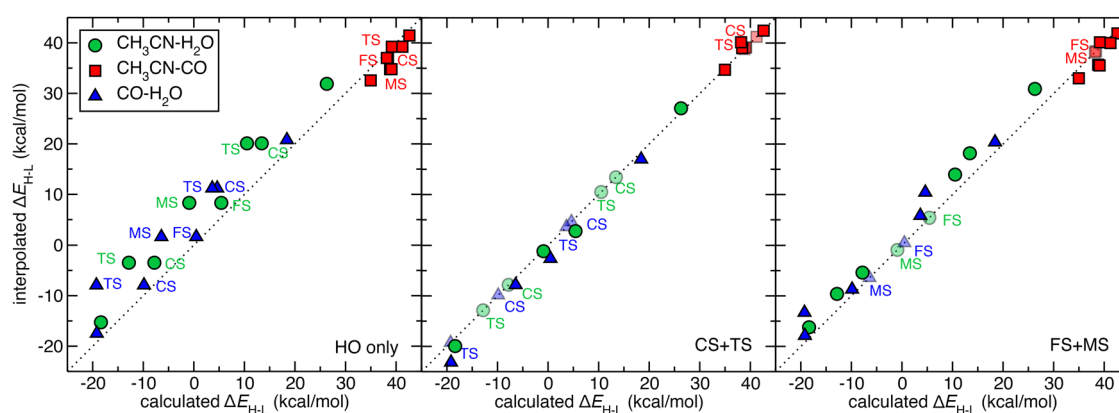


FIG. 5. Calculated vs linearly interpolated ΔE_{H-L} (kcal/mol) for Fe(II) complexes with pairs of any of the three ligands: CH₃CN, H₂O, and CO. From left to right: interpolation between homoleptic complexes (HO only), interpolation using homoleptic complexes and CS and TS complex energies (CS+TS), or 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.

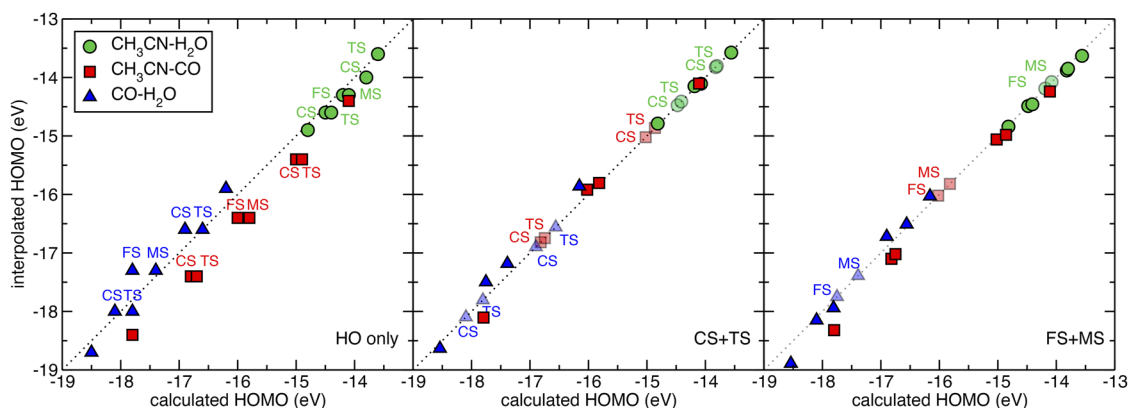


FIG. 6. Calculated vs linearly interpolated HOMO level (eV) for singlet Fe(II) complexes with pairs of any of the three ligands: CH₃CN, H₂O, and CO. From left to right: interpolation between homoleptic complexes (HO only), interpolation using homoleptic complexes and CS and TS complex energies (CS+TS), 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.

more varied results for the HOMO level, with some cases occurring where the TS and FS complexes are less accurately predicted (Fig. 6 and supplementary material, Table S10). Nevertheless, over the full range of data, the outliers in the HOMO level prediction are more modest than what was observed for ΔE_{H-L} . These trends also hold for the complexes with NH₃, H₂O, and CO (supplementary material, Fig. S2 and Tables S10–S11).

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:

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

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

$$E(FS) = \frac{1}{2}(E(CSM(L_1)_4(L_2)_2) + E(CSM(L_2)_4(L_1)_2)), \quad (3)$$

$$E(MS) = \frac{1}{2}(E(TSM(L_1)_4(L_2)_2) + E(TSM(L_2)_4(L_1)_2)), \quad (4)$$

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 (Fig. 5 and

supplementary material, Tables S8–S9). 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 (Fig. 5). The HOMO levels for these and other complexes are also improved (supplementary material, Tables S10–S11). 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 (Fig. 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 (supplementary material, Fig. S1). For the same complex, the HOMO level is equivalently predicted by both interpolation schemes, and most HOMO level estimates are improved (supplementary material, Fig. S2). 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 Fig. 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:

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

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

$$E(\text{CS}) = \frac{1}{3}E(\text{HO}) + \frac{2}{3}E(\text{FSM}(L_1)_3(L_2)_3), \quad (6)$$

$$E(\text{TS}) = \frac{1}{3}E(\text{HO}) + \frac{2}{3}E(\text{MSM}(L_1)_3(L_2)_3), \quad (7)$$

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 (Fig. 5). For example, the higher spin-splitting 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 (Fig. 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 (supplementary material, Fig. S1 and Tables S8–S9). 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 (Fig. 6 and supplementary material, Tables S10–S11). For the set of ligands including ammonia, almost all points are predicted to comparable or slightly improved values (supplementary material, Fig. S2). 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 (Fig. 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 (Fig. 1 and supplementary material, Table S1). 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 CH_3CN , CO, and H_2O , with a large difference between CA and TA complexes of H_2O and CO being treated completely equivalently in this scheme (Fig. 7). Similar differences in TA and CA HOMO levels are also missed in this averaging scheme (Fig. 8). Generally, the CA complex spin-splitting energies are better predicted by the homoleptic averaging than the TA are for both ternary complexes with CH_3CN and with NH_3 (Fig. 7 and supplementary material, Fig. S3 and Tables S12–S13). For the HOMO level, results are more varied, with the HOMO energies of the CH_3CN ternary complexes being underestimated, while those of NH_3 ternary complexes are overestimated (Fig. 8 and supplementary material, Fig. S4 and Tables S14–S15).

Thus, we next investigated interpolation schemes that incorporated 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:

$$E(\text{CAM}(L_1)_4(L_2)_1(L_3)_1) = \frac{1}{3}(E(\text{HOM}(L_1)_6) + E(\text{FSM}(L_1)_3(L_2)_3) + E(\text{FSM}(L_1)_3(L_3)_3)), \quad (8)$$

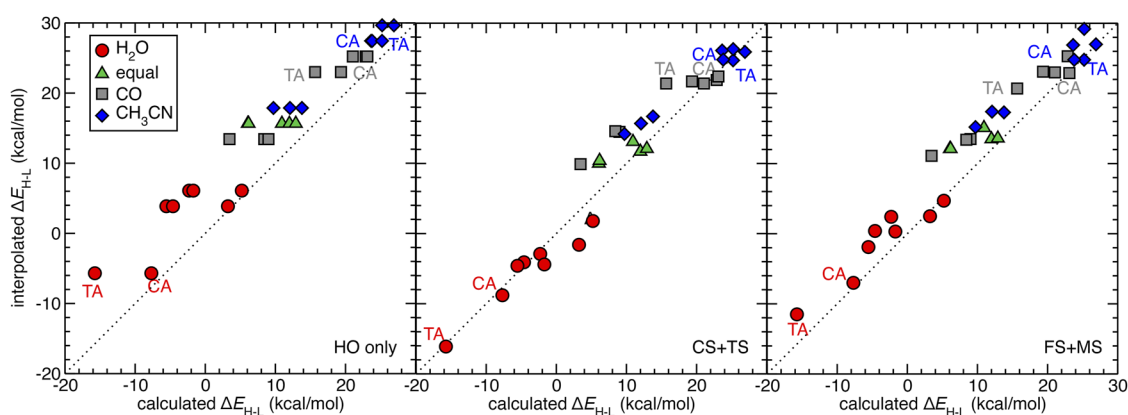


FIG. 7. Calculated vs linearly interpolated $\Delta E_{\text{H-L}}$ (kcal/mol) for Fe(II) complexes with at least one each of three ligands: CH_3CN , H_2O , and CO. From left to right: interpolation from homoleptic complexes (HO only), interpolation using homoleptic complexes and CS and TS complex energies 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: H_2O (red circles), CO (gray squares), and CH_3CN (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.

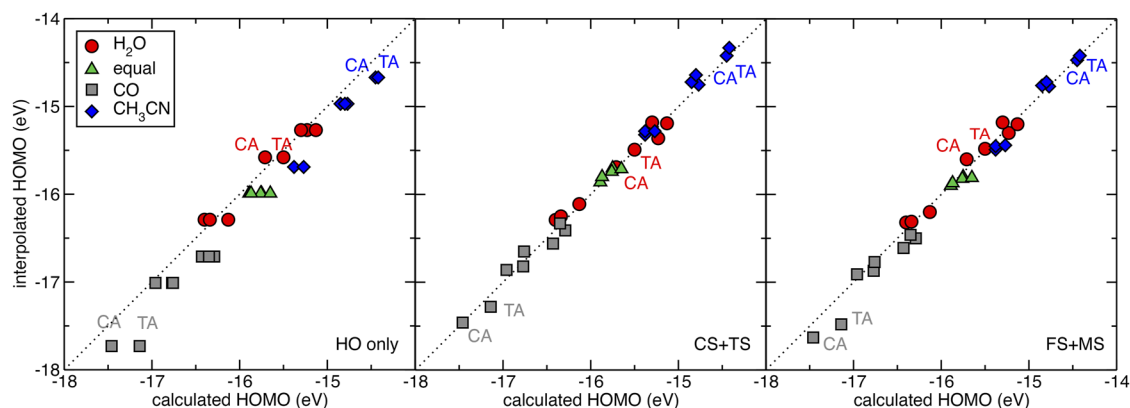


FIG. 8. Calculated vs linearly interpolated HOMO level (eV) for singlet Fe(II) complexes with at least one each of three ligands: CH₃CN, H₂O, and CO. From left to right: interpolation from homoleptic complexes (HO only), interpolation using homoleptic complexes and CS and TS complex energies derived 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.

$$E(TA M(L_1)_4(L_2)_1(L_3)_1) = \frac{1}{3} \left(E(HOM(L_1)_6) + E(MSM(L_1)_3(L_2)_3) + E(MSM(L_1)_3(L_3)_3) \right), \quad (9)$$

$$E(FAM(L_1)_3(L_2)_2(L_3)_1) = \frac{1}{3} \left(\frac{1}{2} E(HOM(L_1)_6) + \frac{1}{2} E(HOM(L_2)_6) + E(FSM(L_1)_3(L_2)_3) + E(FSM(L_1)_3(L_3)_3) \right), \quad (10)$$

$$E(MACM(L_1)_3(L_2)_2(L_3)_1) = \frac{1}{3} \left(\frac{1}{2} E(HOM(L_1)_6) + \frac{1}{2} E(HOM(L_2)_6) + E(FSM(L_1)_3(L_2)_3) + E(MSM(L_1)_3(L_3)_3) \right), \quad (11)$$

$$E(MATM(L_1)_3(L_2)_2(L_3)_1) = \frac{1}{3} \left(\frac{1}{2} E(HOM(L_1)_6) + \frac{1}{2} E(HOM(L_2)_6) + E(MSM(L_1)_3(L_2)_3) + E(MSM(L_1)_3(L_3)_3) \right), \quad (12)$$

$$E(DCSM(L_1)_2(L_2)_2(L_3)_2) = \frac{1}{3} \left(\frac{1}{3} E(HOM(L_1)_6) + \frac{1}{3} E(HOM(L_2)_6) + \frac{1}{3} E(HOM(L_3)_6) + \frac{2}{3} E(FSM(L_1)_3(L_2)_3) + \frac{2}{3} E(FSM(L_1)_3(L_3)_3) + \frac{2}{3} E(FSM(L_2)_3(L_3)_3) \right), \quad (13)$$

$$E(DTSM(L_1)_2(L_2)_2(L_3)_2) = \frac{1}{3} \left(\frac{1}{3} E(HOM(L_1)_6) + \frac{1}{3} E(HOM(L_2)_6) + \frac{1}{3} E(HOM(L_3)_6) + \frac{2}{3} E(MSM(L_1)_3(L_2)_3) + \frac{2}{3} E(MSM(L_1)_3(L_3)_3) + \frac{2}{3} E(MSM(L_2)_3(L_3)_3) \right), \quad (14)$$

$$E(EAM(L_1)_2(L_2)_2(L_3)_2) = \frac{1}{3} \left(\frac{1}{3} E(HOM(L_1)_6) + \frac{1}{3} E(HOM(L_2)_6) + \frac{1}{3} E(HOM(L_3)_6) + \frac{2}{3} E(FSM(L_1)_3(L_2)_3) + \frac{2}{3} E(MSM(L_1)_3(L_3)_3) + \frac{2}{3} E(MSM(L_2)_3(L_3)_3) \right), \quad (15)$$

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 (supplementary material, Text S3). 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 (Figs. 5 and 6 and supplementary material, Figs. S3 and S4 and Tables S14–S16). 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 (Figs. 5 and 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 (Figs. 5 and 6). Overall, errors for the interpolation scheme using FS/MS ligands are sufficiently low to warrant its use in chemical space exploration (supplementary material, Table S16). 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.

C. 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 homoleptic, followed closely by binary complexes (407) and ternary complexes (129), with only six having more ligand types (Fig. 4 and supplementary material, Table S4). 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 (supplementary material, Tables S4–S5). 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 (Fig. 9 and supplementary material, Fig. S5 and Table S17). The majority of spin-splitting energies are in the range of –30 to 0 kcal/mol, whereas a minority of complexes are low-spin (Fig. 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 near-degenerate spin states (Fig. 9). While the preference for high spin could be attributed to the strong sensitivity of ΔE_{H-L} to the choice of the functional,^{85–94} 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 ΔE_{H-L} values

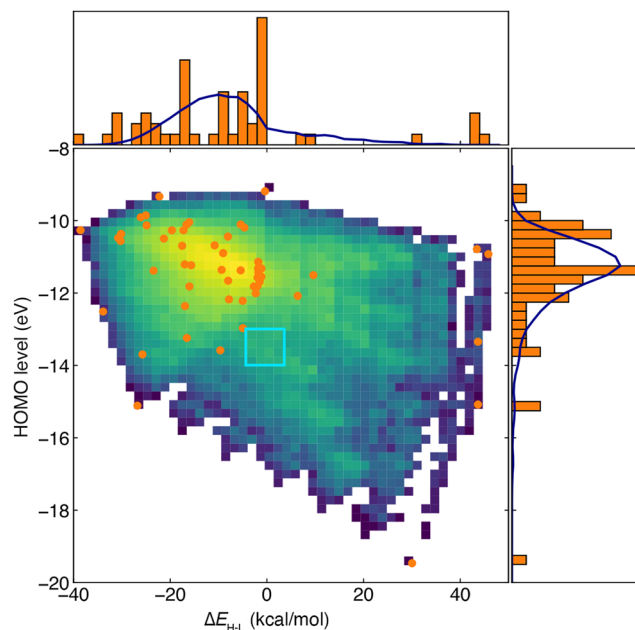


FIG. 9. The ΔE_{H-L} (in kcal/mol) vs singlet HOMO level (in eV) for 56 homoleptic complexes (orange circles) and HO-only interpolation of all possible binary and ternary complexes colored by frequency from purple (low) to yellow (high). 1D histograms of each property are shown at top and right with 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 ΔE_{H-L} and –14.0 to –13.0 eV for the HOMO is annotated as a light blue square.

that homoleptic complexes rarely sample (supplementary material, Fig. S6).

We next interpolated estimates for the 816 200 binary and ternary complexes from the 56 homoleptic complexes using the HO-only averaging scheme (Fig. 9). By definition, this interpolation scheme provides rough estimates of which ligand combinations will enrich ΔE_{H-L} /HOMO energetic pairings not observed in the homoleptic set (Fig. 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 I). 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 ΔE_{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 (Fig. 10 and supplementary material, Table S18 and Fig. S7). 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, supplementary material, Table S18). 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 ΔE_{H-L} properties at higher fidelity than HO-only averaging (see Table I). We carried out geometry optimizations of these 132 complexes and used their properties to evaluate the revised interpolated HOMO level and ΔE_{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 ΔE_{H-L} (supplementary material, Fig. S8 and Table S19). 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 (supplementary material, Fig. S8 and Table S19).

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 (supplementary material, Table S20). 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 ΔE_{H-L} underestimated (predicted: -7.3 kcal/mol vs actual -3.6 kcal/mol), while FS/MS Fe(II)(CH₃CN)₃(MeOH)₃ ΔE_{H-L} was overestimated (predicted: 9.0 kcal/mol vs 2.9 and -0.8 kcal/mol, respectively,

supplementary material, Table S20). For these same complexes, the homoleptic averaging predicted HOMO levels very well, within around 0.1 eV (i.e., lower error than for ΔE_{H-L} , supplementary material, Table S20).

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 (Fig. 11 and supplementary material, Table S20). Indeed, all three of these complexes have HOMO levels in the targeted zone, while the ΔE_{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), supplementary material, Table S20]. For CS/TS Fe(II)(ClPyz)₄(CO)₂, homoleptic averaging had underestimated both the ΔE_{H-L} and HOMO level (i.e., out of the zone at -15.32 eV) and could not distinguish between CS and TS isomers (supplementary material, Table S20). Although FS/MS interpolation slightly underestimated the ΔE_{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 (Fig. 10).

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 (supplementary material, Table S20). The homoleptic averaging predicts the ΔE_{H-L} value for both complexes to be -2.6 kcal/mol, while the calculated ΔE_{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 ΔE_{H-L} value (supplementary material, Table S20). Overall, FS/MS-interpolation MAEs for these binary complexes are low

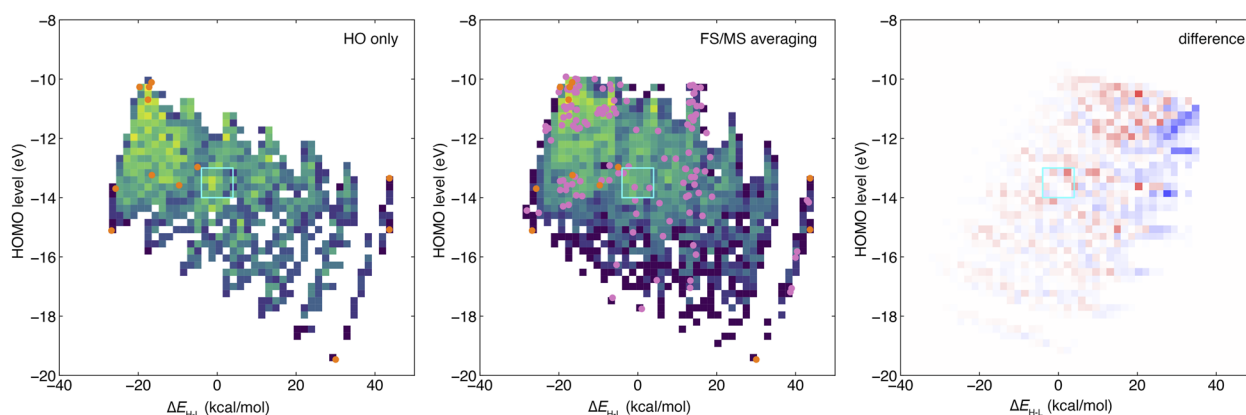


FIG. 10. Left and middle: The ΔE_{H-L} (in kcal/mol) vs singlet HOMO level (in eV) for 12 homoleptic complexes (orange circles) and HO-only (left) or FS/MS-based (middle) interpolation of all possible binary and ternary complexes colored by frequency from purple (low) to yellow (high). The FS/MS complex 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 ΔE_{H-L} and -14.0 to -13.0 eV for the HOMO is annotated as a light blue square.

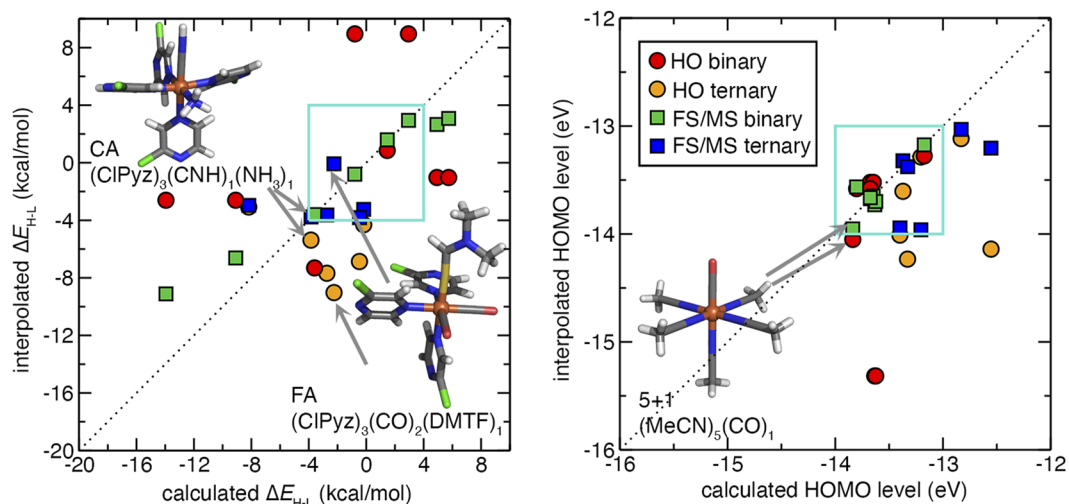


FIG. 11. Properties of eight binary (red or green) and six ternary (orange or blue) complexes in the validation set for HO-interpolation (circles) and FS/MS-augmented interpolation (squares) for ΔE_{H-L} (in kcal/mol, left) and the HOMO level (in eV, right). The targeted zone for each quantity is shown 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.

(2.5 kcal/mol for ΔE_{H-L} and 0.1 eV for HOMO levels) and less than half that observed from HO-only averaging (supplementary material, Table S20).

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 (Fig. 11 and supplementary material, Table S21). 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 (supplementary material, Table S21). Over this set, explicit DFT calculations show that five of the complexes have ΔE_{H-L} values in the targeted zone and four have HOMO levels in the targeted zone (supplementary material, Table S21). These include FA Fe(II)(CIPyz)₃(CO)₂(DMTF), which was predicted to be strongly HS by homoleptic averaging ($\Delta E_{H-L} = -9.0$ kcal/mol) but was much closer to the FS/MS-interpolated value (calculated $\Delta E_{H-L} = -2.2$ kcal/mol vs FS/MS $\Delta E_{H-L} = -0.1$ kcal/mol, supplementary material, Table S21). For the worst-performing example, FA Fe(II)(DMTF)₃(CO)₂(H₂O), FS/MS interpolation overestimates ΔE_{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, supplementary material, Table S21). 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 (~ 2 – 3 kcal/mol errors) to exceptional in the case of CA Fe(II)(CIPyz)₄(CNH)(NH₃) where errors on ΔE_{H-L} are below 0.1 kcal/mol and the HOMO level are around 0.05 eV (supplementary material, Table S21). Overall errors are low from FS/MS interpolation at around 2.5 kcal/mol for ΔE_{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.

IV. 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 ΔE_{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 ΔE_{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 ΔE_{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 ΔE_{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.

SUPPLEMENTARY MATERIAL

See the [supplementary material](#) for the details of mononuclear octahedral complex curation from the CSD; details of the complex symmetry class assignment; statistics of CSD complexes by a number of unique ligands; statistics of CSD complexes by symmetry classes; density of ligands in computation-ready CSD complexes by symmetry classes; density of ligands in Fe(II) CSD complexes by symmetry classes; number of unique monodentate ligands in Fe(II) complexes; sampled vs theoretical number of complexes with monodentate ligands; homoleptic complex properties, ΔE_{H-L} and the HOMO level; interpolated vs calculated ΔE_{H-L} for transition metal complexes with pairs of $\text{NH}_3/\text{H}_2\text{O}/\text{CO}$; interpolated and actual ΔE_{H-L} for pairs of $\text{H}_2\text{O}/\text{CO}$ and NH_3 or CH_3CN ; interpolated vs calculated HOMO for transition metal complexes with pairs of $\text{NH}_3/\text{H}_2\text{O}/\text{CO}$; MAEs for ΔE_{H-L} binary complex estimates from interpolation schemes; interpolated and actual HOMO for pairs of $\text{H}_2\text{O}/\text{CO}$ and NH_3 or CH_3CN ; MAEs for HOMO binary complex estimates from interpolation schemes; interpolated vs calculated ΔE_{H-L} for transition metal complexes with three ligands: $\text{NH}_3/\text{H}_2\text{O}/\text{CO}$; interpolated and actual ΔE_{H-L} for ternary $\text{H}_2\text{O}/\text{CO}/\text{CH}_3\text{CN}$ complexes; interpolated and actual

ΔE_{H-L} for ternary $\text{H}_2\text{O}/\text{CO}/\text{NH}_3$ complexes; interpolated vs calculated HOMO for transition metal complexes with three ligands: $\text{NH}_3/\text{H}_2\text{O}/\text{CO}$; interpolated and actual HOMO for ternary $\text{H}_2\text{O}/\text{CO}/\text{CH}_3\text{CN}$ complexes; interpolated and actual HOMO for ternary $\text{H}_2\text{O}/\text{CO}/\text{NH}_3$ complexes; CS/TS-derived interpolation expressions for ternary complexes; MAEs of HOMO and ΔE_{H-L} for ternary complexes; properties of 56 homoleptic Fe(II) complexes; scatter of HOMO and ΔE_{H-L} for 56 homoleptic Fe(II) complexes; 12 most common ligands in the target zone for Fe(II) complexes; homoleptic-only interpolated target zone for Fe(II) complexes; parity plot for predicted vs calculated 132 FS/MS Fe(II) complexes; MAEs for FS and MS complexes from HO-only averaging; binary complexes in the validation set for interpolation; and ternary complexes in the validation set for interpolation.

ACKNOWLEDGMENTS

This work was primarily supported by the Office of Naval Research under Grant No. N00014-20-1-2150. A.N., N.A., and D.B.K.C. were partially supported by the National Science Foundation Graduate Research Fellowships under Grants No. 1122374 (to A.N.) and Grant No. 1745302 (to N.A. and D.B.K.C.). C.D. was partially supported by a seed fellowship from the Molecular Sciences Software Institute under NSF Grant No. OAC-1547580. The authors thank Adam H. Steeves for a critical reading of the manuscript.

The authors declare no competing financial interest.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

N.A., S.G., and M.G.T. contributed equally to this work.

Naveen Arunachalam: Data curation (equal); Formal analysis (equal); Writing – original draft (equal); Writing – review & editing (equal). **Stefan Gugler:** Conceptualization (supporting); Formal analysis (equal); Methodology (equal). **Michael G. Taylor:** Data curation (equal); Formal analysis (equal); Writing – original draft (equal); Writing – review & editing (equal). **Chenru Duan:** Data curation (equal); Formal analysis (equal); Methodology (equal); Validation (equal); Writing – review & editing (equal). **Aditya Nandy:** Supervision (equal); Writing – review & editing (equal). **Jon Paul Janet:** Data curation (supporting); Formal analysis (supporting); Methodology (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Ralf Meyer:** Data curation (supporting); Formal analysis (supporting); Methodology (supporting); Writing – review & editing (equal). **Jonas Oldenstaedt:** Formal analysis (supporting); Methodology (supporting); Writing – review & editing (equal). **Daniel B. K. Chu:** Formal analysis (supporting); Methodology (supporting); Writing – review & editing (supporting). **Heather J. Kulik:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its [supplementary material](#).

REFERENCES

- ¹Y. N. Shu and B. G. Levine, *J. Chem. Phys.* **142**(10), 104104 (2015).
- ²R. 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).
- ³I. Y. Kanal, S. G. Owens, J. S. Bechtel, and G. R. Hutchison, *J. Phys. Chem. Lett.* **4**(10), 1613–1623 (2013).
- ⁴K. D. Vogiatzis, M. V. Polynski, J. K. Kirkland, J. Townsend, A. Hashemi, C. Liu, and E. A. Pidko, *Chem. Rev.* **119**, 2453–2523 (2018).
- ⁵M. Foscato and V. R. Jensen, *ACS Catal.* **10**(3), 2354–2377 (2020).
- ⁶S. Curtarolo, G. L. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, *Nat. Mater.* **12**(3), 191–201 (2013).
- ⁷S. 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).
- ⁸J. K. Nørskov and T. Bligaard, *Angew. Chem., Int. Ed.* **52**(3), 776–777 (2013).
- ⁹J. P. Janet, C. Duan, A. Nandy, F. Liu, and H. J. Kulik, *Acc. Chem. Res.* **54**(3), 532–545 (2021).
- ¹⁰M. Ceriotti, C. Clementi, and O. A. von Lilienfeld, *Chem. Rev.* **121**(16), 9719–9721 (2021).
- ¹¹J. 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).
- ¹²M. Ceriotti, *J. Chem. Phys.* **150**(15), 150901 (2019).
- ¹³A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf, and M. Ceriotti, *ACS Cent. Sci.* **5**, 57–64 (2018).
- ¹⁴K. Jorner, A. Tomberg, C. Bauer, C. Sköld, and P.-O. Norrby, *Nat. Rev. Chem.* **5**(4), 240–255 (2021).
- ¹⁵M. F. Kasim and S. M. Vinko, *Phys. Rev. Lett.* **127**(12), 126403 (2021).
- ¹⁶J. Westermayr, M. Gastegger, K. T. Schütt, and R. J. Maurer, *J. Chem. Phys.* **154**(23), 230903 (2021).
- ¹⁷J. P. Coe, *J. Chem. Theory Comput.* **14**(11), 5739–5749 (2018).
- ¹⁸A. 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).
- ¹⁹T. Dimitrov, C. Kreisbeck, J. S. Becker, A. Aspuru-Guzik, and S. K. Saikin, *ACS Appl. Mater. Interfaces* **11**(28), 24825–24836 (2019).
- ²⁰A. Glielmo, B. E. Husic, A. Rodriguez, C. Clementi, F. Noé, and A. Laio, *Chem. Rev.* **121**(16), 9722–9758 (2021).
- ²¹L. 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).
- ²²M. 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).
- ²³T. 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).
- ²⁴A. M. Virshup, J. Contreras-García, P. Wipf, W. Yang, and D. N. Beratan, *J. Am. Chem. Soc.* **135**(19), 7296–7303 (2013).
- ²⁵L. Ruddigkeit, R. van Deursen, L. C. Blum, and J.-L. Reymond, *J. Chem. Inf. Model.* **52**(11), 2864–2875 (2012).
- ²⁶R. W. Sterner and J. J. Elser, *Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere* (Princeton University Press, 2002).
- ²⁷H. J. M. Bowen, *Environmental Chemistry of the Elements* (Academic Press, 1979).
- ²⁸T. Fink, H. Bruggesser, and J.-L. Reymond, *Angew. Chem., Int. Ed.* **44**(10), 1504–1508 (2005).
- ²⁹M. 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).
- ³⁰P. G. Polishchuk, T. I. Madzhidov, and A. Varnek, *J. Comput.-Aided Mol. Des.* **27**(8), 675–679 (2013).
- ³¹R. S. Bohacek, C. McMartin, and W. C. Guida, *Med. Res. Rev.* **16**(1), 3–50 (1996).
- ³²N. Fey, S. E. Harris, J. N. Harvey, and A. G. Orpen, *J. Chem. Inf. Model.* **46**(2), 912–929 (2006).
- ³³N. 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).
- ³⁴D. J. Durand and N. Fey, *Chem. Rev.* **119**(11), 6561–6594 (2019).
- ³⁵M. Foscato, G. Occhipinti, V. Venkatraman, B. K. Alsberg, and V. R. Jensen, *J. Chem. Inf. Model.* **54**(3), 767–780 (2014).
- ³⁶M. Foscato, V. Venkatraman, G. Occhipinti, B. K. Alsberg, and V. R. Jensen, *J. Chem. Inf. Model.* **54**(7), 1919–1931 (2014).
- ³⁷M. Foscato, R. J. Deeth, and V. R. Jensen, *J. Chem. Inf. Model.* **55**(6), 1282–1290 (2015).
- ³⁸M. Foscato, B. J. Houghton, G. Occhipinti, R. J. Deeth, and V. R. Jensen, *J. Chem. Inf. Model.* **55**(9), 1844–1856 (2015).
- ³⁹D. Balcells and B. B. Skjelstad, *J. Chem. Inf. Model.* **60**(12), 6135–6146 (2020).
- ⁴⁰J. S. Smith, O. Isayev, and A. E. Roitberg, *Sci. Data* **4**, 170193 (2017).
- ⁴¹R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. Von Lilienfeld, *Sci. Data* **1**, 140022 (2014).
- ⁴²C. A. Gaggioli, S. J. Stoneburner, C. J. Cramer, and L. Gagliardi, *ACS Catal.* **9**(9), 8481–8502 (2019).
- ⁴³W. Jiang, N. J. DeYonker, and A. K. Wilson, *J. Chem. Theory Comput.* **8**(2), 460–468 (2012).
- ⁴⁴F. Liu, C. Duan, and H. J. Kulik, *J. Phys. Chem. Lett.* **11**, 8067–8076 (2020).
- ⁴⁵C. Duan, D. B. K. Chu, A. Nandy, and H. J. Kulik, *Chem. Sci.* **13**, 4962–4971 (2022).
- ⁴⁶M. Feldt, Q. M. Phung, K. Pierloot, R. A. Mata, and J. N. Harvey, *J. Chem. Theory Comput.* **15**(2), 922–937 (2019).
- ⁴⁷Q. M. Phung, M. Feldt, J. N. Harvey, and K. Pierloot, *J. Chem. Theory Comput.* **14**(5), 2446–2455 (2018).
- ⁴⁸S. Gugler, J. P. Janet, and H. J. Kulik, *Mol. Syst. Des. Eng.* **5**, 139–152 (2020).
- ⁴⁹A. 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).
- ⁵⁰P. F. Bernath and S. McLeod, *J. Mol. Spectrosc.* **207**(2), 287 (2001).
- ⁵¹J. P. Janet, C. Duan, T. Yang, A. Nandy, and H. J. Kulik, *Chem. Sci.* **10**, 7913–7922 (2019).
- ⁵²C. 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).
- ⁵³J. Glerup, O. Moensted, and C. E. Schaeffer, *Inorg. Chem.* **15**(6), 1399–1407 (1976).
- ⁵⁴R. J. Deeth, D. L. Foulis, and B. J. Williams-Hubbard, *Dalton Trans.* (20), 3949–3955 (2003).
- ⁵⁵Y. Cytter, A. Nandy, A. Bajaj, and H. J. Kulik, *J. Phys. Chem. Lett.* **13**, 4549–4555 (2022).
- ⁵⁶R. A. Friesner, E. H. Knoll, and Y. Cao, *J. Chem. Phys.* **125**(12), 124107 (2006).
- ⁵⁷D. Rinaldo, L. Tian, J. N. Harvey, and R. A. Friesner, *J. Chem. Phys.* **129**(16), 164108 (2008).
- ⁵⁸T. F. Hughes and R. A. Friesner, *J. Chem. Theory Comput.* **7**(1), 19–32 (2011).
- ⁵⁹R. A. Friesner and S. V. Jerome, *Coord. Chem. Rev.* **344**, 205–213 (2017).
- ⁶⁰C. 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).
- ⁶¹K. U. Lao and J. M. Herbert, *J. Phys. Chem. A* **119**(2), 235–252 (2015).
- ⁶²M. S. Gordon, L. Slipchenko, H. Li, and J. H. Jensen, *Annu. Rep. Comput. Chem.* **3**, 177–193 (2007).
- ⁶³F. Neese, A. Hansen, and D. G. Liakos, *J. Chem. Phys.* **131**(6), 064103 (2009).

- ⁶⁴Q. Ma and H. J. Werner, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **8**(6), e1371 (2018).
- ⁶⁵M. 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).
- ⁶⁶S. Vela, R. Laplaza, Y. Cho, and C. Corminboeuf, *npj Comput. Mater.* **8**(1), 188 (2022).
- ⁶⁷S. 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).
- ⁶⁸S. 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).
- ⁶⁹A. D. Becke, *J. Chem. Phys.* **98**(7), 5648–5652 (1993).
- ⁷⁰C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785–789 (1988).
- ⁷¹P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, *J. Phys. Chem.* **98**(45), 11623–11627 (1994).
- ⁷²P. J. Hay and W. R. Wadt, *J. Chem. Phys.* **82**(1), 270–283 (1985).
- ⁷³R. Ditchfield, W. J. Hehre, and J. A. Pople, *J. Chem. Phys.* **54**(2), 724 (1971).
- ⁷⁴V. R. Saunders and I. H. Hillier, *Int. J. Quantum Chem.* **7**(4), 699–705 (1973).
- ⁷⁵L.-P. Wang and C. Song, *J. Chem. Phys.* **144**(21), 214108 (2016).
- ⁷⁶E. I. Ioannidis, T. Z. H. Gani, and H. J. Kulik, *J. Comput. Chem.* **37**, 2106–2117 (2016).
- ⁷⁷See <https://github.com/hjkgrp/molsimplify> for the KulikGroup.
- ⁷⁸N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, *J. Cheminf.* **3**, 33 (2011).
- ⁷⁹N. M. O'Boyle, C. Morley, and G. R. Hutchison, *Chem. Cent. J.* **2**, 5 (2008).
- ⁸⁰N. 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>.
- ⁸¹G. Pólya and R. C. Read, *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds* (Springer, New York, NY, 1987).
- ⁸²I. Baraldi, C. Fiori and D. Vanossi, *J. Math. Chem.* **25**, 23–30 (1999).
- ⁸³B. A. Kennedy, D. A. McQuarrie, and C. H. Brubaker, *Inorg. Chem.* **3**(2), 265–268 (1964).
- ⁸⁴A. 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).
- ⁸⁵M. Reiher, *Inorg. Chem.* **41**(25), 6928–6935 (2002).
- ⁸⁶G. Ganzenmüller, N. Berkaine, A. Fouqueau, M. E. Casida, and M. Reiher, *J. Chem. Phys.* **122**(23), 234321 (2005).
- ⁸⁷M. Reiher, O. Salomon, and B. A. Hess, *Theor. Chem. Acc.* **107**(1), 48–55 (2001).
- ⁸⁸O. Salomon, M. Reiher, and B. A. Hess, *J. Chem. Phys.* **117**(10), 4729–4737 (2002).
- ⁸⁹E. I. Ioannidis and H. J. Kulik, *J. Chem. Phys.* **143**(3), 034104 (2015).
- ⁹⁰D. N. Bowman and E. Jakubikova, *Inorg. Chem.* **51**(11), 6011–6019 (2012).
- ⁹¹A. Droghetti, D. Alfè, and S. Sanvito, *J. Chem. Phys.* **137**(12), 124303 (2012).
- ⁹²D. M. Smith, M. Dupuis, and T. Straatsma, *Mol. Phys.* **103**(2–3), 273–278 (2005).
- ⁹³P. Verma, Z. Varga, J. E. Klein, C. J. Cramer, L. Que, and D. G. Truhlar, *Phys. Chem. Chem. Phys.* **19**(20), 13049–13069 (2017).
- ⁹⁴K. P. Kepp, *Inorg. Chem.* **55**(6), 2717–2727 (2016).