Identifying Underexplored and Untapped Regions in the Chemical Space of Transition Metal Complexes

Aditya Nandy[1,2], Michael G. Taylor[1], and Heather J. Kulik[1,2,*]

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

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

*Corresponding author email: hjkulik@mit.edu, phone: 617-253-4584

ABSTRACT: We survey over 230,000 crystallized mononuclear transition metal complexes (TMCs) to identify trends in preferred geometric structure and metal coordination. While we observe d-filling to influence coordination preference, with late TMCs preferring lower coordination number, we also note exceptions. We also observe that $4d$ and $5d$ transition metals and $3p$-coordinating ligands are systematically undersampled. For the roughly one third of the set of mononuclear TMCs that are octahedral, analysis of the 67 symmetry classes of their ligand environments reveals that complexes most commonly contain monodentate ligands that may likely be removable to leave an open site amenable to catalysis. Due to their frequent use in transition metal catalysts, we analyze trends in coordination by tetradentate ligands in terms of the capacity to support multiple metals and the variability of coordination geometry. We identify promising tetradentate ligands that co-occur in crystallized complexes with labile monodentate ligands, indicating their ability to generate reactive sites. Literature mining suggests that many of these tetradentate ligands are untapped as ligands in catalytic complexes, motivating proposal of a promising octa-functionalized porphyrin in this set as a candidate ligand for catalysis.

TOC Graphic:

2
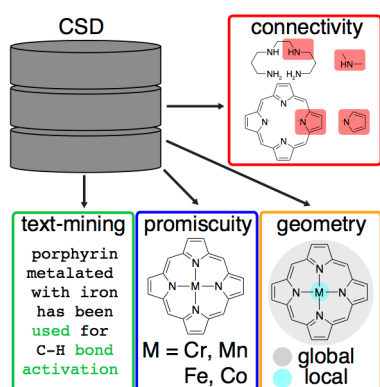
Transition metal chemical space remains challenging to traverse due to oxidation and spin state variability, a variety of metal-coordinating environments, and limited experimental data relative to organic chemistry.[1-7] This chemical space holds the potential to discover new catalysts for known reactions[8,9], catalysts with new mechanisms[10,11], or functional materials with promising optical[12], sensing[13], or magnetic[14-17] properties. Over the past two decades, this opportunity has motivated efforts to improve transition metal complex (TMC) design by mapping the space of known TMCs.[4,18] Many of these efforts have been localized to specific metals[2,19,20] or ligand classes to develop ligand knowledge bases (e.g. on phosphine[21-25] or carbene[26] ligands) to understand how to design new TMCs with improved properties.

The Cambridge Structural Database[27] (CSD) is a centralized repository of structural data that contains TMCs, enabling quantitative analysis of crystal structures over a broad chemical space. Recent studies have leveraged experimental TMC structures in the CSD to understand spin crossover[28], redox behavior in bimetallic complexes[29], oxidation state[30] and charge[2] assignment, and electronic properties.[3] Previously curated subsets of TMCs from the CSD analyzed either a subset of metals (e.g. cell2mol[2]), complexes that are closed-shell in nature (e.g. tmQM[3]), or specific TMCs for spin-crossover applications.[28] However, understanding the diversity of TMCs in terms of ligand characteristics (i.e. metal-coordinating atoms and ring substructures) and metal characteristics (i.e. metal distributions and coordination geometries) without restrictions on elemental identity would enable identification of areas for future *in silico* design[5,31-33] while maintaining synthesizability.

Here, we examine the metal coordination geometries, overall connectivities, and chemical composition for mononuclear transition metal complexes in the CSD, considering all 3*d*, 4*d*, or 5*d* transition metals and coordinating atoms. We curated a data set which we refer to as the

mononuclear CSD data set, that comprises 242,829 mononuclear transition metal complexes (TMCs) from the Cambridge Structural Database[27] (CSD) version 5.41 (November 2019) combined with the March and May 2020 updates (Supporting Information Figure S1). To assign the coordination geometry, we categorized the shape of the primary coordination sphere using the coordination number and all possible $L_1–M–L_2$ angles. We reported a subset of this data with octahedral geometries (n = 85,575) in prior work[34], which we call the *octahedral mononuclear CSD* data set. We also identified a set of unique TMCs by computing their Weisfeiler–Lehman Z-atom weighted connectivity graph hashes.[35] The CSD primarily contains $3d$ TMCs with $2p$ coordinating atoms (Figure 1). Overall, we find that Fe metal centers with N/C coordinating atoms are most heavily sampled across the full *mononuclear CSD* data set (Figure 1). These observations shift only slightly when analyzing unique molecules, with Cu appearing in a high number of distinct complexes (Figure 1). Later-group metals more frequently feature $3p$ metal-coordinating atoms, recapitulating expectations from hard-soft acid base theory (Supporting Information Figure S2). There are numerous exceptions, however, that motivate further analysis across the full data set. Our data set is larger than both the tmQM[3] and cell2mol[2] data sets because we do not place constraints on the net charge or elemental identity of transition metal atoms, thus spanning the full set of mononuclear compounds in the CSD.
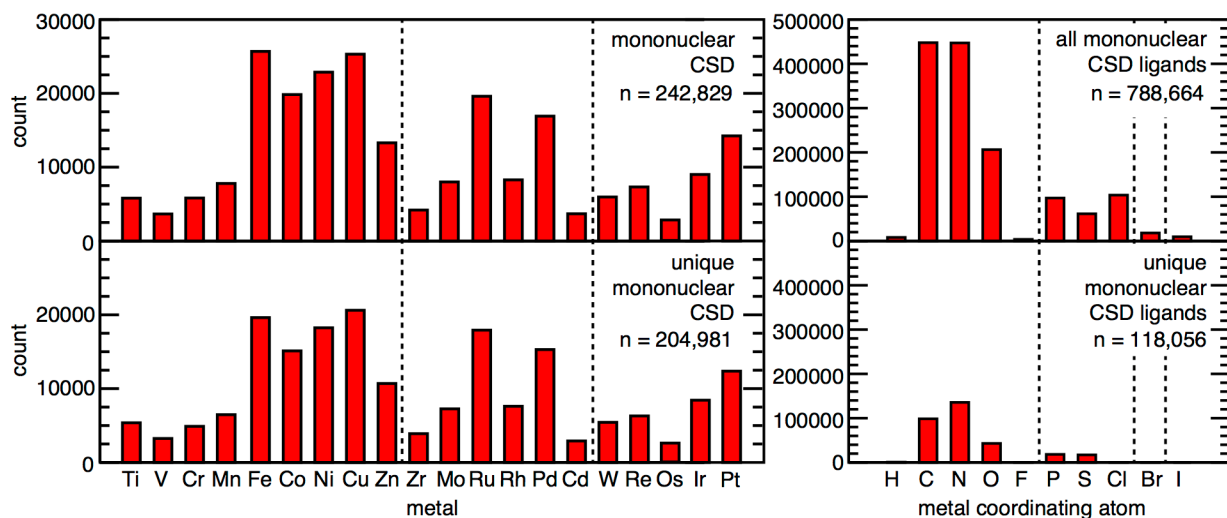
Figure 1. The top 20 metal (top left) counts in complexes and top 10 coordinating atom (top right) counts in ligands over the *mononuclear CSD* data set. Counts over the full *mononuclear CSD* data set are shown at top, and counts over unique transition metal complexes or ligands are shown at bottom, as determined by the Weisfeiler–Lehman graph hash. Dashed lines indicate different periods of the periodic table, and elements are ordered by their atomic number.

We observe 10 distinct coordination geometries with varying coordination numbers (CN) across all mononuclear compounds: CN 4 (i.e., tetrahedral, square planar, seesaw), CN 5 (i.e., trigonal bipyramidal and square pyramidal), CN 6 (i.e., octahedral and trigonal prismatic), CN 7 (i.e., pentagonal bipyramidal), and ambiguous coordination number sandwich (hapticity, η, > 2) or edge ($\eta = 2$) compounds. Of these, CN 6 octahedral and CN 4 seesaw are the most and least common geometries, respectively (Supporting Information Figure S3 and Table S1). In accordance with molecular orbital theory, we find that metal identity influences the preferred coordination geometries. We find that 50% (29,540 out of 59,037) of mid-row $3d$ (i.e., Cr, Mn, Fe, and Co) TMCs are in octahedral geometries, whereas only 2% (1,066) are in square planar geometries and < 0.5% (244) are in seesaw geometries (Supporting Information Table S2). This is in contrast to later $3d$ (i.e., Ni, Cu, Zn) TMCs that are comparably represented in our data set (i.e., mid-row: 59,037, late: 61,432) but sample geometries with lower coordination numbers. Only 28% (17,071

5

out of 61,432) of late TMCs are in octahedral geometries, with a comparable 28% (17,427 out of 61,432) in square planar geometries (Supporting Information Table S2).

Because octahedral geometries comprise a significant fraction of TMCs in the *mononuclear CSD* data set (35%, 85,575 out of 242,829), we carried out an analysis of the symmetry classes[34] present in the *octahedral mononuclear CSD* data set (Supporting Information Text S1 and Figure S4). We find that TMCs with two identical bidentate ligands in the equatorial plane and two identical axial monodentate ligands (i.e., |22||11|t where "||" groups identical ligands and t indicates ligands *trans* to each other, as detailed further in Supporting Information Text S1) are the most common symmetry class (Figure 2). When preserving only unique cases, the |22|2 symmetry class, which contains three bidentate ligands with two unique ligand identities, is the most common (Figure 2). Homoleptic monodentate compounds (i.e., |111111|) are the sixth-most common symmetry class, but the number of these shrinks dramatically when eliminating duplicates since the majority of such complexes are solvated metal ions (Figure 2). Similarly, the homoleptic bidentate transition metal complexes (i.e., |222|) drop from the third-most frequent in absolute counts to the ninth-most common, due to multiple repeat cases of $Fe(bpy)_3$ and $Ru(bpy)_3$ (Figure 2). Intriguingly, we find that many of the most common symmetry classes (e.g., |22||11|t, |111111|, |1111||11|t, and 4|11|t) contain monodentate ligands that may likely be removable to leave an open site in the complex amenable for catalysis.
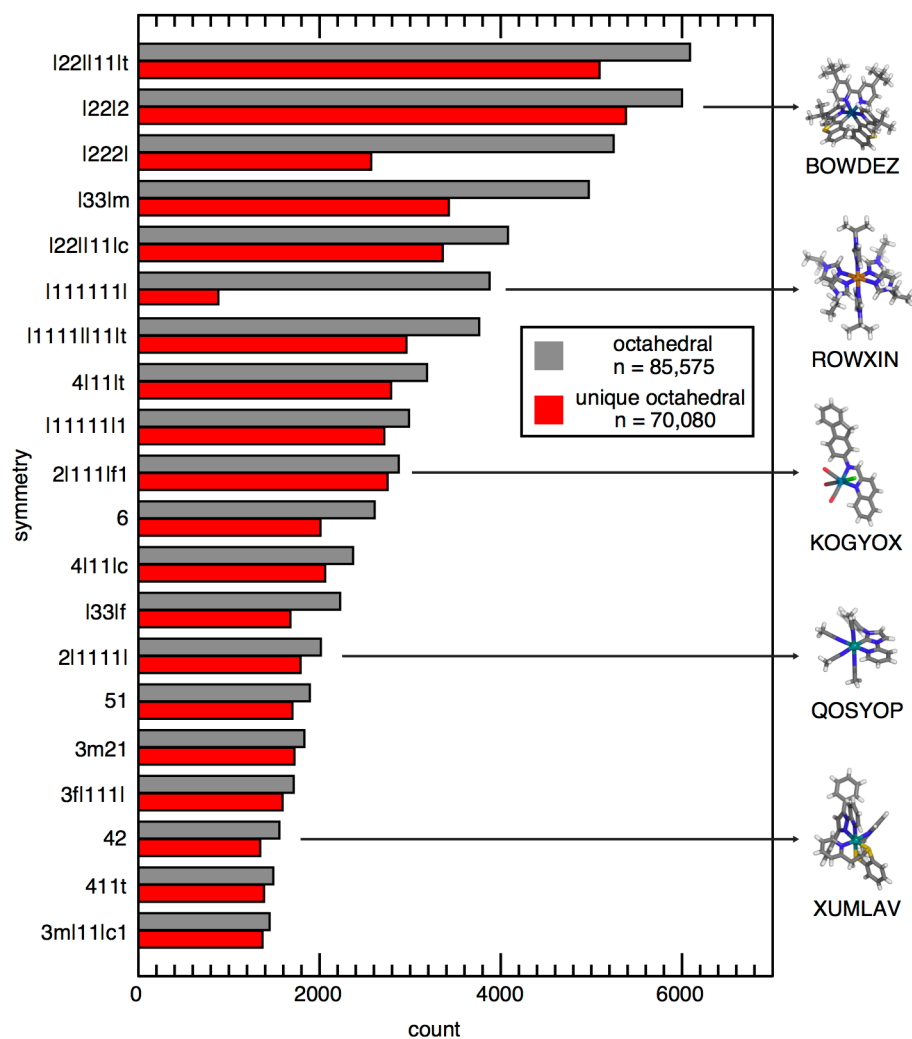
Figure 2. Bar plots of the 20 most common symmetry classes over the *octahedral mononuclear CSD* data set (gray) and accounting for repeat complexes (red). Representative structures for a subset of symmetry classes are shown with their six-digit CSD refcodes indicated. Colors are as follows: white for H, gray for C, blue for N, yellow for S, teal for Ir, brown for Cu, green for Re, and dark green for Ru.

TMCs with tetradentate ligands account for 12% (8,520 out of 70,080) of unique octahedral complexes. The 4|11|t symmetry class with a planar tetradentate ligand and identical monodentate axial ligands is the sixth-most frequent configuration when accounting for unique compounds (Figure 2). Overall, we find that tetradentate macrocycles are evenly divided between planar (49%) and seesaw (51%) configurations (Supporting Information Table S3). We find that 7% (143 out of 2,091) of the unique tetradentate macrocycles that appear in planar configurations (i.e., the 4|11|t,

7

411t, or 4planar2 symmetry classes) also appear in configurations in which the macrocycle is bent (i.e., the 4|11|c, 411c, or 42 symmetry classes, Supporting Information Table S4). Macrocycles that appear most frequently in both seesaw and planar configurations for the same metal center tend to be ligands containing saturated elements with rotatable bonds (Figure 3). For instance, we find that two complexes containing the same cyclam ligand coordinated to Mn exist in different symmetry classes (refcode SIQMUA: 411c, refcode AFEJOL: 4|11|t, Figure 3). More surprisingly, we find exceptions to this general rule, with some rigid, planar ligands (e.g. tetraphenylporphyrin) appearing in seesaw configurations with certain metals (e.g. Zr) that have a larger coordination sphere and can accommodate higher coordination numbers (Figure 3). Here, the seesaw configuration arises as a result of out-of-plane distortion for the metal–ligand bonds with the macrocycle.
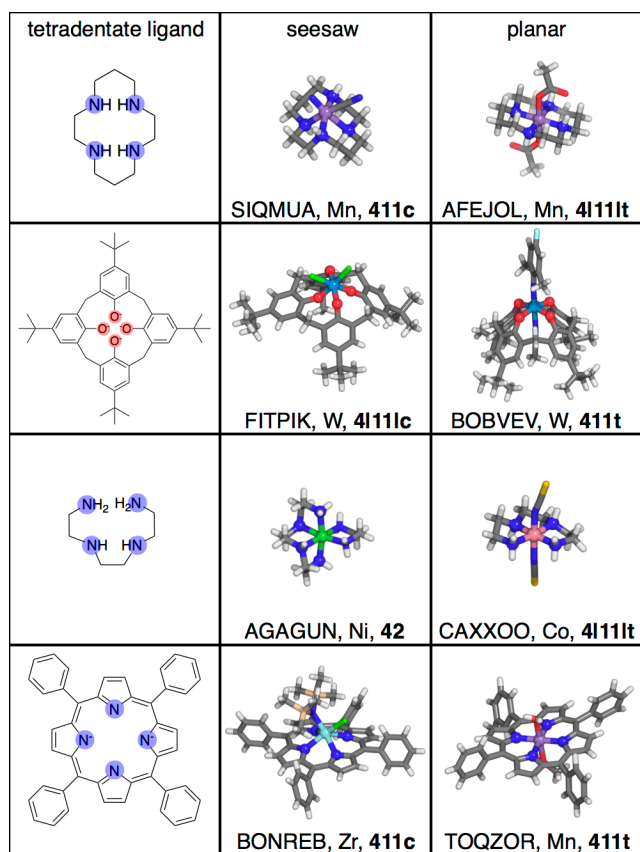
tetradentate ligand	seesaw	planar
	SIQMUA, Mn, **411c**	AFEJOL, Mn, **4|11|t**
	FITPIK, W, **4|11|c**	BOBVEV, W, **411t**
	AGAGUN, Ni, **42**	CAXXOO, Co, **4|11|t**
	BONREB, Zr, **411c**	TOQZOR, Mn, **411t**

Figure 3. Representative examples of tetradentate ligands that appear in seesaw and planar configurations. The skeleton of each macrocycle is shown, with metal-coordinating atoms highlighted with translucent circles colored by coordinating atom identity: blue for N and red for O. Representative structures are shown with their CSD refcodes, metal, and assigned symmetry class in bold. Colors in structures are as follows: white for H, gray for C, blue for N, red for O, cyan for F, yellow for S, light green for Cl, purple for Mn, pink for Co, green for Ni, light blue for Zr, and aqua for W.

We next assessed the metal-promiscuity of tetradentate ligands. Certain privileged ligands, such as tetraphenylporphyrin, have been characterized in complex with up to 15 distinct metals (Figure 4 and Supporting Information Figure S5 and Tables S5–S6). On average, we find that each unique tetradentate ligand from the set of octahedral complexes has been metalated with 1.3 metals (standard deviation: 1.0, median: 1.0), with the majority of ligands only appearing with a single metal (Figure 4 and Supporting Information Table S7). Ligands metalated with $3d$ transition metals are typically only found with other $3d$ transition metals, whereas complexes that can coordinate $5d$ metals frequently coordinate $3d$ metals as well (Figure 4 and Supporting Information Figure S5). In particular, macrocycles that have been characterized in complex with at least two different $5d$ metals have also been coordinated with $4d$ or $3d$ metals (Figure 4 and Supporting Information Figure S5).
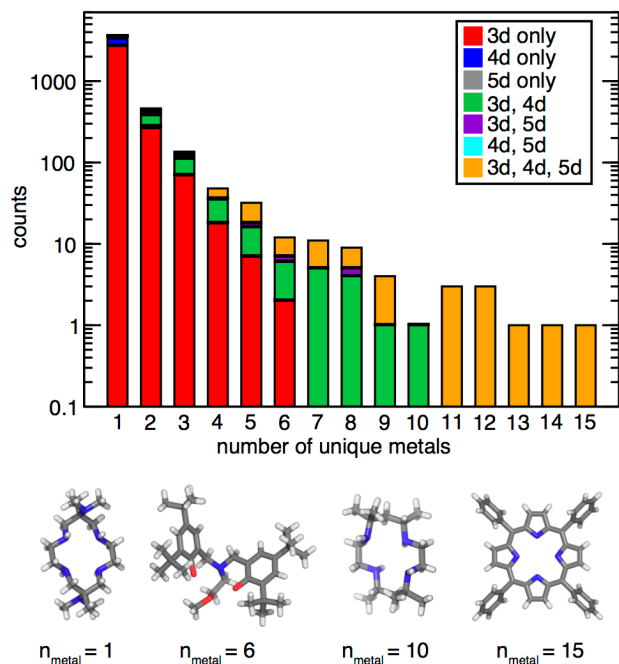
Figure 4. Bar plot of number of unique metals coordinating a tetradentate ligand. The plot is colored by the type of transition metal(s) supported. Representative examples of a very low (n_{metal} = 1) promiscuity ligand (trans-6,13-dimethyl-6,13-bis(dimethylamino)-1,4,8,11-tetra-azacyclotetradecane), low (n_{metal} = 6) promiscuity ligand (2,2'-[{[2-(methoxy)ethyl]azanediyl}bis(methylene)]bis(4,6-di-t-butylphenolato)), moderate (n_{metal} = 10) promiscuity ligand (5,5,7,12,12,14-hexamethyl-1,4,8,11-tetraazacyclotetradecane), and high (n_{metal} = 15) promiscuity ligand (tetraphenylporphyrin) are shown. Colors are as follows: white for H, gray for C, blue for N, red for O.

Because the majority of structurally characterized tetradentate ligands are only paired with a single metal, we next evaluated the likelihood of a ligand to coordinate a specific metal based on its relative binding pocket size with respect to that of the most promiscuous tetradentate ligand (i.e., tetraphenylporphyrin). This analysis could identify ligands that are compatible with other metals but for which crystallization of the resulting complexes has not been carried out. We evaluated the pocket sizes, which we define by the smallest and largest N–N distances of the metal-coordinating atoms in each instance (i.e., with each different metal) of tetraphenylporphyrin (Supporting Information Table S8). The relatively small variation in pocket sizes in tetraphenylporphyrin but high promiscuity with metal identity suggests that a large number of

existing tetradentate ligands may be able to coordinate a different metal. Indeed, we find 1,544

ligands that fall within the bond length ranges of the tetraphenylporphyrin cases with different

metals, suggesting they can coordinate other metals they have not yet been crystallized in complex

with (Supporting Information Table S8 and a list of refcodes is provided in the Zenodo repository).

For instance, we find that a Zr complex (refcode: MITZEW) with an N- and O-coordinating

tetradentate ligand has bond lengths that are likely to support an isovalent series of group 8 metals

(e.g., Fe, Ru, Os), which do not appear in the CSD in complex with that macrocycle (Figure 5 and

Supporting Information Table S8). In contrast, we find a V complex (refcode: KARRAX) with a

smaller pocket size that can likely support Fe but is too small to support Ru or Os (Figure 5 and
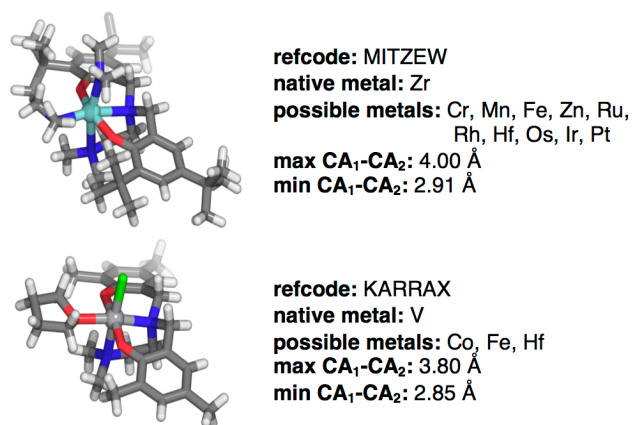
Supporting Information Table S8).

refcode: MITZEW
native metal: Zr
possible metals: Cr, Mn, Fe, Zn, Ru,
                     Rh, Hf, Os, Ir, Pt
max CA_1-CA_2: 4.00 Å
min CA_1-CA_2: 2.91 Å

refcode: KARRAX
native metal: V
possible metals: Co, Fe, Hf
max CA_1-CA_2: 3.80 Å
min CA_1-CA_2: 2.85 Å

Figure 5. Two representative examples of complexes that are likely to be remetalated based on ligand pocket size relative to tetraphenylporphyrin. The minimum CA_1–CA_2 and maximum CA_1–CA_2 distances (where CA refers to metal-coordinating atom) are indicated directly in the figure. Atom colors are as follows: white for H, gray for C, blue for N, red for O, light green for Cl, silver for V, and teal for Zr.

In an octahedral coordination environment, the tetradentate ligand must be combined with

either two monodentate ligands, or in a minority of cases a bidentate ligand, to complete metal

coordination. As described earlier, a significant number of complexes in the CSD were observed

to have tetradentate ligands in combination with two monodentate ligands. To further identify if

these complexes have a labile monodentate ligand making them amenable to catalysis, we inspected the monodentate ligands that are paired with tetradentate ligands (Supporting Information Figure S6). The most common monodentate ligands include anions (e.g., chloride and bromide) and neutral small molecules (e.g., methanol and acetonitrile) that are common solvents and thus are likely labile (Supporting Information Figure S6). We curated a list of likely labile ligands corresponding to solvents and find that 34% (1,429 out of 4,255) of unique tetradentate ligands in an octahedral geometry are paired with a labile solvent monodentate ligand (Supporting Information Table S9).

To identify tetradentate ligands that have coordination variability making them amenable to be prepared with removable axial ligands, e.g., for catalysis, we quantified the overlap between tetradentate ligands that appear in octahedral, square pyramidal, and square planar geometries. Although octahedral geometries ($n = 85{,}575$) significantly outnumber both square planar ($n = 49{,}643$) and square pyramidal ($n = 12{,}782$) geometries, they have comparable numbers of unique tetradentate ligands (Supporting Information Table S10). We find that 10% (957 out of 9,855) of these tetradentate ligands appear in two coordination geometries and 176 appear in all three (Supporting Information Figure S7). For example, a dianionic N_4 ligand that appears with Fe in an octahedral geometry with two cyanopyridine ligands (refcode: CESBEH) also appears with Fe in a square pyramidal geometry with a methanol ligand (refcode: KARFUF) and without any axial ligands, in a square planar geometry (refcode: KARFOZ), indicating that the additional ligands are removable (Supporting Information Figure S8). Nevertheless, complexes with intrinsically labile ligands may simply have not been crystallized without these ligands, suggesting the identified complex counts in multiple coordination geometries represent a lower bound on such complexes.

It is not generally known which coordination geometry a given metal and ligand combination will favor. We might expect larger metals to prefer ligands with lower connectivity if it allows for more flexibility around the metal. To identify these relationships, we defined "closed" and "open" tetradentate ligands as those with or without a closed ring containing all four metal-coordinating atoms, respectively. The relative frequency of closed and open structures is roughly comparable among octahedral, square pyramidal, and square planar coordination geometries, indicating that overall tetradentate ligand connectivity alone does not dictate coordination geometry (Figure 6 and Supporting Information Figure S9). Because many metal-coordinating atoms appear within rings (i.e., are cyclic), we investigated the frequencies of ring substructures. Octahedral TMCs less frequently contain cyclic coordinating atoms relative to square planar or square pyramidal TMCs, indicating that metal-local atom connectivity influences crystallized geometry more than the global tetradentate connectivity (Figure 6 and Supporting Information Figure S10). Pyrrole and pyridine substructures are frequently observed in tetradentate ligands, whereas the $3p$ equivalents of these substructures (i.e., phospholes or phosphorines) appear infrequently (Supporting Information Figure S10). Additionally, we find that $3p$ coordinating atoms more frequently appear in non-cyclic environments relative to $2p$ coordinating atoms ($2p$: 43% cyclic, $3p$: 4% cyclic). In line with hard-soft acid-base theory, we find that ligands with any $3p$ coordinating atoms tend to coordinate heavier (e.g., $4d$ or $5d$) metals, which are also less frequently sampled overall (Supporting Information Figure S2). Thus, re-metalating existing $3p$-coordinating tetradentate ligands with new $4d$ or $5d$ transition metals is a promising strategy for chemical space exploration.
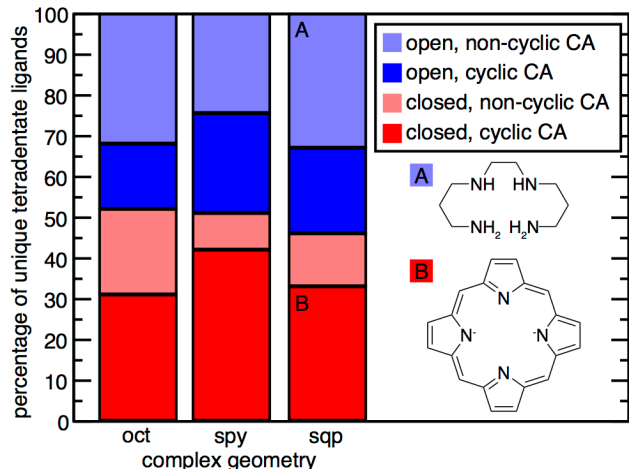
Figure 6. Percentage of unique tetradentate ligands with closed rings containing all four metal coordination atoms (CA), denoted "closed," relative to ligands that do not have a closed ring containing all four metal coordination atoms, denoted "open." CAs contained within ring substructures (e.g. pyrroles) are indicated as "cyclic," in contrast to those outside of ring substructures, indicated as "non-cyclic." Two representative chemical structures are shown inset: A) an open cyclam-like tetradentate ligand with non-cyclic CAs, and B) a porphyrin, which is a closed tetradentate ligand with cyclic CAs. Relative fractions are split by geometry: octahedral (oct, coordination number: 6), square pyramidal (spy, coordination number: 5), and square planar (sqp, coordination number: 4) are shown.

Finally, given that we have identified tetradentate ligands paired with one or more monodentate ligands as candidates for catalytic applications, we searched for TMCs that have not yet been used for catalysis but are good candidates for such purposes. Using our previous approach for text mining manuscripts[28,36,37], we curated a corpus for octahedral and square planar TMCs that contain tetradentate ligands. Using a lexicon selected by trial and error to perform keyword matching, we identified when the publications corresponding to the deposited structures did not draw conclusions on catalytic activity (Supporting Information Table S11 and a list of DOIs are provided in the Zenodo repository). We examined keywords in the main text as well as in article titles. Surprisingly, only a minority (28%) of octahedral compounds containing planar tetradentate ligands with downloadable manuscripts had been studied for catalysis, and the fraction was even lower for other coordination geometries (square planar: 22%, Supporting Information Table S12).

A large fraction of the compounds studied for catalysis contain the ligands we previously categorized as labile (243 out of 698, 35%, Supporting Information Table S9 and a list of refcodes is provided in the Zenodo repository). For the set of octahedral compounds that have not been used for catalysis, we found that a significant fraction (366 out of 1,810, 20%) also contain these labile ligands (Supporting Information Table S9). From the 191 unique tetradentate ligands that appear in octahedral complexes in combination with labile ligands, 79 (41%) also appear in square planar geometries, indicating that they have been isolated in other geometries and are correspondingly potentially suitable for catalysis. Of these, 50 have downloadable associated manuscripts, and we identified 48 that were not used for catalysis applications in their corresponding CSD depositions. A list of the DOIs and refcodes is provided in the Zenodo repository. The tetradentate porphyrinic macrocycle in a square planar Zn complex (CSD refcode: ODIPIC) also appears in an octahedral complex (refcode: ODINUM), suggesting that its axial ligands are removable, which would allow for the formation of a reactive active site. This macrocycle has been studied for its supramolecular chemistry but has not yet been studied for catalysis.[38] A search of the CSD with the name of the compound ("(5,10,15,20-tetrakis(4-hydroxyphenyl)-2,3,7,8,12,13,17,18-octaphenylporphyrinato-zinc(II)") yields no other compounds. Similarly, the use of highly substituted Zn-metalated octaethylporphyrins for their reactivity[39] implies that ligands such as that in ODIPIC may be repurposed as catalyst candidates (Supporting Information Figure S11). This octa-functionalized porphyrin ligand has not been crystallized in combination with other metals that may be more catalytically relevant than Zn depending on the reaction, but our earlier analysis suggests that the pocket size would be amenable to other metals. For instance, this pocket size is amenable to supporting Fe, a metal commonly employed for catalysis. The resulting complex may have potential for use as a C–H activation or CO_2 reduction catalyst.

In summary, we mapped the chemical space of over 230,000 crystallized TMCs and highlighted areas for future effort. Across mononuclear TMCs, we found that $3p$ coordinating ligands and $4d$ and $5d$ transition metals are systematically undersampled. We observed that tetradentate ligands are studied in various geometries and are frequently promiscuous, supporting many metals. We quantified the pocket size of the most promiscuous ligand, tetraphenylporphyrin, and identified over 1,500 ligands that are likely to be similarly promiscuous, although they have not been crystallized with many metals indicating potential for TMC synthesis opportunities. By analyzing tetradentate ligand connectivity and metal coordination geometry, we found that whether the ligand forms a closed ring through all metal-coordinating atoms does not dictate the metal coordination geometry that it is crystallized in, whereas the nature of coordinating atoms in ring substructures (e.g., N in pyrrole vs. N in an amine) holds a greater influence. After identifying ligands that are observed in the presence and absence of likely labile monodentate ligands, we identified whether some of these ligands could be useful, untapped catalysts. We identified an example of an octa-functionalized porphyrin that does not have documented use for catalysis but is a promising candidate for coordinating multiple metals. Further analysis of compounds in the CSD in additional coordination geometries is expected to present other opportunities for repurposing synthesized TMCs in future experimental and computational studies.

ASSOCIATED CONTENT
Supporting Information. The following files are available free of charge.

Workflow for CSD structure curation; cooccurrence statistics between metals and coordinating atoms; statistics of metal coordination geometries; statistics of symmetry class distributions; examples and statistics of promiscuous tetradentate ligand; character and counts for labile monodentate ligands; statistics for tetradentate ligands that appear in multiple geometries; distribution of ligand ring sizes split by metal coordination geometry; Delaunay triangulated hull volumes and buried volumes by metal coordination geometry; standard deviations of Delaunay hull volumes by atom count; comparison of hull volumes for ligands in multiple coordination

geometries; text mining keywords used for identifying catalysts; statistics for keyword matches over downloadable manuscripts; representative example of TMC to repurpose for catalysis (PDF)

Details and refcodes for all mononuclear CSD data, including graph hashes, names, chemical formulae, metals, metal-coordinating atom symbols, and deposited manuscript DOI; octahedral subset of CSD data with corresponding symmetry assignment; Delaunay triangulated volumes for octahedral (with and without axial ligands), square pyramidal, and square planar structures; ligand mol2 file. All data associated with this manuscript are also available on Zenodo at `10.5281/zenodo.7853466`. (ZIP)

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

AUTHOR INFORMATION

Notes

The authors declare no competing financial interest.

References

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

(2)   Vela, S.; Laplaza, R.; Cho, Y.; Corminboeuf, C. Cell2mol: Encoding Chemistry to Interpret Crystallographic Data. *npj Comput. Mater.* **2022,** *8*.

(3)   Balcells, D.; Skjelstad, B. B. Tmqm Dataset—Quantum Geometries and Properties of 86k Transition Metal Complexes. *J. Chem. Inf. Model.* **2020,** *60*, 6135-6146.

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

(5)   Moosavi, S. M.; Nandy, A.; Jablonka, K. M.; Ongari, D.; Janet, J. P.; Boyd, P. G.; Lee, Y.; Smit, B.; Kulik, H. J. Understanding the Diversity of the Metal-Organic Framework Ecosystem. *Nat. Commun.* **2020,** *11*.

(6)   Janet, J. P.; Duan, C.; Nandy, A.; Liu, F.; Kulik, H. J. Navigating Transition-Metal Chemical Space: Artificial Intelligence for First-Principles Design. *Acc. Chem. Res.* **2021,** *54*, 532-545.

(7)   Mroz, A. M.; Posligua, V.; Tarzia, A.; Wolpert, E. H.; Jelfs, K. E. Into the Unknown: How Computation Can Help Explore Uncharted Material Space. *J. Am. Chem. Soc.* **2022,** *144*, 18730-18743.

(8)   Goldsmith, J. I.; Hudson, W. R.; Lowry, M. S.; Anderson, T. H.; Bernhard, S. Discovery and High-Throughput Screening of Heteroleptic Iridium Complexes for Photoinduced Hydrogen Production. *J. Am. Chem. Soc.* **2005,** *127*, 7502-7510.

(9)   Vogiatzis, K. D.; Polynski, M. V.; Kirkland, J. K.; Townsend, J.; Hashemi, A.; Liu, C.; Pidko, E. A. Computational Approach to Molecular Catalysis by 3d Transition Metals: Challenges and Opportunities. *Chem. Rev.* **2018,** *119*, 2453-2523.

(10)   Cho, K.-B.; Nam, W. A Theoretical Study into a Trans-Dioxo Mn^v Porphyrin Complex That Does Not Follow the Oxygen Rebound Mechanism in C–H Bond Activation Reactions. *Chem. Commun.* **2016,** *52*, 904-907.

(11)   Jung, J.; Kim, S.; Lee, Y. M.; Nam, W.; Fukuzumi, S. Switchover of the Mechanism between Electron Transfer and Hydrogen-Atom Transfer for a Protonated Manganese(Iv)–Oxo Complex by Changing Only the Reaction Temperature. *Angew. Chem. Int. Ed.* **2016,** *55*, 7450-7454.

(12)   DiLuzio, S.; Mdluli, V.; Connell, T. U.; Lewis, J.; VanBenschoten, V.; Bernhard, S. High-Throughput Screening and Automated Data-Driven Analysis of the Triplet Photophysical Properties of Structurally Diverse, Heteroleptic Iridium(Iii) Complexes. *J. Am. Chem. Soc.* **2021,** *143*, 1179-1194.

(13)   Miller, R. G.; Brooker, S. Reversible Quantitative Guest Sensing Via Spin Crossover of an Iron(Ii) Triazole. *Chem. Sci.* **2016,** *7*, 2501-2505.

(14)   Ashley, D. C.; Jakubikova, E. Ironing out the Photochemical and Spin-Crossover Behavior of Fe(Ii) Coordination Compounds with Computational Chemistry. *Coord. Chem. Rev.* **2017,** *337*, 97-111.

(15) Shaffer, D. W.; Bhowmick, I.; Rheingold, A. L.; Tsay, C.; Livesay, B. N.; Shores, M. P.; Yang, J. Y. Spin-State Diversity in a Series of Co(Ii) Pnp Pincer Bromide Complexes. *Dalton Trans*. **2016,** *45*, 17910-17917.

(16) Bowman, D. N.; Jakubikova, E. Low-Spin Versus High-Spin Ground State in Pseudo-Octahedral Iron Complexes. *Inorg*. *Chem*. **2012,** *51*, 6011-6019.

(17) Phan, H.; Hrudka, J. J.; Igimbayeva, D.; Lawson Daku, L. M.; Shatruk, M. A Simple Approach for Predicting the Spin State of Homoleptic Fe(Ii) Tris-Diimine Complexes. *J. Am*. *Chem*. *Soc*. **2017,** *139*, 6437-6447.

(18) Tarzia, A.; Jelfs, K. E. Unlocking the Computational Design of Metal–Organic Cages. *Chem*. *Commun*. **2022,** *58*, 3717-3730.

(19) Cundari, T. R. Computational Studies of Transition Metal–Main Group Multiple Bonding. *Chem*. *Rev*. **2000,** *100*, 807-818.

(20) Sinha, V.; Laan, J. J.; Pidko, E. A. Accurate and Rapid Prediction of Pk_a of Transition Metal Complexes: Semiempirical Quantum Chemistry with a Data-Augmented Approach. *Phys*. *Chem*. *Chem*. *Phys*. **2021,** *23*, 2557-2567.

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

(22) Jover, J.; Fey, N.; Harvey, J. N.; Lloyd-Jones, G. C.; Orpen, A. G.; Owen-Smith, G. J. J.; Murray, P.; Hose, D. R. J.; Osborne, R.; Purdie, M. Expansion of the Ligand Knowledge Base for Chelating P,P-Donor Ligands (Lkb-Pp). *Organometallics* **2012,** *31*, 5302-5306.

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

(24) Fey, N.; Harvey, J. N.; Lloyd-Jones, G. C.; Murray, P.; Orpen, A. G.; Osborne, R.; Purdie, M. Computational Descriptors for Chelating P,P- and P,N-Donor Ligands[1]. *Organometallics* **2008,** *27*, 1372-1383.

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

(26) Fey, N.; Haddow, M. F.; Harvey, J. N.; McMullin, C. L.; Orpen, A. G. A Ligand Knowledge Base for Carbenes (Lkb-C): Maps of Ligand Space. *Dalton Trans*. **2009,** DOI:10.1039/b909229c 10.1039/b909229c, 8183.

(27) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta*. *Crystallogr*. *B*. *Struct*. *Sci*. *Cryst*. *Eng*. *Mater*. **2016,** *72*, 171-179.

(28) 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. *J. Phys*. *Chem*. *A* **2020,** *124*, 3286-3299.

(29) Taylor, M. G.; Nandy, A.; Lu, C. C.; Kulik, H. J. Deciphering Cryptic Behavior in Bimetallic Transition-Metal Complexes with Machine Learning. *J. Phys*. *Chem*. *Lett*. **2021,** *12*, 9812-9820.

(30) Jablonka, K. M.; Ongari, D.; Moosavi, S. M.; Smit, B. Using Collective Knowledge to Assign Oxidation States of Metal Cations in Metal–Organic Frameworks. *Nat*. *Chem*. **2021,** *13*, 771-777.

(31) Foscato, M.; Venkatraman, V.; Jensen, V. R. Denoptim: Software for Computational De Novo Design of Organic and Inorganic Molecules. *J. Chem. Inf. Model.* **2019,** *59*, 4077-4082.

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

(33) Duan, C.; Chen, S.; Taylor, M. G.; Liu, F.; Kulik, H. J. Machine Learning to Tame Divergent Density Functional Approximations: A New Path to Consensus Materials Design Principles. *Chem. Sci.* **2021,** *12*, 13021-13036.

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

(35) Shervashidze, N.; Schweitzer, P.; Leeuwen, E. J. v.; Mehlhorn, K.; Borgwardt, K. M. Weisfeiler-Lehman Graph Kernels. *J. Mach. Learn. Res.* **2011,** *12*, 2539-2561.

(36) Nandy, A.; Duan, C.; Kulik, H. J. Using Machine Learning and Data Mining to Leverage Community Knowledge for the Engineering of Stable Metal–Organic Frameworks. *J. Am. Chem. Soc.* **2021,** *143*, 17535-17547.

(37) Nandy, A.; Terrones, G.; Arunachalam, N.; Duan, C.; Kastner, D. W.; Kulik, H. J. Mofsimplify, Machine Learning Models with Extracted Stability Data of Three Thousand Metal–Organic Frameworks. *Sci. Data* **2022,** *9*.

(38) Ishizuka, T.; Sankar, M.; Kojima, T. Control of the Spatial Arrangements of Supramolecular Networks Based on Saddle-Distorted Porphyrins by Intermolecular Hydrogen Bonding. *Dalton Trans.* **2013,** *42*, 16073.

(39) Pognon, G.; Mamardashvilli, N. Z.; Weiss, J. Convenient Preparation of 5-Ethynyl-Octaethylporphyrin Free Base and Zinc Complex. *Tetrahedron Lett.* **2007,** *48*, 6174-6176.