

## Machine-learning-assisted Synthesis of Polar Racemates

Matthew L. Nisbet,<sup>†</sup> Ian M. Pendleton,<sup>‡</sup> Gene M. Nolis,<sup>§</sup> Kent J. Griffith,<sup>†</sup> Joshua Schrier,<sup>#</sup> Jordi Cabana,<sup>§</sup> Alexander J. Norquist,<sup>‡</sup> Kenneth R. Poeppelmeier<sup>\*,†</sup>

<sup>†</sup> Department of Chemistry, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, USA

<sup>‡</sup> Department of Chemistry, Haverford College, 370 Lancaster Avenue, Haverford, Pennsylvania 19041, USA

<sup>§</sup> Department of Chemistry, University of Illinois at Chicago, 845 W. Taylor Street, Chicago, Illinois 60607, USA

<sup>#</sup> Department of Chemistry, Fordham University, 441 E. Fordham Road, The Bronx, New York, New York 10458, USA

**Abstract.** Racemates have recently received attention as nonlinear optical and piezoelectric materials. Here, a machine-learning-assisted composition space approach was applied to synthesize the missing M = Ti, Zr members of the  $\Delta, \Lambda$ -[Cu(bpy)<sub>2</sub>(H<sub>2</sub>O)]<sub>2</sub>[MF<sub>6</sub>]<sub>2</sub>·3H<sub>2</sub>O (M = Ti, Zr, Hf; bpy = 2,2'-bipyridine) family (space group: *Pna*2<sub>1</sub>). In each (CuO, MO<sub>2</sub>)/bpy/HF(*aq*) (M = Ti, Zr, Hf) system, the polar noncentrosymmetric racemate (M-NCS) forms in competition with a centrosymmetric one-dimensional chain compound (M-CS) based on alternating Cu(bpy)(H<sub>2</sub>O)<sub>2</sub><sup>2+</sup> and MF<sub>6</sub><sup>2-</sup> basic building units (space groups: Ti-CS (*Pnma*), Zr-CS (*P* $\bar{1}$ ), Hf-CS (*P2*/*n*)). Machine learning models were trained on reaction parameters to gain unbiased insight into the underlying statistical trends in each composition space. A human-interpretable decision tree shows that phase selection is driven primarily by the bpy:Cu molar ratio for reactions containing Zr or Hf, and predicts that formation of the Ti-NCS compound requires that the amount

of HF present be decreased to raise the pH, which we verified experimentally. Predictive leave-one-metal-out (LOO) models further confirm that behavior in the Ti system is distinct from that of the Zr and Hf systems. The chemical origin of this distinction was probed via fluorine K-edge X-ray absorption spectroscopy. Pre-edge features in the F1s X-ray absorption spectra reveal the strong ligand-to-metal  $\pi$  bonding between Ti( $3d - t_{2g}$ ) and F( $2p$ ) states that distinguishes the  $\text{TiF}_6^{2-}$  anion from the  $\text{ZrF}_6^{2-}$  and  $\text{HfF}_6^{2-}$  anions.

**Introduction.** Noncentrosymmetric (NCS) materials are of wide commercial and academic interest owing to exciting properties, including ferroelectricity, piezoelectricity, and nonlinear optical activity, that are only allowed in materials lacking inversion symmetry.<sup>1,2</sup> Synthesis of NCS materials involves assembling functional basic building units (BBUs) and controlling their packing into an NCS crystal structure. In prominent NCS materials, such as  $\text{KTiOPO}_4$ ,<sup>3,4</sup>  $\text{BaTiO}_3$ ,<sup>5</sup> and  $\text{LiNbO}_3$ ,<sup>6</sup> inversion symmetry breaking originates from the cooperative alignment of second-order Jahn–Teller (SOJT) distortions in  $d^0$  early transition metal (ETM) octahedra.<sup>7–12</sup> Developing new NCS materials derived from  $d^0$  ETM octahedra is a challenge, however, as polar ETM octahedra tend to anti-align and adopt centrosymmetric arrangements. Although chiral templating agents have often been introduced to guarantee inversion symmetry breaking, further work is required to optimize and control interactions between polar and chiral BBUs to realize NCS materials with efficient properties.<sup>13–18</sup>

Recent reports indicate that racemic compounds, which contain both left- and right-handed enantiomers of a chiral BBU, can possess nonlinear optical and piezoelectric properties comparable to those of commercial NCS materials.<sup>19,20</sup> Racemates have long been observed to crystallize in NCS structures and are by no means rare, yet synthetic examinations of inversion symmetry breaking in racemic compounds have been limited.<sup>21–28</sup> Here, we present the targeted

synthesis of a series of polar racemates based on  $d^0$  ETM octahedra and racemic combinations of chiral copper-bipyridine (bpy) coordination complexes.

We applied a composition space approach to target the M = Ti, Zr members of the  $\Delta,\Lambda$ -[Cu(2,2'-bpy)<sub>2</sub>(H<sub>2</sub>O)]<sub>2</sub>[MF<sub>6</sub>]<sub>2</sub>·3H<sub>2</sub>O (M = Ti, Zr, Hf<sup>29-31</sup>; space group *Pna2<sub>1</sub>*) family. Composition space diagrams are useful tools for planning reactions and understanding reaction outcomes in terms of chemical trends by plotting reaction outcomes as a function of two or more variables, commonly the initial reactant mole fractions.<sup>32,33</sup> By examining the composition space of three (CuO, MO<sub>2</sub>)/bpy/HF(*aq*) (M = Ti, Zr, Hf) systems, we were able to locate the crystallization field, or region of selective crystallization, of each racemic compound and structurally characterize compounds that form under similar conditions.

In tandem with our experimental approach, supervised machine learning (ML) models were trained on reaction parameters to gain unbiased insight into each system. Supervised ML models are functions mapping between a set of features (i.e., model inputs), such as descriptions of a chemical experiment, to a known output (e.g., the reaction outcome). The data used to train the model must be representative of the problem as a whole, requiring sampling of the relevant dimensions of chemical space. This includes sampling the chemical properties, reagent concentrations, and reaction conditions in order to capture reactivity variations.<sup>34,35</sup> In practice, the datasets generated for use in ML need to be prepared and normalized such that they can be easily read into statistical analysis software packages. Models such as decision tree classifiers provide an unbiased prediction based upon underlying statistical patterns in the datasets using a human-interpretable function that can confirm or improve scientific insight.<sup>36</sup>

Following this approach, we found that polar  $\Delta,\Lambda$ -[Cu(bpy)<sub>2</sub>(H<sub>2</sub>O)]<sub>2</sub>[MF<sub>6</sub>]<sub>2</sub>·3H<sub>2</sub>O compounds (M = Ti, Zr, Hf), denoted as M-NCS, form in competition with a series of one-

dimensional chain compounds with the formula  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})\text{MF}_6$  ( $\text{M} = \text{Ti}$  ( $Pnma$ ),  $\text{Zr}$  ( $P\bar{1}$ ),  $\text{Hf}$  ( $P2/n$ )), denoted as M-CS. Decision tree classification of reaction outcomes indicates that phase selection in each system is driven by the molar ratio of 2,2'-bipyridine (bpy) to  $\text{Cu}^{2+}$ , with M-CS phases forming when  $\text{bpy}:\text{Cu}$  is less than 1.5 and M-NCS forming when  $\text{bpy}:\text{Cu}$  is greater than 1.5 for reactions containing Zr and Hf. This parameter does not accurately capture the crystallization boundary between Ti-CS and Ti-NCS, however, as Ti-NCS only forms when  $\text{bpy}:\text{Cu}$  is greater than 1.5 and less than 0.0025 moles of HF are present. Predictive models trained using a leave-one-metal-out (LOO) strategy were able to predict the outcomes of reactions containing Zr or Hf with high performance but had low performance when predicting reaction outcomes in the  $(\text{CuO}, \text{TiO}_2)/\text{bpy}/\text{HF}(aq)$  system. Fluorine K-edge X-ray absorption spectroscopy was performed to experimentally discern the distinction between reactions carried out with Ti and those with Zr and Hf. Examination of pre-edge features in the F1s spectra reveals significant  $\pi$  bonding between  $\text{Ti}(3d - t_{2g})$  and  $\text{F}(2p)$  states,<sup>37,38</sup> which also manifests in the diffraction data as short Ti–F bond distances and in the NMR data as a high frequency  $^{19}\text{F}$  shift relative to Zr–F and Hf–F. This strong  $\pi$  bonding in  $\text{TiF}_6^{2-}$  is emblematic of a suitably small energy gap that facilitates SOJT activity, giving the  $\text{TiF}_6^{2-}$  anion distinct behavior compared to  $\text{ZrF}_6^{2-}$  and  $\text{HfF}_6^{2-}$ .

## **Methods**

**Caution.** Hydrofluoric acid (HF) is toxic and corrosive! HF must be handled with extreme caution and the appropriate protective gear.

**Materials.**  $\text{TiO}_2$  (Aldrich, 99.9+%) ,  $\text{ZrO}_2$  (Alfa Aesar, 99.978%),  $\text{HfO}_2$  (Aldrich, 98%),  $\text{CuO}$  (Sigma-Aldrich,  $\geq 99.0\%$ ), 2,2'-bipyridine (bpy) (Sigma-Aldrich,  $\geq 99\%$ ), and  $\text{HF}(aq)$  (Sigma-

Aldrich, 48% wt. in H<sub>2</sub>O,  $\geq 99.99\%$  trace metals basis) were used as received. Reagent amounts of deionized water were used.

**Hydrothermal Synthesis.** Reactions were carried out following the hydrothermal pouch method.<sup>39</sup> The composition spaces of the (CuO, MO<sub>2</sub>)/bpy/HF(*aq*) (M = Ti, Zr, Hf) systems were explored by varying the amounts of bpy and HF(*aq*) used in each reaction, while the amounts of CuO and MO<sub>2</sub> (M = Ti, Zr, Hf) were held constant. Deionized water was added to each pouch to achieve a final solution volume of 1.1 mL. Full details of the reactions can be found in a spreadsheet included as Supporting Information. After heat sealing, six pouches were placed into a 125 mL Teflon-lined Parr autoclave with 40 mL distilled water as backfill. This larger pressure vessel was heated at a rate of 5 °C/min to 150 °C and held at 150 °C for 24 h, then allowed to cool at a rate of 6 °C/h. Solid products were recovered via vacuum filtration.

**Powder X-ray Diffraction.** Powder X-ray diffraction was used to assign reaction outcomes and assess phase purity. Measurements were carried out using Cu K $\alpha$ 1 ( $\lambda = 1.5418$  Å) radiation on STOE STADI-P and Rigaku IV Ultima diffractometers.

**Single-Crystal X-ray Diffraction.** Single-crystal X-ray diffraction was used to determine the structure of crystalline products. Diffraction patterns were recorded on Bruker-APEX II CCD diffractometers at 100 K with Mo K $\alpha$  ( $\lambda = 0.71073$  Å) radiation. Structures were solved with SHELXS and SHELXT and refined with SHELXL.<sup>40</sup> The diffraction data was integrated using SAINT.<sup>41</sup> Multi-scan absorption corrections were applied with SADABS.<sup>42</sup> No higher symmetry was found using symmetry checks in PLATON.<sup>43</sup> Hydrogen atom positions were located from difference maps and refined freely for the three M-CS structures. Hydrogen atoms bound to carbon atoms in 2,2'-bipyridine were attached in Olex2 using a riding model.<sup>44</sup> Hydrogen atoms of free

water molecules and water molecules bound to  $\text{Cu}(\text{bpy})_2(\text{H}_2\text{O})^{2+}$  units in the three M-NCS compounds could not be located from the difference maps and were omitted.

**Machine Learning.** The dataset used here consists of 51 experiments transcribed from laboratory notebook records, using the ESCALATE "entity, materials, actions, observations" ontology.<sup>45</sup> In addition to this raw experimental data, additional calculated stoichiometric properties and computed electronic structure properties were added. Stoichiometric features, such as molar amounts and molar ratios, were calculated directly from the experimental observations. Electronic structure calculations were performed on the  $\text{TiF}_6^{2-}$ ,  $\text{ZrF}_6^{2-}$ , and  $\text{HfF}_6^{2-}$  anionic building units to provide data on geometry, energetics, and charges using Gaussian 09, Revision D.01.<sup>46</sup> The B3LYP/LANL2DZ model chemistry was used as it provides good estimations ( $\pm 10$  pm) for bond lengths of transition metal oxides and halides.<sup>47</sup> Atomic charges were assessed for the optimized geometries using Mulliken, Hirshfeld, CM5,<sup>48</sup> Natural Bond Orbital (NBO), and electrostatic potential fitting (Merz–Singh–Kollman using UFF radii, MKUFF) methods. The optimized geometries and output files are in the Supplementary Information. The optimized geometries were used for bond valence sum calculations. Features that are constant in all experiments, such as  $\text{HF}(aq)$  concentration, or those that are potentially misleading, such as the mass of the early transition metal were removed before performing ML calculations, reducing the number of features from 76 to 33. To improve model clarity, we performed feature selection using ANOVA F-value analysis using the `f_classif` function in SciKitLearn to isolate the three most important features: bipyridine-to-copper molar ratio, fluorine average NBO, and bpy normalized molar amount. The decision tree trained using only the top three features outperformed one trained on the full feature set on precision, accuracy, and recall. Other more sophisticated models are always

provided the full feature set. Additional results and discussion can be found in the Supporting Information.

ML modeling was performed using SciKitLearn version-0.21.2.<sup>49</sup> The relevant hyperparameters (e.g., tree depth, leaf sample splits, and minimum samples per leaf for decision trees) were optimized through a permuted grid search varying and compared across two different hold-out regimes. In a standard test train split (STTS) hold-out regime, a 5-fold cross validation strategy is used; for each of the five cycles of the cross validation, 80% of the data are used for training the model (i.e., the model is given inputs AND outputs) and 20% is used for testing the model predictions. In a leave-one-metal-out (LOO) hold-out regime, the dataset is divided into three groups based on the metal (Ti, Zr, Hf); two metals are used for training and the remaining metal is used for testing. For example, in one iteration of LOO model testing, Zr and Hf data is used to train the model and the performance is tested on how well the model predicted the outcomes of Ti. Model performance is reported using the average performance across all of the ‘unseen’ test groups (i.e., the left-out metal for LOO or the test samples for STTS).

Class imbalances, meaning divergent observation rates for different outcomes, can lead to challenges when relying upon model efficacy parameters such as precision and accuracy because they can be deceptive for datasets with imbalanced outcomes, such as the prevalence of M-CS or M-NCS phases in this dataset. Instead, Matthews correlation coefficients (MCC) are reported, as this metric is not affected by class imbalances. Models that only predict the majority class will have an MCC of zero; an MCC of ‘1’ corresponds to perfect prediction of both M-CS and M-NCS. To establish the quality of the model we considered four performance baselines, corresponding to predicting the majority class for every task (‘Majority (B)’), randomly guessing with the probability as the training data (‘Probability (B)’), a support vector classification (SVC) model

trained on randomly shuffled data ('Shuffled, SVC (B)'), and a model that uses the closest example in the training set as a proxy for memorization ('kNN, k=1'). A detailed description can be found in the Supporting Information.

**X-ray Absorption Spectroscopy.** X-ray absorption spectroscopy measurements were performed at Argonne National Laboratory Advanced Photon Source beamline 4-ID-C. The fluorine K-edge was scanned from 680 to 710 eV. Data was collected in both the total-electron-yield and total-fluorescence-yield mode utilizing photocurrent for the electron yield and a silicon drift diode detector for the fluorescence yield. For all samples, the fluorescence yield produced more signal than the electron yield and was more reliable between runs. Three scans were performed on each sample and averaged to maximize the signal-to-noise ratio.

**Solid-State Nuclear Magnetic Resonance Spectroscopy.**  $^1\text{H}$  and  $^{19}\text{F}$  solid-state nuclear magnetic resonance (NMR) spectra were recorded in a static magnetic field of 9.4 T with a Bruker Avance III spectrometer. The samples were packed into 1.6 mm diameter zirconia rotors and spectra were recorded at 40 kHz magic-angle spinning (MAS) in a Phoenix narrow-bore 1.6 mm HFX probe.  $T_1$  (spin-lattice) relaxation was measured with a saturation recovery pulse.  $^1\text{H}$  spectra were measured with a rotor-synchronized Hahn-echo ( $\pi/2-\tau-\pi-\tau$ -acquire) pulse sequence using a  $90^\circ$  RF pulse of 1.22  $\mu\text{s}$  and a recycle delay of 15 s.  $^{19}\text{F}$  spectra were measured with the same pulse sequence with a  $90^\circ$  RF pulse of 1.3–1.75  $\mu\text{s}$  and a recycle delay of 0.02 s.  $^{19}\text{F}$  spectra were also collected with a recycle delay of 15 s and no difference was observed. For each  $^1\text{H}$  measurement, 64 scans were co-added; for  $^{19}\text{F}$ , 1024 scans were co-added. Variable temperature spectra were collected with a Bruker Cooling Unit (BCU) and heater.  $^1\text{H}$  and  $^{19}\text{F}$  spectra were externally referenced to adamantane at +1.8 ppm and the center of the doublet in  $\text{NaPF}_6$  at -82.5 ppm, respectively. The actual sample temperature under MAS and external temperature control was

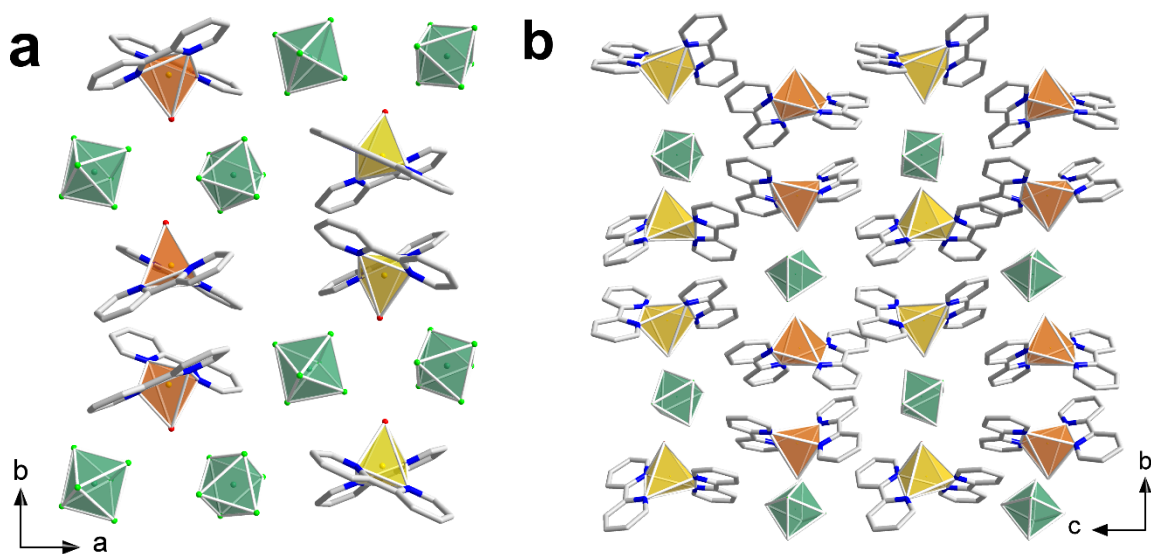


calibrated with the  $^{207}\text{Pb}$  shift of lead nitrate.<sup>50,51</sup> Frictional heating at 40 kHz MAS leads to an internal sample temperature of 44 °C. Temperatures under heating and cooling are given in the figures.

## Results

### Structure Descriptions

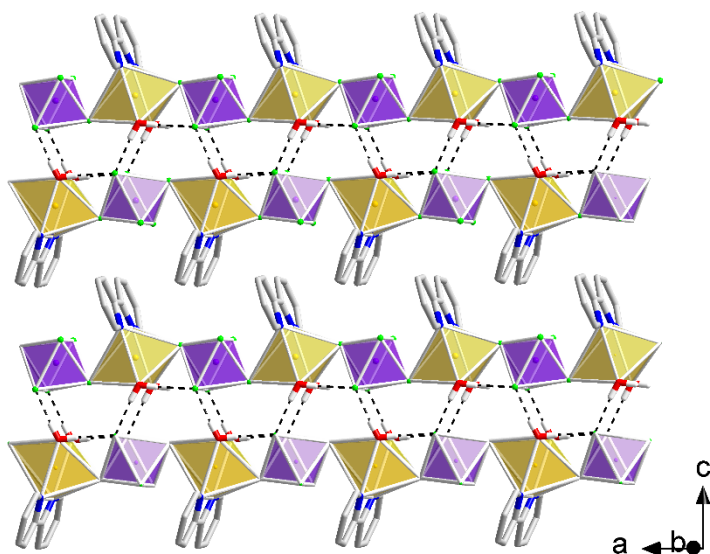
*Noncentrosymmetric polar racemates.* The  $\Delta,\Lambda$ -[Cu(bpy)<sub>2</sub>(H<sub>2</sub>O)]<sub>2</sub>[MF<sub>6</sub>]<sub>2</sub>·3H<sub>2</sub>O (M = Ti, Zr, Hf<sup>29</sup>) family of isostructural racemic compounds crystallizes in the polar space group *Pna*2<sub>1</sub>. These compounds are denoted as M-NCS (M = Ti, Zr, Hf). The structure class (Figure 1) contains both  $\Delta$ - and  $\Lambda$ -Cu(bpy)<sub>2</sub>(H<sub>2</sub>O)<sup>2+</sup> molecular cations and MF<sub>6</sub><sup>2-</sup> (M = Ti, Zr, Hf) anions that are arranged via hydrogen bonding and  $\pi$ -stacking interactions. Two independent BBUs of each type are present in the asymmetric unit, with all atoms located on general positions. Homochiral pairs of



**Figure 1:** (a) The crystal structure of  $\Delta,\Lambda$ -[Cu(bpy)<sub>2</sub>(H<sub>2</sub>O)]<sub>2</sub>[MF<sub>6</sub>]<sub>2</sub>·3H<sub>2</sub>O (M-NCS, M = Ti, Zr, Hf) along the *c* axis. The polar structure is composed of chiral  $\Delta/\Lambda$ -Cu(bpy)<sub>2</sub>(H<sub>2</sub>O)<sup>2+</sup> ( $\Delta$  – orange,  $\Lambda$  – yellow) cations and weakly polar MF<sub>6</sub><sup>2-</sup> (green) anions. Free water molecules and all hydrogen atoms have been omitted for clarity. (b) View of the M-NCS structure along the *a* axis. Free water molecules and all hydrogen atoms have been omitted for clarity.

$\text{Cu}(\text{bpy})_2(\text{H}_2\text{O})^{2+}$  cations pack in an alternating fashion with pairs of  $\text{MF}_6^{2-}$  BBUs along the  $b$  direction (Figure 1a). These homochiral columns are related only by glide planes to columns containing  $\text{Cu}(\text{bpy})_2(\text{H}_2\text{O})^{2+}$  cations with the opposite handedness. Hydrogen bonding contacts with each  $\text{MF}_6^{2-}$  unit give rise to differences in M–F bond lengths (between 1.799 Å and 1.889 Å for Ti–F bonds, 1.938 Å and 2.024 Å for Zr–F bonds, and 1.942 Å and 2.021 Å for Hf–F bonds). The differences in bond lengths result in small polar distortions, with the central M atom being displaced less than 0.006 Å from the center of each octahedron (Table S1). The polar moments of the  $\text{MF}_6^{2-}$  anions are partially aligned along  $c$ .

*Centrosymmetric Chain Compounds.* Three distinct centrosymmetric compounds were observed in this work, denoted as M-CS (M = Ti, Zr, Hf). In contrast to the three M-NCS compounds described above,  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2\text{TiF}_6$  (Ti-CS),<sup>52</sup>  $[\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2]_2[\text{Zr}_2\text{F}_{12}]$  (Zr-CS), and  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2\text{HfF}_6$  (Hf-CS) are not isostructural, although the M-CS compounds do share two

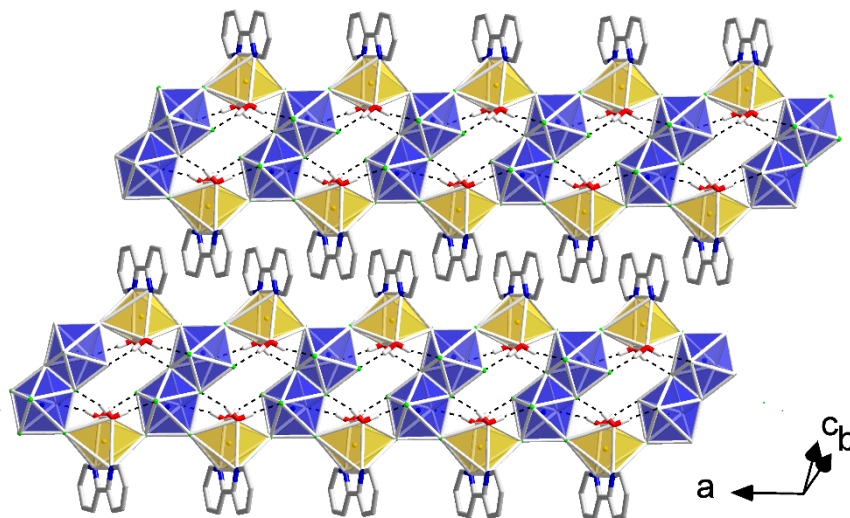


**Figure 2:** The crystal structure of  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2\text{TiF}_6$  (Ti-CS) features polar one-dimensional chains. Orange and purple polyhedra represent  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2^{2+}$  and  $\text{TiF}_6^{2-}$  building units, respectively.

structural motifs: (1) each Cu center is bound to a single bpy ligand and (2) each of the three M-CS phases contains one-dimensional chains of alternating  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2^{2+}$  and ETM-fluoride BBUs.

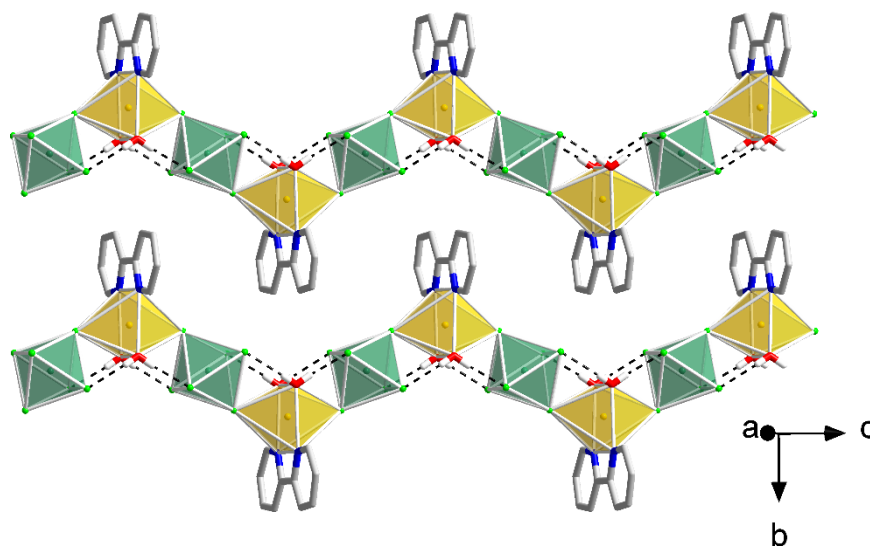
The structure of  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2\text{TiF}_6$  (Ti-CS) (Figure 2), originally reported in 2013, contains polar zigzag chains composed of alternating  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2^{2+}$  and  $\text{TiF}_6^{2-}$  units.<sup>52</sup> Ti-CS crystallizes in the space group *Pnma* and is a member of an isostructural series of 1D chain compounds with the general formula  $\text{A}(\text{bpy})(\text{H}_2\text{O})_2\text{A}'\text{O}_x\text{F}_{6-x}$  ( $\text{A}/\text{A}' = \text{Cu}/\text{Ti}, \text{Cu}/\text{V}, \text{Cu}/\text{Nb}, \text{Cu}/\text{Mo}, \text{Zn}/\text{Mo}, \text{Zn}/\text{W}$ ).<sup>52</sup> Differences in Ti–F bond distances in the  $\text{TiF}_6^{2-}$  anion reveal a  $\text{C}_2$ -type second-order Jahn–Teller (SOJT) distortion of the Ti atom toward the edge of the coordination octahedron. The central Ti atom of the  $\text{TiF}_6^{2-}$  unit is displaced 0.011 Å out of the center of the coordination octahedron. The displacement occurs perpendicular to the chain direction, resulting in two short Ti–F1 bonds with distances of 1.7945(7) Å and two long Ti–F2 bonds with distances of 1.9414(7) Å. The Ti–F3 bonds along the chain have distances of 1.8749(9) Å and 1.8600(10) Å. The distortion is facilitated by inter- and intrachain hydrogen bonding interactions with F2 and water molecules in  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2^{2+}$  units, with O1–F2 distances of 2.639 Å and 2.700 Å, respectively. Adjacent chains are arranged in the *ac* plane via hydrogen bonding interactions without stacking of bpy ligands.

Zr-CS crystallizes in the space group  $P\bar{1}$  and has the formula  $[\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2]_2[\text{Zr}_2\text{F}_{12}]$ . The structure of Zr-CS (Figure 3) features bridged “ladder” chains based on alternating  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2^{2+}$  and  $\text{Zr}_2\text{F}_{12}^{4-}$  BBU's. The anionic  $\text{Zr}_2\text{F}_{12}^{4-}$  cluster is similar to the  $\text{V}_2\text{O}_2\text{F}_8^{2-}$  cluster found in  $[\text{Cu}(\text{bpy})(\text{H}_2\text{O})]_2[\text{V}_2\text{O}_2\text{F}_8]$ , which contains bridged chains wherein each cluster is involved in six bridging V–F–Cu linkages.<sup>53</sup> The  $\text{Zr}_2\text{F}_{12}^{4-}$  clusters in Zr-CS bridge between  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2^{2+}$  units to form chains through four Zr–F–Cu linkages. Unlike the  $\text{V}_2\text{O}_2\text{F}_8^{2-}$  cluster, which features edge-sharing  $\text{VOF}_5^{2-}$  octahedra, the  $\text{Zr}_2\text{F}_{12}^{4-}$  cluster contains edge-sharing pentagonal bipyramidal  $\text{ZrF}_7^{3-}$  units. The Zr–F bond distances range from 2.0021(7) to 2.2050(7) Å. The central Zr atom is displaced 0.016 Å from the center of the pentagonal bipyramid as a result of cation-cation repulsion between neighboring Zr atoms in the edge-sharing  $\text{Zr}_2\text{F}_{12}^{4-}$  cluster. Hydrogen bonding interactions are present along the chain direction linking the  $\text{Zr}_2\text{F}_{12}^{4-}$  clusters and the coordinated water molecules in  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2^{2+}$ , with O–F distances of 2.593 and 2.639 Å. Adjacent chains are interlocked to form sheets via face-on  $\pi$  stacking contacts.



**Figure 3:** Crystal structure of Zr-CS (formula:  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2]_2[\text{Zr}_2\text{F}_{12}]$ ). Zr-CS features 1D chains composed of  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2^{2+}$  (orange) and  $\text{Zr}_2\text{F}_{12}^{4-}$  (blue) units.

The structure of  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2\text{HfF}_6$  (Hf-CS) (Figure 4) features nonpolar zigzag chains composed of alternating  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2^{2+}$  and  $\text{HfF}_6^{2-}$  units. The octahedral  $\text{HfF}_6^{2-}$  anion is undistorted with Hf–F bond distances between 1.9992(11) and 2.0052(10) Å and the Hf atom occupying an inversion center. Each equatorial fluorine participates in two hydrogen bonding interactions along the chain with O–F distances of 2.628 and 2.646 Å.

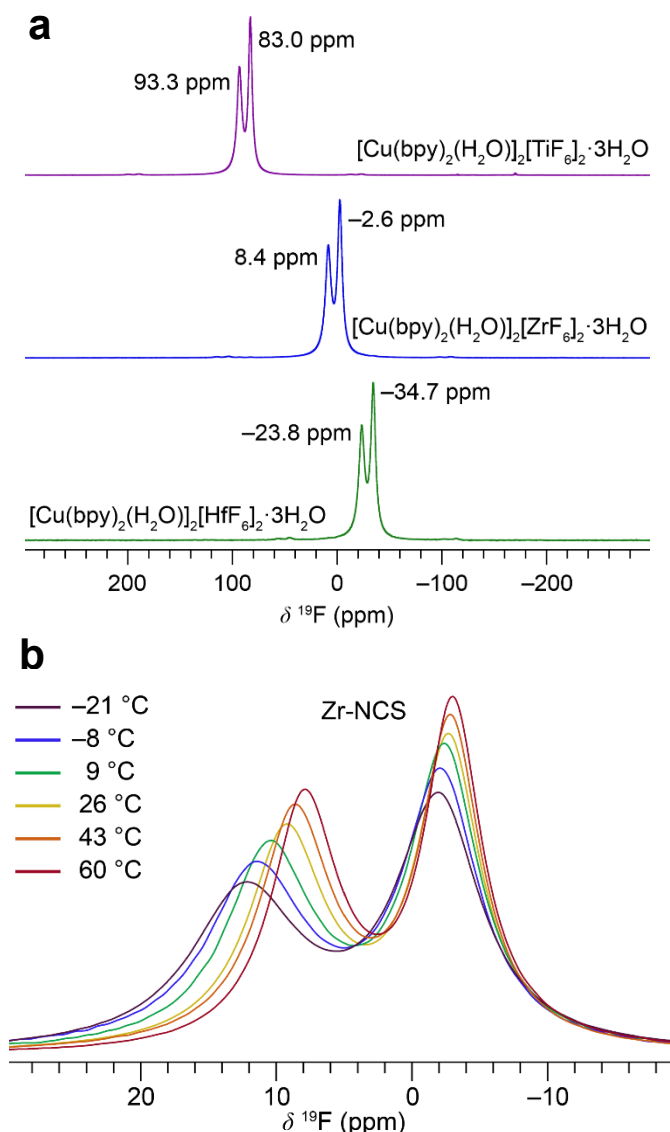


**Figure 4:** Crystal structure of  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2\text{HfF}_6$  (Hf-CS). Hf-CS contains nonpolar zig-zag chains based on  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2^{2+}$  (orange) and  $\text{HfF}_6^{2-}$  (green) basic building units.

### Solid-State NMR

Solid-state  $^{19}\text{F}$  MAS NMR measurements were performed to analyze the local environment of the  $\text{MF}_6^{2-}$  anions in each structure.  $^{19}\text{F}$  MAS NMR spectra of the isostructural paramagnetic M-NCS (M = Ti, Zr, Hf) compounds are shown in Figure 5. All three compounds show two resonances separated by about 10 ppm that integrate to 1:1 with a less than 0.5% error (Figure 5a). According to the crystal structure, there are 12 unique F atoms. Thus, the two  $^{19}\text{F}$  signals are assigned to the two distinct  $\text{TiF}_6^{2-}$  octahedra, each signal representing six crystallographically-unique but motionally-averaged fluorine atoms. This hypothesis is supported by variable-

temperature  $^{19}\text{F}$  MAS NMR spectra measured from  $-21$  to  $+60$   $^{\circ}\text{C}$  (Figure 5b, Figure S1), which show that the linewidth decreases from low to high temperatures. Furthermore, the small spinning sidebands (Figure 5a) in the  $^{19}\text{F}$  spectra indicate that the chemical shift anisotropy (CSA) is almost completely suppressed by dynamic averaging even at ambient temperature. Spectral fitting suggests a CSA of no more than  $\pm 35$  ppm, while typical early transition metal fluoride CSAs are much larger: e.g., 170 ppm in  $\beta\text{-ZrF}_4$ ,<sup>54</sup> 300 ppm in  $\text{NbF}_5$  and  $\text{TaF}_5$ ,<sup>55</sup> and 900 ppm in  $\text{TiF}_4$ .<sup>56</sup> The



**Figure 5:** (a)  $^{19}\text{F}$  solid-state MAS NMR spectra for the three M-NCS (M = Ti, Zr, Hf) compounds. (b) Variable temperature  $^{19}\text{F}$  solid-state MAS NMR spectra for Zr-NCS.

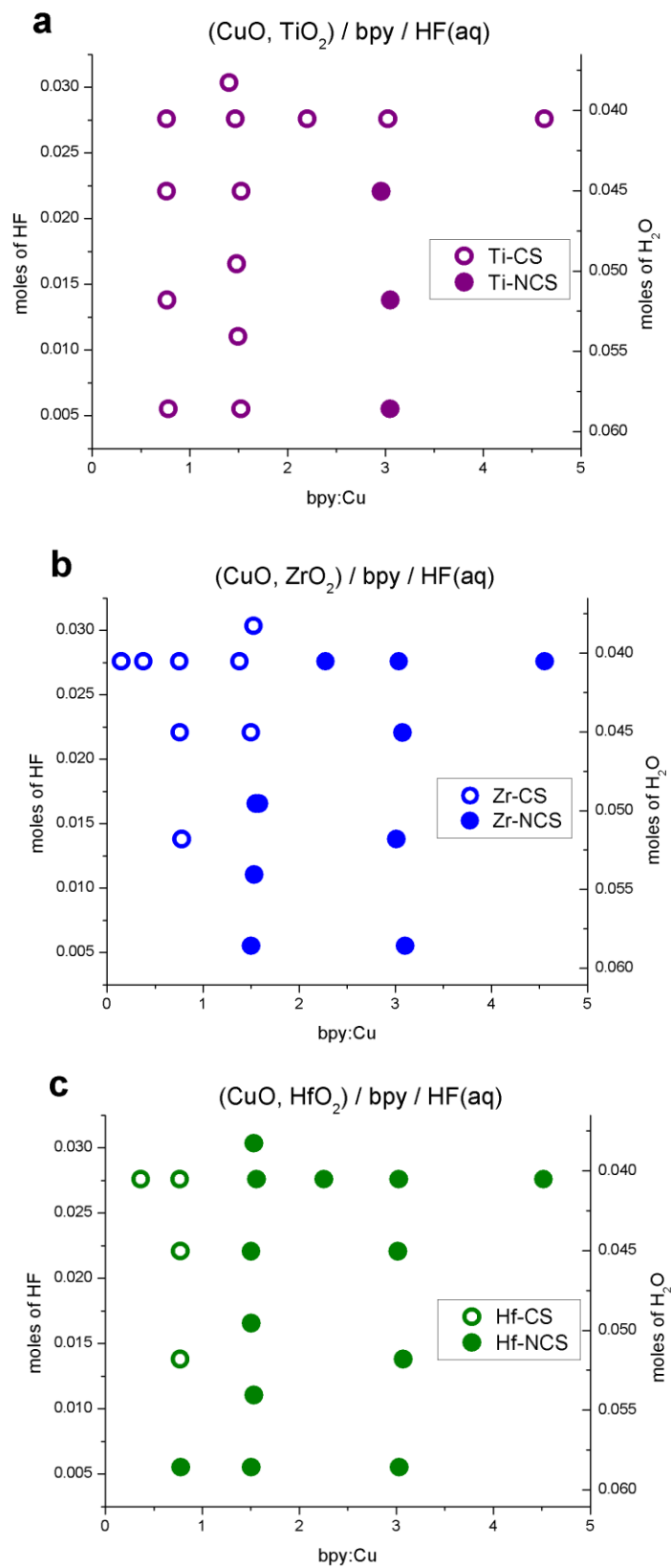
$^{19}\text{F}$  resonances systematically shift to lower frequency from Ti to Zr to Hf. In each case, the two  $^{19}\text{F}$  resonances are separated by 10–11 ppm, which is most likely attributable to the difference in average bond distance within the  $\text{MF}_6^{2-}$  octahedra (always about 0.005–0.01 Å) that arise from differences in hydrogen bonding between the two sites. These  $^{19}\text{F}$  spectra indicate that the local environments and dynamic behavior of the  $\text{MF}_6^{2-}$  anions in the three M-NCS compounds are identical, consistent with the single crystal structures. The motional dynamics of the isolated octahedra will likely influence the properties of these materials, which may be of interest for future studies. The  $^1\text{H}$  spectra of all three NCS compounds are identical, displaying broad resonances centered around 11 and 16 ppm (Figure S2), and do not vary with temperature (Figure S3).

The situation in the non-isostructural paramagnetic M-CS phases is considerably different. Evidently the paramagnetic effects are so strong in Ti-CS that neither  $^1\text{H}$  (Figure S2) nor  $^{19}\text{F}$  (Figure S4) signals were observed. In the  $^{19}\text{F}$  NMR of Zr-CS, four resonances of equal integrated intensity were observed. The compound has six distinct fluorine sites in the asymmetric unit; however, two of these fluorine atoms are directly bound to copper as  $\text{Zr-F-Cu}$  bridges whereas three are terminal  $\text{Zr-F}$  and one is bridging  $\text{Zr-F-Zr}$ . The  $\text{Zr-F-Cu}$  fluorine atoms are likely not observed due to rapid paramagnetic relaxation while the terminal F are assigned to the three resonances at 37, 25, and 15 ppm and the signal at –63 ppm is assigned to  $\text{Zr-F-Zr}$ . These assignments are made on the basis of the bond lengths; shorter M–F bonds (such as terminal fluorine) experience a larger paramagnetic contribution to shielding and thus appear at higher frequencies than longer M–F bonds (such as bridging fluorine).<sup>54</sup> N.B. the paramagnetic term described here is not directly related to Cu, rather it is the second-order paramagnetic contribution to the local magnetic field from the orbital motion of valence electrons as originally described by Ramsey, Saika, and Slichter.<sup>57,58</sup> Finally, in the Hf-CS compound, there are three distinct fluorine

sites, again one is bridging Hf–F–Cu while two are terminal Hf–F. It appears that the two distinct terminal fluorine are overlapping with an isotropic shift of 2 ppm. The  $^{19}\text{F}$  Hf-CS assignment is based on the nearly identical bond distances of Hf–F2 (1.9992(11) Å) and Hf–F3 (1.9994(11) Å), the similarity to the shifts of the terminal Zr–F, and the broader linewidth of the Hf-CS  $^{19}\text{F}$  signal as compared to the individual terminal  $^{19}\text{F}$  signals in Zr–F. Zr-CS and Hf-CS both showed three broad proton signals from 10 to 70 ppm.

Given the presence of  $d^9$  Cu(II), paramagnetic shifts and paramagnetic relaxation enhancement are expected. The  $^1\text{H}$  shifts are well outside the normal shift range though the  $^{19}\text{F}$  signals that are observable are within the range of diamagnetic fluoride compounds. The spin–lattice relaxation of all  $^1\text{H}$  and  $^{19}\text{F}$  compounds was below 20 ms, which is consistent with the expected dipolar relaxation rate for  $d^9$  Cu(II).<sup>59</sup>





**Figure 6:** Composition space diagrams for the  $(\text{CuO}, \text{MO}_2)/\text{bpy}/\text{HF(aq)}$  ( $M =$  (a) Ti, (b) Zr, (c) Hf) systems.

## Composition Space Analysis

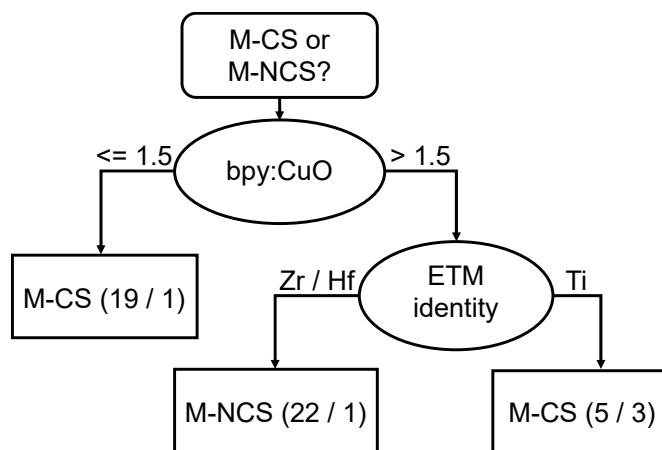
Machine learning efforts can provide considerable insight into the domain under inspection. This process begins with an evaluation of the dataset, which was manifested in two ways in the work presented here. First, a series of visualizations were created to help evaluate the extent of the explorations. Second, working with the experimental data allowed for the identification of outliers and provided insight into data input errors, which has improved the future reusability and relevance of the dataset for use in larger machine learning projects.

The formation of  $\text{Cu}(\text{bpy})_2^{2+}$  species is favored either by increasing pH or raising the bpy:Cu molar ratio, as seen in the relevant distribution diagrams from previous studies of aqueous copper-bipyridine systems.<sup>60</sup> The initial experiments investigated the *independent* roles of the bpy:Cu ratio and pH in each  $(\text{CuO}, \text{MO}_2)/\text{bpy}/\text{HF}(\text{aq})$  ( $\text{M} = \text{Ti}, \text{Zr}, \text{Hf}$ ) system, which lead to the synthesis of Ti-CS, Zr-NCS, Zr-CS, Hf-NCS, and Hf-CS. Visualizing this data revealed the need for additional experiments where *both* the bpy:Cu ratio and the amount of HF were varied. Ultimately the Ti-NCS compound was discovered in the region of composition space with simultaneously high bpy:Cu ratio and high pH (low amount of HF) as a direct result of the data visualizations that were performed, and enabled the subsequent machine learning work to probe the formation of this additional phase.

In the preliminary data visualizations, outlier experiments consisting of isolated reaction products surrounded by other phases were observed. A re-analysis of the powder X-ray diffraction phase identification revealed incorrect phase identity determinations, which would have negatively impacted the resulting machine learning models. The cleaned and curated dataset has also been provided along with the relevant characterization in a machine readable format on the materials data facility.<sup>61–63</sup>

Composition space diagrams (Figure 6) were constructed for each (CuO, MO<sub>2</sub>)/bpy/HF(aq) (M = Ti, Zr, Hf) system. A total of 51 reactions were conducted across the three systems, with varying HF and bpy concentrations targeted for exploration. Identification of the dominant solid product for each individual reaction led to the observation of both a polar racemate (M-NCS) and a centrosymmetric one-dimensional chain compound (M-CS) (M = Ti, Zr, Hf) in each system. The crystallization fields for these six compounds vary in terms of both position and breadth.

A decision tree algorithm was employed to provide interpretable quantification of key experimental properties and the associated numerical cutoffs. The decision tree model shown in Figure 7 was fit on bipyridine-to-copper ratio, fluorine average NBO, and bpy normalized molar amount; these were the three features selected algorithmically through feature selection to avoid overfitting (see methods section). The model targeted classification of each experiment based on whether it produced M-NCS or M-CS.



**Figure 7:** Decision tree classification of M-CS versus M-NCS. The three entries in terminal leaves indicate the symmetry of largest class in the leaf (either CS or NCS), along with the number of correct predictions and the number of incorrect predictions, respectively.

The decision tree model (Figure 7) captures the primary division in the dataset; M-CS compounds are observed when the ratio of bipyridine to copper is less than or equal to 1.5, with

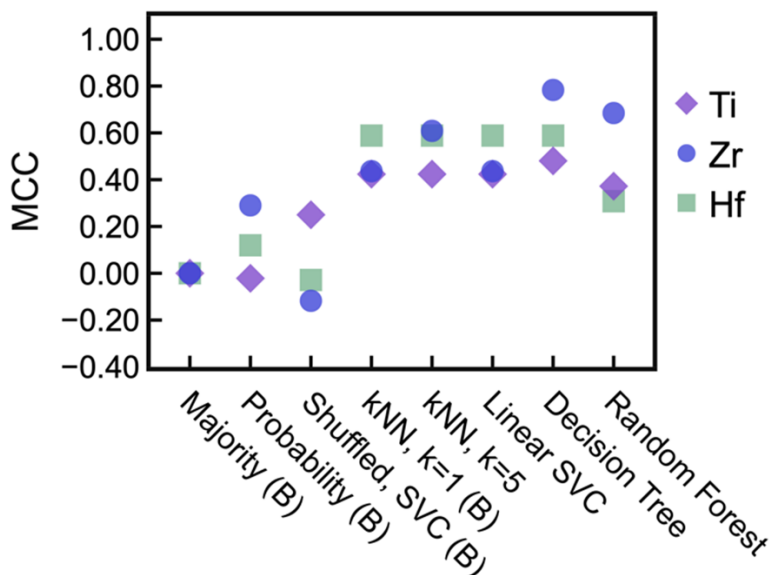
19 of the 20 reactions performed with a bpy:Cu ratio less than 1.5 being correctly classified as producing M-CS. Continuing to the right internal node, the identity of the transition metal distinguishes the reaction outcome when the bpy:Cu ratio is above 1.5. The identity of the metal was described via the metal-NBO feature, the numerical value of which corresponds to the charge on the  $[\text{MF}_6]^{2-}$  metal center. All electronic structure features related to the  $[\text{MF}_6]^{2-}$  anions are linearly correlated (see Feature Covariance Matrix Figure S5), indicating that the predominant characteristic of the feature is not the numerical value but rather the correspondence to a given metal. The decision tree correctly classifies 22 out of 23 reactions involving Zr and Hf with bpy:Cu > 1.5 as producing M-NCS. For Ti reactions, however, five out of the eight reactions classified in this leaf produced Ti-CS, meaning that the tree does not produce a leaf in which the majority phase is Ti-NCS. Therefore, the decision tree model is able to effectively classify reactions for Zr and Hf but does not fully capture the crystallization boundary between Ti-CS and Ti-NCS.

**Table 1:** Matthews correlation coefficient (MCC) for each leave-one-metal-out (LOO) model. “B” in the model name signifies a baseline comparison.

| Model Name        | Ti    | Zr    | Hf    |
|-------------------|-------|-------|-------|
| Majority (B)      | 0.00  | 0.00  | 0.00  |
| Probability (B)   | −0.02 | 0.29  | 0.12  |
| Shuffled, SVC (B) | 0.25  | −0.12 | −0.03 |
| kNN, k = 1 (B)    | 0.42  | 0.44  | 0.59  |
| kNN, k = 5        | 0.42  | 0.61  | 0.59  |
| Linear SVC        | 0.42  | 0.44  | 0.59  |
| Decision Tree     | 0.48  | 0.78  | 0.59  |
| Random Forest     | 0.37  | 0.68  | 0.31  |

The limited efficacy of the decision tree for differentiating the boundary between Ti-CS and Ti-NCS motivated us to develop and compare against three other more sophisticated ML algorithms (i.e., support vector classifiers [SVCs], random forests, and k-nearest neighbors [kNN]), as well as different hold-out regimes including leave-one-metal-out in addition to the five-fold test-train split (STTS) (see Methods for details). The leave-one-out (LOO) models are denoted as “extrapolative” as they are constructed using data from two early transition metals as the training set and predicting crystallization outcomes for the third metal that was “left out” as the test set. Standard test train split models are denoted as ‘interpolative’ owing to the high degree of overlap between the data within each fold of the cross-validation. Interpolative STTS models are generally expected to perform better than comparable ‘extrapolative’ LOO models. Baseline models are

denoted with a “(B)” and are explained in the methods section. MCC for each fold of the leave-one-metal-out models are shown in Table 1 and Figure 8.



**Figure 8:** Visualization of the Matthews correlation coefficient (MCC) for extrapolative leave-one-metal-out (LOO) models. “B” in the model name signifies a baseline comparison.

**Table 2.** Matthews correlation coefficient (MCC) for each standard test train split (STTS) model. “B” in the model name signifies a baseline comparison.

| Model Name        | MCC              |
|-------------------|------------------|
| Majority (B)      | 0.00             |
| Probability (B)   | $-0.05 \pm 0.18$ |
| Shuffled, SVC (B) | $-0.14 \pm 0.25$ |
| kNN, k = 1 (B)    | $0.78 \pm 0.15$  |
| kNN, k = 5        | $0.52 \pm 0.16$  |
| Linear SVC        | $0.79 \pm 0.17$  |
| Decision Tree     | $0.74 \pm 0.29$  |
| Random Forest     | $0.31 \pm 0.36$  |

Evidence for a distinction between the Ti reactions with respect to the Zr and Hf reactions is found in the decision tree shown in Figure 7, as discussed above. This evidence is found in the presence of a node that separates Ti from the Zr and Hf reactions. Support for this distinction can also be found in the LOO models. Models trained through LOO hold-out regimes generally perform slightly worse when predicting reaction outcomes when Ti is the target, better for Hf, and best for Zr as the target (Figure 8). For the LOO task, the decision tree has the highest MCC of 0.78 for Zr, 0.59 for Hf, and 0.48 for Ti. From these analyses we conclude that the interpretable decision tree model is competent at differentiating M-CS and M-NCS and is statistically competitive with other more sophisticated models for LOO training regimes.

In contrast with the LOO models, the interpolative STTS models are trained with data from all three early transition metals (Ti, Zr, Hf) and consequently are expected to perform better than LOO models across all metrics. A plot detailing the MCCs for the STTS models is shown in Table 2. The highest observed MCCs,  $0.78 \pm 0.15$  for kNN  $k = 1$  and  $0.79 \pm 0.17$  for linear SVC, were similar to the performance of the interpretable decision tree, which yielded an MCC of  $0.74 \pm 0.29$ , but with a lower standard deviation. The similar predictive power of other ML methods compared with decision tree performance indicates that three features (bipyridine-to-copper molar ratio, fluorine average NBO, and moles of bipyridine) are sufficient to describe this dataset and there is no loss in predictive power by using an interpretable decision tree model. This confirms what is observed via human inspection, namely that phase selection is primarily driven by the bpy:Cu ratio but that this parameter does not fully describe the phase boundary between Ti-NCS and Ti-CS. Furthermore, all tested models outperform the benchmarks with the exception of the memorization benchmark, which indicates that the search space is sufficiently sampled to distinguish the

crystallization boundary between the M-NCS and M-CS phases using solely the most similar previous reactions.

### **X-ray Absorption Spectroscopy.**

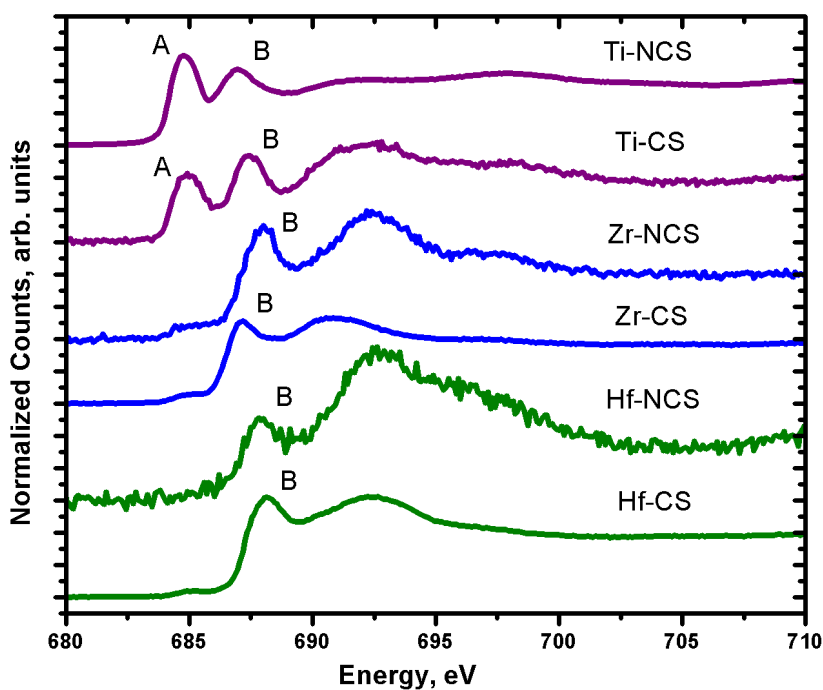
Fluorine K-edge X-ray absorption spectroscopy measurements were performed to examine covalency in each  $\text{MF}_6^{2-}$  anion (Figure 9).<sup>38,64–67</sup> Previous studies have shown that fluorine K-edge spectroscopy, which probes electronic transitions from the filled F(1s) orbitals to empty orbitals with F(np) character, can be used to probe metal-ligand covalency.<sup>65</sup> Because the F(1s) orbital is highly localized, excitations to empty metal *d* states can only be observed when the metal *d* orbitals are hybridized with F(2p) orbitals.<sup>66</sup>

In these spectra, two regions can be distinguished. First are the features in the pre-edge region (< 690 eV) corresponding to F(2p) orbitals hybridizing with empty Ti(3*d*), Zr(4*d*), and Hf(5*d*) states. And second, in the main edge region (> 690 eV), these features correspond to F(3p) orbitals hybridizing with empty Ti(4*sp*), Zr(5*sp*), and Hf(6*sp*) states, and, at higher energies, multiple scattering events of the completely unbound electron. Since the transition metals are octahedrally coordinated, features in the pre-edge region provide direct evidence of crystal field splitting effects on the metal *d* orbitals.

F1s spectra for the Ti-based compounds, Ti-NCS and Ti-CS, are in good agreement with previous measurements of the F K-edge in  $\text{K}_2\text{TiF}_6$ .<sup>68–70</sup> The pre-edge peak at 685 eV (A) represents covalent mixing between Ti(3*d* – *t*<sub>2g</sub>) and F(2p) states, while the pre-edge peak at 687 eV (B) represents covalent mixing between Ti(3*d* – *e*<sub>g</sub>) and F(2p) states. The presence of pre-edge features A and B indicates significant  $\sigma$  and  $\pi$  character in the bonding of the  $\text{TiF}_6^{2-}$  anion.



Examining the spectra recorded for Zr- and Hf-based compounds, a dominant pre-edge feature B at 687 eV (B) is observed. For Zr-NCS, Hf-NCS, and Hf-CS, which contain octahedral  $\text{ZrF}_6^{2-}$  and  $\text{HfF}_6^{2-}$  anions, the strong pre-edge peak B is consistent with  $\text{Zr}(4d - e_g)$  and  $\text{Hf}(5d - e_g)$  hybridization with  $\text{F}(2p)$  states. The presence of a weak pre-edge peak at 685 eV illustrates the limited hybridization between  $\text{Zr}(4d - t_{2g})/\text{Hf}(5d - t_{2g})$  and  $\text{F}(2p)$  states. There was too much noise in the spectrum of Hf-NCS to observe any pre-edge feature at 685 eV. Peak B is dominant in Zr-NCS, Hf-CS, and Hf-NCS because orbital overlap is higher for the primary  $\sigma$  crystal field splitting interactions in octahedral configurations, which occur along the bonding axis ( $e_g$ ), compared to  $\pi$  bonding interactions ( $t_{2g}$ ). The dominant pre-edge feature B in the spectrum for Zr-CS is consistent with the presence of  $\sigma$  interactions through the primary crystal field splitting in a 7-coordinate pentagonal bipyramidal environment via mixing of  $\text{Zr}(4d - a_1')$  and  $\text{F}(2p)$  states.<sup>71</sup> The low



**Figure 9:** F1s absorption spectra for Ti-NCS, Ti-CS, Zr-NCS, Zr-CS, Hf-NCS, and Hf-CS. The pre-edge feature A represents  $\text{M}(nd - t_{2g}) - \text{F}(2p)$  covalency, while the pre-edge feature B represents  $\text{M}(nd - e_g) - \text{F}(2p)$  covalency.

intensity of the feature at 685 eV indicates the limited hybridization of the  $\text{Zr}(4d - e_2')$  and  $\text{Zr}(4d - e_1'')$  orbitals with  $\text{F}(2p)$  states. The observations clearly indicate that the  $\pi$  interactions are much weaker for the Hf and Zr compounds compared to Ti,<sup>72</sup> leading to  $t_{2g}$  states that have little F character.

The spectra remain consistent for both structures containing each anion, indicating that the electronic interactions are inherent to each  $\text{MF}_6^{2-}$  BBU. These spectra indicate qualitative differences in the electronic structure of the  $\text{TiF}_6^{2-}$  anion compared to the  $\text{ZrF}_6^{2-}$  and  $\text{HfF}_6^{2-}$  anions, with  $\text{TiF}_6^{2-}$  featuring strong  $\sigma$  and  $\pi$  bonding while  $\text{ZrF}_6^{2-}$  and  $\text{HfF}_6^{2-}$  have strong  $\sigma$  bonding and weak  $\pi$  bonding.

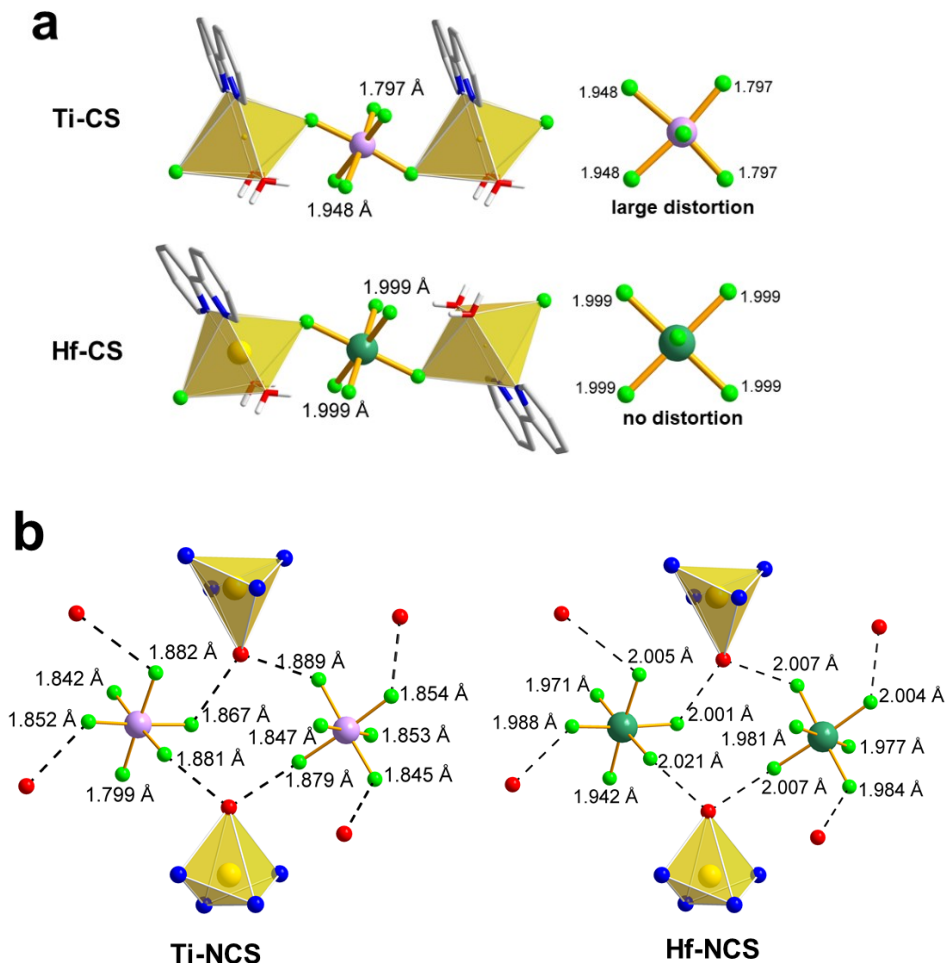
## **Discussion**

Visualization of each composition space aided by machine learning modeling captures the unique character of the  $\text{TiF}_6^{2-}$  anion. The decision tree classification and predictive machine learning models capture statistical differences in the three composition spaces examined in this study that point to chemical differences between reactions containing Ti and those containing Zr or Hf.

XAS measurements provide experimental evidence that the  $\text{MF}_6^{2-}$  ( $\text{M} = \text{Ti}, \text{Zr}, \text{Hf}$ ) anions are distinguished by the strength of ligand-to-metal  $\pi$  bonding. Pre-edge features in the fluorine K-edge XAS spectra demonstrate that the  $\text{TiF}_6^{2-}$  anion possesses strong  $\pi$  bonding interactions that are weak in the  $\text{ZrF}_6^{2-}$  or  $\text{HfF}_6^{2-}$  anions. The strong ligand-to-metal  $\pi$  bonding gives the  $\text{TiF}_6^{2-}$  anion second-order Jahn–Teller (SOJT) character, as the energy gap between filled  $\text{F}(2p)$  orbitals and empty  $\text{Ti}(3d - t_{2g})$  orbitals is small enough to allow for mixing to occur between these states.<sup>9,73</sup> Conversely, weak  $\pi$  bonding in the  $\text{ZrF}_6^{2-}$  and  $\text{HfF}_6^{2-}$  anions reflects the fact that these anions lack

SOJT activity because the energy gap between filled  $F(2p)$  and unoccupied  $Zr(4d - t_{2g})/Hf(5d - t_{2g})$  states is too large.

The SOJT character of the  $TiF_6^{2-}$  anion allows for out-of-center distortions to occur in a suitable structure.<sup>73,74</sup> Therefore, interactions with the extended structure must be considered to



**Figure 10: (a)** Configurations of octahedral  $MF_6^{2-}$  units in Ti-CS and Hf-CS. The large structural distortion of  $TiF_6^{2-}$  in Ti-CS reflects the presence of SOJT effects in the  $TiF_6^{2-}$  anion, while the lack of a distortion in Hf-CS is consistent with a lack of SOJT effects in the  $HfF_6^{2-}$  anion. **(b)** Smaller octahedral distortions are present in both Ti-NCS and Hf-NCS as a result of hydrogen bonding interactions (shown as dashed lines). Bond lengths of M–F bonds are labeled. Bipyridine ligands and hydrogen atoms have been omitted for clarity.

understand the presence or absence of distortions in the  $\text{TiF}_6^{2-}$  anion. Previous studies have shown that  $\text{TiF}_6^{2-}$  anions may experience distortions of the central Ti atom toward an edge ( $\text{C}_2$ -type), face ( $\text{C}_3$ -type), or vertex ( $\text{C}_4$ -type), or remain undistorted, depending on the nature of hydrogen bonding interactions and other contacts with the extended structure.<sup>75</sup> To understand the interplay of structural and electronic factors in the Ti-CS and Ti-NCS compounds, we will compare the structures of Ti-CS and Ti-NCS with Hf-CS and Hf-NCS. We limit the following discussion to Ti and Hf because the 7-coordinate environment of Zr-CS complicates direct comparison with the other M-CS structures. In the structure of Zr-CS, the edge-sharing nature of the  $\text{Zr}_2\text{F}_{12}^{4-}$  anion gives rise to distortions from cation-cation repulsion that are not present for Ti-CS or Hf-CS.

The structure of Ti-CS features a  $\text{C}_2$ -type distortion in the  $\text{TiF}_6^{2-}$  anion, which is supported by the asymmetric distribution of hydrogen bonding contacts on one side of the octahedron (Figure 10a). These contacts mitigate the large residual charge on the two fluoride ligands away from which the central Ti atom is displaced, making the Ti-CS structure suitable for SOJT activity in the  $\text{TiF}_6^{2-}$  anion. Hf-CS and Ti-CS reflect the structural differences that arise from the SOJT activity of the  $\text{TiF}_6^{2-}$  anion. Hf-CS features the same connectivity as Ti-CS yet the environment of the  $\text{HfF}_6^{2-}$  anion is undistorted, as the  $\text{Hf}(5d)\text{--F}(2p)$  energy gap is too large to allow for SOJT activity (Figure 10b).

The structures of Ti-NCS and Hf-NCS, on the other hand, feature small distortions in the  $\text{MF}_6^{2-}$  anions that are comparable in magnitude for  $\text{TiF}_6^{2-}$  and  $\text{HfF}_6^{2-}$  (Figure 10b). In both cases, the central M atom is displaced less than 0.006 Å from the center of the octahedron (Table S1). As the distortions are comparable for SOJT-active  $\text{TiF}_6^{2-}$  and SOJT-inactive  $\text{HfF}_6^{2-}$ , the distortions in the M-NCS structures result from hydrogen bonding interactions rather than SOJT activity.<sup>29</sup> The

small distortion in Ti-NCS indicates that the arrangement of hydrogen bonding interactions in this structure does not support the SOJT activity of the  $\text{TiF}_6^{2-}$  anion.

Our machine-learning-assisted composition space investigation revealed that behavior in each  $(\text{CuO}, \text{MO}_2)/\text{bpy}/\text{HF}(\text{aq})$  ( $\text{M} = \text{Ti}, \text{Zr}, \text{Hf}$ ) system is strongly dependent on the choice of ETM. Analysis of each composition space shows that phase selection in the three  $(\text{CuO}, \text{MO}_2)/\text{bpy}/\text{HF}(\text{aq})$  ( $\text{M} = \text{Ti}, \text{Zr}, \text{Hf}$ ) systems is driven by the  $\text{bpy}:\text{Cu}$  ratio, with lower values of  $\text{bpy}:\text{Cu}$  favoring the formation of  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2^{2+}$  cations in M-CS compounds and higher values of  $\text{bpy}:\text{Cu}$  favoring the formation of  $\Delta$ - and  $\Lambda$ - $\text{Cu}(\text{bpy})_2(\text{H}_2\text{O})_2^{2+}$  cations in M-NCS compounds. Phase selection for  $\text{M} = \text{Zr}$  and  $\text{Hf}$  can be achieved by modifying reaction stoichiometry to reflect the ratio of the desired phase, as Zr-NCS and Hf-NCS feature a  $\text{bpy}:\text{Cu}$  ratio of 2 while Zr-CS and Hf-CS have a  $\text{bpy}:\text{Cu}$  ratio of 1. The presence of the  $\text{TiF}_6^{2-}$  anion shifts the speciation equilibrium toward the formation of  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2^{2+}$  species and away from  $\Delta$ - and  $\Lambda$ - $\text{Cu}(\text{bpy})_2(\text{H}_2\text{O})_2^{2+}$  complexes, as demonstrated by the formation of Ti-NCS only when  $\text{bpy}:\text{Cu}$  greater than 1.5 and the amount of HF is less than 0.0025 moles. The shift in speciation equilibrium indicates an increase in the stability of Ti-CS relative to Zr-CS and Hf-CS. Ti-CS is stabilized by two factors: (1) increased covalency via the out-of-center SOJT distortion and (2) hydrogen bonding interactions that compensate for the decrease in covalency in the two long Ti–F bonds. The stability of the Ti-CS structure shifts speciation to favor  $\text{Cu}(\text{bpy})(\text{H}_2\text{O})_2^{2+}$  cations and form the one-dimensional polar chains that allow for the SOJT distortion to occur, rather than  $\Delta$ - and  $\Lambda$ - $\text{Cu}(\text{bpy})_2(\text{H}_2\text{O})_2^{2+}$  cations that form the Ti-NCS structure and do not allow for a SOJT distortion. These factors make Ti-CS more competitive with Ti-NCS during crystallization than the Zr-CS and Hf-CS counterparts.

Using ML on the experimental data provided many advantages. First, the data visualization efforts resulted in additional experiments, during which Ti-NCS was first synthesized. Additionally, these data visualizations enabled the identification of outlier points, which were addressed using data cleaning techniques. Working with the experimental data also provided insight into data input errors which has improved reproducibility and replicability of the experiments reported. Second, modeling the experiments provided unbiased statistical support to human intuition in the form of visualizations, an interpretable decision tree, and predictive LOO models that emphasize distinct properties of Ti compared to Hf and Zr. The machine learning models presented here provide statistical significance to the qualitative data inferences made by human inspection and can quantify traditionally qualitative arguments. We emphasize that these models are not intended to be generalizable to M-CS and M-NCS prediction for a broader range of early transition metals. Instead, the models were used to demonstrate how interpretable ML can capture the salient phase boundary information in small experimental datasets. These results support the distinction between composition spaces outlined for Ti and those for Zr and Hf, a point which is further corroborated from data visualization and decision tree analysis.

## Conclusion

A strategy combining high-throughput hydrothermal synthesis with machine learning was employed to synthesize two new polar compounds, Ti-NCS and Zr-NCS, based on racemic combinations of  $\Delta$ - and  $\Lambda$ - $\text{Cu}(\text{bpy})_2(\text{H}_2\text{O})^{2+}$  cations, as well as two related centrosymmetric compounds, Zr-CS and Hf-CS. Machine-learning analysis of reaction outcomes shows that phase selection between M-NCS and M-CS phases in each  $(\text{CuO}, \text{MO}_2)/\text{bpy}/\text{HF}(\text{aq})$  ( $\text{M} = \text{Ti}, \text{Zr}, \text{Hf}$ ) composition space is primarily driven by the bpy:Cu molar ratio for reactions containing Zr and

Hf, while phase selection between Ti-NCS and Ti-CS shows an additional dependence on the amount of HF present in the reaction. Fluorine K-edge XAS reveals that strong ligand-to-metal  $\pi$  bonding between  $\text{Ti}(3d - t_{2g})$  and  $\text{F}(2p)$  states distinguishes  $\text{TiF}_6^{2-}$  from  $\text{ZrF}_6^{2-}$  and  $\text{HfF}_6^{2-}$ . Crystallographic data and  $^{19}\text{F}$  NMR spectra are consistent with this distinction. Furthermore, the solid-state NMR data indicate rapid motion in the isolated  $\text{MF}_6^{2-}$  octahedra for all three compounds in the M-NCS series. The present study illustrates the value of combining synthetic methods with data visualization, which improves reproducibility and replicability of the experiments reported, and machine learning techniques that provide unbiased insight from ensembles of experiments. With these findings we hope to stimulate interest in racemic compounds with important properties that are dependent on the lack of crystallographic inversion symmetry.

## **Acknowledgements:**

The authors would like to thank Professor Andrew Maverick for valuable discussions regarding speciation of copper-bipyridine complexes. This work was supported by funding from the National Science Foundation (DMR-1904701). Single-crystal and powder X-ray diffraction data and solid-state NMR spectra were acquired at IMSERC at Northwestern University, which has received support from the Soft and Hybrid Nanotechnology Experimental (SHyNE) Resource (NSF ECCS-1542205), the State of Illinois, the International Institute for Nanotechnology (IIN), and the National Science Foundation (DMR-0521267). This work made use of the J.B. Cohen X-ray Diffraction Facility supported by the MRSEC program of the National Science Foundation (DMR-1720139). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Computational

resources were provided by the MERCURY consortium (<http://mercuryconsortium.org/>) under NSF grants CHE-1229354 and CHE-1662030. Contributions from A.J.N., I.M.P., and J.S. are supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR001118C0036. J.S. acknowledges the Henry Dreyfus Teacher-Scholar Award (TH-14-010). K.J.G., G.M.N., and J.C. were supported as part of the Joint Center for Energy Storage Research (JCESR), an Energy Innovation Hub funded by the U.S. Department of Energy (DOE), Office of Science, Basic Energy Sciences (BES).

#### **Author Information:**

Corresponding Author:

\* Email: [krp@northwestern.edu](mailto:krp@northwestern.edu)

#### **Supporting Information:**

Table of crystallographic data. Solid-state NMR spectra for M-CS compounds. Table of out-of-center distortions. Description of methods used in construction of machine learning models.

ML Files - SI: folder containing the files used to construct machine learning models

PXRD Data: powder X-ray diffraction data used to determine the outcome of each reaction

CCDC 1972390-1972395 contain the supplementary crystallographic data for this paper. The data can be obtained free of charge from The Cambridge Crystallographic Data Centre via [www.ccdc.cam.ac.uk/structures](http://www.ccdc.cam.ac.uk/structures).

#### **References:**

- (1) Halasyamani, P. S.; Poeppelmeier, K. R. Noncentrosymmetric Oxides. *Chem. Mater.* **1998**, *10* (10), 2753–2769. <https://doi.org/10.1021/cm980140w>.
- (2) Nye, J. F. *Physical Properties of Crystals: Their Representation by Tensors and Matrices*; Clarendon Press - Oxford, 1985.



- (3) Jacco, J. C.  $\text{KTiOPO}_4$  (KTP) Past, Present, And Future. In *Ceramics and Inorganic Crystals for Optics, Electro-Optics, and Nonlinear Conversion*; International Society for Optics and Photonics, 1989; Vol. 0968, pp 93–100. <https://doi.org/10.1117/12.948132>.
- (4) Stucky, G. D.; Phillips, M. L. F.; Gier, T. E. The Potassium Titanyl Phosphate Structure Field: A Model for New Nonlinear Optical Materials. *Chem. Mater.* **1989**, *1* (5), 492–509. <https://doi.org/10.1021/cm00005a008>.
- (5) von Hippel, A. Ferroelectricity, Domain Structure, and Phase Transitions of Barium Titanate. *Rev. Mod. Phys.* **1950**, *22* (3), 221–237. <https://doi.org/10.1103/RevModPhys.22.221>.
- (6) Nassau, K.; Levinstein, H. J.; Loiacono, G. M. Ferroelectric Lithium Niobate. 1. Growth, Domain Structure, Dislocations and Etching. *J. Phys. Chem. Solids* **1966**, *27* (6), 983–988. [https://doi.org/10.1016/0022-3697\(66\)90070-9](https://doi.org/10.1016/0022-3697(66)90070-9).
- (7) Cohen, R. E. Origin of Ferroelectricity in Perovskite Oxides. *Nature* **1992**, *358* (6382), 136. <https://doi.org/10.1038/358136a0>.
- (8) *Physics of Ferroelectrics: A Modern Perspective*; Rabe, K. M., Ahn, C. H., Triscone, J.-M., Eds.; Topics in applied physics; Springer: Berlin ; New York, 2007.
- (9) Bersuker, I. B. Pseudo-Jahn–Teller Effect—A Two-State Paradigm in Formation, Deformation, and Transformation of Molecular Systems and Solids. *Chem. Rev.* **2013**, *113* (3), 1351–1390. <https://doi.org/10.1021/cr300279n>.
- (10) Pearson, R. G. The Second-Order Jahn-Teller Effect. *J. Mol. Struct. THEOCHEM* **1983**, *103* (Supplement C), 25–34. [https://doi.org/10.1016/0166-1280\(83\)85006-4](https://doi.org/10.1016/0166-1280(83)85006-4).
- (11) Opik, U.; Pryce, M. H. L. Studies of the Jahn-Teller Effect. I. A Survey of the Static Problem. *Proc. R. Soc. Math. Phys. Eng. Sci.* **1957**, *238* (1215), 425–447. <https://doi.org/10.1098/rspa.1957.0010>.
- (12) Wheeler, R. A.; Whangbo, M. Hwan.; Hughbanks, Timothy.; Hoffmann, Roald.; Burdett, J. K.; Albright, T. A. Symmetric vs. Asymmetric Linear M-X-M Linkages in Molecules, Polymers, and Extended Networks. *J. Am. Chem. Soc.* **1986**, *108* (9), 2222–2236. <https://doi.org/10.1021/ja00269a018>.
- (13) Muller, E. A.; Cannon, R. J.; Sarjeant, A. N.; Ok, K. M.; Halasyamani, P. S.; Norquist, A. J. Directed Synthesis of Noncentrosymmetric Molybdates. *Cryst. Growth Des.* **2005**, *5* (5), 1913–1917. <https://doi.org/10.1021/cg050184z>.
- (14) Veltman, T. R.; Stover, A. K.; Narducci Sarjeant, A.; Ok, K. M.; Halasyamani, P. S.; Norquist, A. J. Directed Synthesis of Noncentrosymmetric Molybdates Using Composition Space Analysis. *Inorg. Chem.* **2006**, *45* (14), 5529–5537. <https://doi.org/10.1021/ic060558t>.
- (15) Gutnick, J. R.; Muller, E. A.; Narducci Sarjeant, A.; Norquist, A. J.  $[\text{C}_5\text{H}_{14}\text{N}_2][(\text{MoO}_3)_3(\text{SO}_4)] \cdot \text{H}_2\text{O}$ : Sulfated  $\alpha$ -Molybdena Chains. *Inorg. Chem.* **2004**, *43* (21), 6528–6530. <https://doi.org/10.1021/ic049192f>.
- (16) Hubbard, D. J.; Johnston, A. R.; Casalongue, H. S.; Sarjeant, A. N.; Norquist, A. J. Synthetic Approaches for Noncentrosymmetric Molybdates. *Inorg. Chem.* **2008**, *47* (19), 8518–8525. <https://doi.org/10.1021/ic800572g>.
- (17) Glor, E. C.; Blau, S. M.; Yeon, J.; Zeller, M.; Shiv Halasyamani, P.; Schrier, J.; Norquist, A. J.  $[\text{R-C}_7\text{H}_{16}\text{N}_2][\text{V}_2\text{Te}_2\text{O}_{10}]$  and  $[\text{S-C}_7\text{H}_{16}\text{N}_2][\text{V}_2\text{Te}_2\text{O}_{10}]$ ; New Polar Templated Vanadium Tellurite Enantiomers. *J. Solid State Chem.* **2011**, *184* (6), 1445–1450. <https://doi.org/10.1016/j.jssc.2011.04.016>.

- (18) Masse, R.; Zyss, J. A New Approach in the Design of Polar Crystals for Quadratic Nonlinear Optics Exemplified by the Synthesis and Crystal Structure of 2-Amino-5-Nitropyridinium Dihydrogen Monophosphate (2A5NPDP). *Mol. Eng.* **1991**, *1* (2), 141–152. <https://doi.org/10.1007/BF00420050>.
- (19) Guerin, S.; O'Donnell, J.; Haq, E. U.; McKeown, C.; Silien, C.; Rhen, F. M. F.; Soulimane, T.; Tofail, S. A. M.; Thompson, D. Racemic Amino Acid Piezoelectric Transducer. *Phys. Rev. Lett.* **2019**, *122* (4), 047701. <https://doi.org/10.1103/PhysRevLett.122.047701>.
- (20) Martin Britto Dhas, S. A.; Natarajan, S. Growth and Characterization of Dl-Alanine — A New NLO Material from the Amino Acid Family. *Mater. Lett.* **2008**, *62* (17), 2633–2636. <https://doi.org/10.1016/j.matlet.2008.01.034>.
- (21) Levy, H. A.; Corey, R. B. The Crystal Structure of Dl-Alanine. *J. Am. Chem. Soc.* **1941**, *63* (8), 2095–2108. <https://doi.org/10.1021/ja01853a020>.
- (22) Dalhus, B.; Görbitz, C. H. Non-Centrosymmetric Racemates: Space-Group Frequencies and Conformational Similarities between Crystallographically Independent Molecules. *Acta Crystallogr. B* **2000**, *56* (4), 715–719. <https://doi.org/10.1107/S0108768100002172>.
- (23) Flack, H. D. Chiral and Achiral Crystal Structures. *Helv. Chim. Acta* **2003**, *86* (4), 905–921. <https://doi.org/10.1002/hlca.200390109>.
- (24) Rieckhoff, K. E.; Peticolas, W. L. Optical Second-Harmonic Generation in Crystalline Amino Acids. *Science* **1965**, *147* (3658), 610–611. <https://doi.org/10.1126/science.147.3658.610>.
- (25) Lo Presti, L.; Sist, M.; Loconte, L.; Pinto, A.; Tamborini, L.; Gatti, C. Rationalizing the Lacking of Inversion Symmetry in a Noncentrosymmetric Polar Racemate: An Experimental and Theoretical Study. *Cryst. Growth Des.* **2014**, *14* (11), 5822–5833. <https://doi.org/10.1021/cg501074x>.
- (26) Carvalho, P. S.; Ellena, J.; Yufit, D. S.; Howard, J. A. K. Rare Case of Polymorphism in a Racemic Fluoxetine Nitrate Salt: Phase Behavior and Relative Stability. *Cryst. Growth Des.* **2016**, *16* (7), 3875–3883. <https://doi.org/10.1021/acs.cgd.6b00442>.
- (27) Ravikumar, B.; Sridhar, B.; Rajaram, R. K. Dl-Phenyl-alaninium Di-hydrogen Phosphate. *Acta Crystallogr. Sect. E Struct. Rep. Online* **2001**, *57* (11), o1078–o1080. <https://doi.org/10.1107/S1600536801016993>.
- (28) Kumar, S.; Subramanian, K.; Schreurs, A. M. M.; Kroon, J.; Steiner, T. Racemic  $\gamma$ -2-Cis-6-Diphenyl-Trans-3-Ethylpiperidin-4-One Oxime, a Crystal Structure with  $Z' = 8$ . *J. Chem. Crystallogr.* **2001**, *31* (9), 411–416. <https://doi.org/10.1023/A:1015675418257>.
- (29) Gautier, R.; Norquist, A. J.; Poeppelmeier, K. R. From Racemic Units to Polar Materials. *Cryst. Growth Des.* **2012**, *12* (12), 6267–6271. <https://doi.org/10.1021/cg301539q>.
- (30) Gautier, R.; Klingsporn, J. M.; Van Duyne, R. P.; Poeppelmeier, K. R. Optical Activity from Racemates. *Nat. Mater.* **2016**, *15* (6), 591–592. <https://doi.org/10.1038/nmat4628>.
- (31) Kahr, B.; Martin, A. T.; Ernst, K.-H. On the Chiroptical Properties of Racemic Crystals. *Chirality* **2018**. <https://doi.org/10.1002/chir.22820>.
- (32) Halasyamani, P.; Willis, M. J.; Stern, C. L.; Lundquist, P. M.; Wong, G. K.; Poeppelmeier, K. R. Composition Space of the  $(\text{CuO}, 1/2\text{Nb}_2\text{O}_5)/(\text{HF})_x \cdot \text{Pyridine}/\text{H}_2\text{O}$  System. Structure and Synthesis of  $[\text{PyH}^+]_2[\text{CuNb}_2(\text{Py})_4\text{O}_2\text{F}_{10}]^{2-}$  and  $\text{CuNb}(\text{Py})_4\text{OF}_5$ . *Inorg. Chem.* **1996**, *35* (5), 1367–1371.

- (33) Norquist, A. J.; Heier, K. R.; Stern, C. L.; Poeppelmeier, K. R. Composition Space Diagrams for Mixed Transition Metal Oxide Fluorides. *Inorg. Chem.* **1998**, *37* (25), 6495–6501. <https://doi.org/10.1021/ic9807238>.
- (34) Mueller, T.; Kusne, A. G.; Ramprasad, R. Machine Learning in Materials Science: Recent Progress and Emerging Applications. In *Reviews in Computational Chemistry*; Parrill, A. L., Lipkowitz, K. B., Eds.; John Wiley & Sons, Inc: Hoboken, NJ, 2016; pp 186–273. <https://doi.org/10.1002/9781119148739.ch4>.
- (35) Abu-Mostafa, Y. S.; Magdon-Ismail, M.; Lin, H.-T. *Learning from Data: A Short Course*; AMLbook.com: S.I., 2012.
- (36) Raccuglia, P.; Elbert, K. C.; Adler, P. D. F.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J. Machine-Learning-Assisted Materials Discovery Using Failed Experiments. *Nature* **2016**, *533* (7601), 73–76. <https://doi.org/10.1038/nature17439>.
- (37) Vinogradov, A. S.; Fedoseenko, S. I.; Krasnikov, S. A.; Preobrajenski, A. B.; Sivkov, V. N.; Vyalikh, D. V.; Molodtsov, S. L.; Adamchuk, V. K.; Laubschat, C.; Kaindl, G. Low-Lying Unoccupied Electronic States in 3d Transition-Metal Fluorides Probed by NEXAFS at the F 1s Threshold. *Phys. Rev. B* **2005**, *71* (4). <https://doi.org/10.1103/PhysRevB.71.045127>.
- (38) de Groot, F.; Vankó, G.; Glatzel, P. The 1s x-Ray Absorption Pre-Edge Structures in Transition Metal Oxides. *J. Phys. Condens. Matter* **2009**, *21* (10), 104207. <https://doi.org/10.1088/0953-8984/21/10/104207>.
- (39) Harrison, W. T. A.; Nenoff, T. M.; Gier, T. E.; Stucky, G. D. Tetrahedral-Atom 3-Ring Groupings in 1-Dimensional Inorganic Chains: Beryllium Arsenate Hydroxide Hydrate ( $\text{Be}_2\text{AsO}_4\text{OH} \cdot 4\text{H}_2\text{O}$ ) and Sodium Zinc Hydroxide Phosphate Hydrate ( $\text{Na}_2\text{ZnPO}_4\text{OH} \cdot \text{H}_2\text{O}$ ). *Inorg. Chem.* **1993**, *32* (11), 2437–2441. <https://doi.org/10.1021/ic00063a039>.
- (40) Sheldrick, G. M. Crystal Structure Refinement with SHELXL. *Acta Crystallogr. Sect. C Struct. Chem.* **2015**, *71* (1), 3–8. <https://doi.org/10.1107/S2053229614024218>.
- (41) *SAINT V8.38A*; Bruker Analytical X-ray Instruments: Madison, WI, USA, 2016.
- (42) Sheldrick, G. M. *SADABS*; University of Göttingen: Göttingen, Germany, 2002.
- (43) Spek, A. L. Single-Crystal Structure Validation with the Program PLATON. *J. Appl. Crystallogr.* **2003**, *36* (1), 7–13. <https://doi.org/10.1107/S0021889802022112>.
- (44) Dolomanov, O. V.; Bourhis, L. J.; Gildea, R. J.; Howard, J. a. K.; Puschmann, H. OLEX2: A Complete Structure Solution, Refinement and Analysis Program. *J. Appl. Crystallogr.* **2009**, *42* (2), 339–341. <https://doi.org/10.1107/S0021889808042726>.
- (45) Pendleton, I. M.; Cattabriga, G.; Li, Z.; Najeeb, M. A.; Friedler, S. A.; Norquist, A. J.; Chan, E. M.; Schrier, J. Experiment Specification, Capture and Laboratory Automation Technology (ESCALATE): A Software Pipeline for Automated Chemical Experimentation and Data Management. *MRS Commun.* **2019**, 1–14. <https://doi.org/10.1557/mrc.2019.72>.
- (46) Frisch, M. J.; Trucks, G. W.; Cheeseman, J. R.; Scalmani, G.; Caricato, M.; Hratchian, H. P.; Li, X.; Barone, V.; Bloino, J.; Zheng, G.; Vreven, T.; Montgomery, J. A.; Petersson, G. A.; Scuseria, G. E.; Schlegel, H. B.; Nakatsuji, H.; Izmaylov, A. F.; Martin, R. L.; Sonnenberg, J. L.; Peralta, J. E.; Heyd, J. J.; Brothers, E.; Ogliaro, F.; Bearpark, M.; Robb, M. A.; Mennucci, B.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Rendell, A.; Gomperts, R.; Zakrzewski, V. G.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H. *Gaussian 09*.

- (47) Bühl, M.; Kabrede, H. Geometries of Transition-Metal Complexes from Density-Functional Theory. *J. Chem. Theory Comput.* **2006**, *2* (5), 1282–1290. <https://doi.org/10.1021/ct6001187>.
- (48) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J. Chem. Theory Comput.* **2012**, *8* (2), 527–541. <https://doi.org/10.1021/ct200866d>.
- (49) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (50) Bielecki, A.; Burum, D. P. Temperature Dependence of  $^{207}\text{Pb}$  MAS Spectra of Solid Lead Nitrate. An Accurate, Sensitive Thermometer for Variable-Temperature MAS. *J. Magn. Reson. A* **1995**, *116* (2), 215–220. <https://doi.org/10.1006/jmra.1995.0010>.
- (51) Beckmann, P. A.; Dybowski, C. A Thermometer for Nonspinning Solid-State NMR Spectroscopy. *J. Magn. Reson.* **2000**, *146* (2), 379–380. <https://doi.org/10.1006/jmre.2000.2171>.
- (52) Gautier, R.; Poeppelmeier, K. R. Alignment of Acentric Units in Infinite Chains: A “Lock and Key” Model. *Cryst. Growth Des.* **2013**, *13* (9), 4084–4091. <https://doi.org/10.1021/cg400908f>.
- (53) Gautier, R.; Oka, K.; Kihara, T.; Kumar, N.; Sundaresan, A.; Tokunaga, M.; Azuma, M.; Poeppelmeier, K. R. Spin Frustration from *Cis* -Edge or -Corner Sharing Metal-Centered Octahedra. *J. Am. Chem. Soc.* **2013**, *135* (51), 19268–19274. <https://doi.org/10.1021/ja409831h>.
- (54) Legein, C.; Fayon, F.; Martineau, C.; Body, M.; Buzaré, J.-Y.; Massiot, D.; Durand, E.; Tressaud, A.; Demourgues, A.; Péron, O.; Boulard, B.  $^{19}\text{F}$  High Magnetic Field NMR Study of  $\beta\text{-ZrF}_4$  and  $\text{CeF}_4$ : From Spectra Reconstruction to Correlation between Fluorine Sites and  $^{19}\text{F}$  Isotropic Chemical Shifts. *Inorg. Chem.* **2006**, *45* (26), 10636–10641. <https://doi.org/10.1021/ic061339a>.
- (55) Biswal, M.; Body, M.; Legein, C.; Sadoc, A.; Boucher, F.  $\text{NbF}_5$  and  $\text{TaF}_5$ : Assignment of  $^{19}\text{F}$  NMR Resonances and Chemical Bond Analysis from GIPAW Calculations. *J. Solid State Chem.* **2013**, *207*, 208–217. <https://doi.org/10.1016/j.jssc.2013.09.001>.
- (56) Murakami, M.; Noda, Y.; Takegoshi, K. Terminal and Bridging Fluorine Ligands in  $\text{TiF}_4$  as Studied by  $^{19}\text{F}$  NMR in Solids. *Solid State Nucl. Magn. Reson.* **2019**, *101*, 82–88. <https://doi.org/10.1016/j.ssnmr.2019.05.007>.
- (57) Ramsey, N. F. Magnetic Shielding of Nuclei in Molecules. *Physica* **1951**, *17* (3), 303–307. [https://doi.org/10.1016/0031-8914\(51\)90070-5](https://doi.org/10.1016/0031-8914(51)90070-5).
- (58) Saika, A.; Slichter, C. P. A Note on the Fluorine Resonance Shifts. *J. Chem. Phys.* **1954**, *22* (1), 26–28. <https://doi.org/10.1063/1.1739849>.
- (59) Pell, A. J.; Pintacuda, G.; Grey, C. P. Paramagnetic NMR in Solution and the Solid State. *Prog. Nucl. Magn. Reson. Spectrosc.* **2019**, *111*, 1–271. <https://doi.org/10.1016/j.pnmrs.2018.05.001>.
- (60) Garribba, E.; Micera, G.; Sanna, D.; Strinna-Erre, L. The  $\text{Cu(II)}$ -2,2'-Bipyridine System Revisited. *Inorganica Chim. Acta* **2000**, *299* (2), 253–261. [https://doi.org/10.1016/S0020-1693\(99\)00508-3](https://doi.org/10.1016/S0020-1693(99)00508-3).

- (61) Nisbet, M. L.; Pendleton, I. M.; Nolis, G. M.; Griffith, K. J.; Schrier, J.; Cabana, J.; Norquist, A. J.; Poeppelmeier, K. R. Machine-Learning-Assisted Synthesis of Polar Racemates. *Materials Data Facility* 2020. <https://doi.org/10.18126/VNHJ-JAZ5>.
- (62) Blaiszik, B.; Chard, K.; Pruyne, J.; Ananthakrishnan, R.; Tuecke, S.; Foster, I. The Materials Data Facility: Data Services to Advance Materials Science Research. *JOM* **2016**, 68 (8), 2045–2052. <https://doi.org/10.1007/s11837-016-2001-3>.
- (63) Blaiszik, B.; Ward, L.; Schwarting, M.; Gaff, J.; Chard, R.; Pike, D.; Chard, K.; Foster, I. A Data Ecosystem to Support Machine Learning in Materials Science. *MRS Commun.* **2019**, 9 (4), 1125–1133. <https://doi.org/10.1557/mrc.2019.118>.
- (64) de Groot, F. M. F.; Grioni, M.; Fuggle, J. C.; Ghijsen, J.; Sawatzky, G. A.; Petersen, H. Oxygen 1s x-Ray-Absorption Edges of Transition-Metal Oxides. *Phys. Rev. B* **1989**, 40 (8), 5715–5723. <https://doi.org/10.1103/PhysRevB.40.5715>.
- (65) Solomon, E. I.; Hedman, B.; Hodgson, K. O.; Dey, A.; Szilagy, R. K. Ligand K-Edge X-Ray Absorption Spectroscopy: Covalency of Ligand–Metal Bonds. *Coord. Chem. Rev.* **2005**, 33.
- (66) Glaser, T.; Hedman, B.; Hodgson, K. O.; Solomon, E. I. Ligand K-Edge X-Ray Absorption Spectroscopy: A Direct Probe of Ligand–Metal Covalency. *Acc. Chem. Res.* **2000**, 33 (12), 859–868. <https://doi.org/10.1021/ar990125c>.
- (67) de Groot, F.; Kotani, A. *Core Level Spectroscopy of Solids*; Advances in Condensed Matter Science; CRC Press, 2008; Vol. 6. <https://doi.org/10.1201/9781420008425>.
- (68) Vinogradov, A. S.; Dukhanyakov, A. Yu.; Ipatov, V. M.; Onopko, D. E.; Pavlychev, A. A.; Titov, S. A. Fine Structure of X-Ray Absorption of the  $(\text{TiF}_6)^{2-}$  Radical. *Sov Phys Solid State* **1982**, 24 (5), 803.
- (69) Vinogradov, A. S.; Fedoseenko, S. I.; Vyalikh, D. V.; Molodtsov, S. L.; Adamchuk, V. K.; Laubschat, C.; Kaindl, G. High Resolution F1s Absorption Spectra of Solid Fluorides of 3d Elements. *Opt. Spectrosc.* **2002**, 93 (6), 862–869. <https://doi.org/10.1134/1.1531709>.
- (70) Vinogradov, A. S.; Fedoseenko, S. I.; Vyalikh, D.; Molodtsov, S. L.; Adamchuk, V. K.; Laubschat, C.; Kaindl, G. F 1s Absorption Study of Hybridization and Collapse of 3d Electronic States of First-Row Transition-Metal Fluorides. *BESSY Annu. Rep.* **2001**, 168.
- (71) Levenson, R. A.; Dominguez, R. J. G. Crystal Field Approach to Pentagonal-Bipyramidal Seven-Coordinate Complexes. Electronic Structure of Heptacyanovanadate(III). *Inorg. Chem.* **1973**, 12 (10), 2342–2345. <https://doi.org/10.1021/ic50128a025>.
- (72) Lucovsky, G.; Fulton, C. C.; Zhang, Y.; Zou, Y.; Luning, J.; Edge, L. F.; Whitten, J. L.; Nemanich, R. J.; Ade, H.; Schlom, D. G.; Afanasev, V. V.; Stesmans, A.; Zollner, S.; Triyoso, D.; Rogers, B. R. Conduction Band-Edge States Associated with the Removal of d-State Degeneracies by the Jahn-Teller Effect. *IEEE Trans. Device Mater. Reliab.* **2005**, 5 (1), 65–83. <https://doi.org/10.1109/TDMR.2005.845804>.
- (73) Kunz, M.; Brown, I. D. Out-of-Center Distortions around Octahedrally Coordinated  $d^0$  Transition Metals. *J. Solid State Chem.* **1995**, 115 (2), 395–406. <https://doi.org/10.1006/jssc.1995.1150>.
- (74) Welk, M. E.; Norquist, A. J.; Arnold, F. P.; Stern, C. L.; Poeppelmeier, K. R. Out-of-Center Distortions in  $d^0$  Transition Metal Oxide Fluoride Anions. *Inorg. Chem.* **2002**, 41 (20), 5119–5125. <https://doi.org/10.1021/ic025622v>.
- (75) Kim, E.; Lee, D. W.; Ok, K. M. Centrosymmetric  $[\text{N}(\text{CH}_3)_4]_2\text{TiF}_6$  vs. Noncentrosymmetric Polar  $[\text{C}(\text{NH}_2)_3]_2\text{TiF}_6$ : A Hydrogen-Bonding Effect on the out-of-

Center Distortion of  $\text{TiF}_6$  Octahedra. *J. Solid State Chem.* **2012**, 195, 149–154.  
<https://doi.org/10.1016/j.jssc.2011.11.027>.

