THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

Cite This: J. Phys. Chem. Lett. 2018, 9, 1064−1071

pubs.acs.org/JPCL

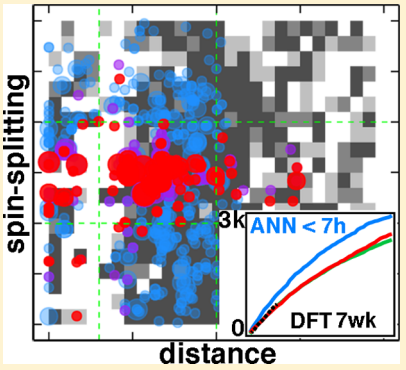
Accelerating Chemical Discovery with Machine Learning: Simulated Evolution of Spin Crossover Complexes with an Artificial Neural Network

Jon Paul Janet, Lydia Chan, and Heather J. Kulik*

Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

Supporting Information

ABSTRACT: Machine learning (ML) has emerged as a powerful complement to simulation for materials discovery by reducing time for evaluation of energies and properties at accuracy competitive with first-principles methods. We use genetic algorithm (GA) optimization to discover unconventional spin-crossover complexes in combination with efficient scoring from an artificial neural network (ANN) that predicts spin-state splitting of inorganic complexes. We explore a compound space of over 5600 candidate materials derived from eight metal/oxidation state combinations and a 32-ligand pool. We introduce a strategy for error-aware ML-driven discovery by limiting how far the GA travels away from the nearest ANN training points while maximizing property (i.e., spin-splitting) fitness, leading to discovery of 80% of the leads from full chemical space enumeration. Over a 51-complex subset, average unsigned errors (4.5 kcal/mol) are close to the ANN's baseline 3 kcal/mol error. By obtaining leads from the trained ANN within seconds rather than days from a DFT-driven GA, this strategy demonstrates the power of ML for accelerating inorganic material discovery.



Although increases in computing power and efficient algorithms[1−6] have cemented first-principles screening[7−16] (e.g., with density functional theory or DFT) as a critical component of materials and chemical discovery,[9−13,17−22] further acceleration is needed to overcome the combinatorial challenge of vast regions of unexplored chemical space.[23,24] With the increased availability of large training sets,[25] machine learning (ML) has emerged[26−30] as a tool to replace first-principles characterization, demonstrating improvement over linear structure−property relationships[31,32] and, where large data sets are available, predicting energies with an accuracy that approaches or exceeds the baseline accuracy of approximate DFT.[29,33,34] ML models have excelled in design for narrow composition spaces (e.g., alloys[31,32] or phase stabilities[35]). Descriptors used in ML model training can have strong size and domain dependence[36−39] that restrict discovery to a specific size range and chemical composition. Following the successes of force-field development,[40] group additivity,[41] and chem-informatics,[42,43] major ML-driven advances have been made in organic molecule design and discovery,[44,45] where structure−property relationships are well-defined. Inorganic chemistry represents a challenging case where few[46] force fields are available, informatics approaches are less well-developed,[14,16,47−50] and properties of interest such as spin-state ordering or redox potential require robust first-principles characterization.

Nevertheless, the enlarged chemical space afforded by inorganic chemistry motivates ML model development as a tool to accelerate discovery. We recently trained[51] an artificial neural network (ANN) on 2690 geometry optimized transition-metal complexes to predict transition-metal complex adiabatic high-spin to low-spin state splitting ($\Delta E_{H\text{-}L}$) with root-mean-square error (RMSE) of 3 kcal/mol along with its Hartree−Fock exchange sensitivity and metal−ligand bond length. We selected 25 mixed continuous (i.e., oxidation state) and discrete (i.e., metal identity) local descriptors (MCDL-25) that focused on metal-proximal effects and demonstrated superior transferability over whole complex representations to the prediction on diverse molecules from experimental databases (Figure 1). This feature set was selected from seven candidate feature sets, as evaluated by retained features and errors with LASSO,[52] and the inclusion of discrete features was made possible by their compatibility with an ANN. As suggested by the success of ligand field theory,[53−55] our representation[51,56] is ideal for predicting inherently local, electronic properties such as spin state splitting.

Now we turn to the outstanding challenge of using ML models to enable chemical discovery in inorganic chemistry. An open question for the use of ML models[28,32,37,39] in discovery[57−60] is the manner in which we should optimally balance exploration of new compounds with model confidence. Although ML model predictions are of virtually no computational cost versus direct simulation, if the ML model lacks extrapolative power to previously unstudied complexes, then its

Received: January 17, 2018
Accepted: February 9, 2018
Published: February 9, 2018

DOI: 10.1021/acs.jpclett.8b00170
J. Phys. Chem. Lett. 2018, 9, 1064−1071

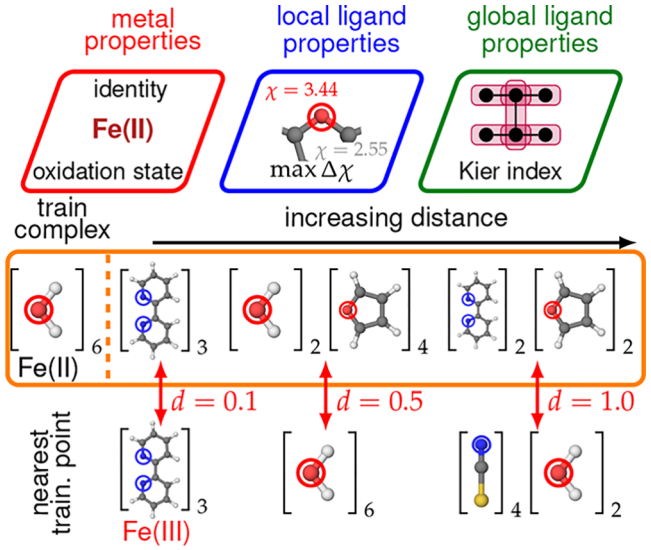


Figure 1. (top) Representative descriptors in MCDL-25: metal properties, metal-adjacent (i.e., local ligand properties), and global ligand properties. (bottom) Representative complexes including $Fe(II)(H_2O)_6$ in training data and increasingly distant complexes from the training data (left to right): $Fe(II)(bpy)_3$, $Fe(II)(H_2O)_2(furan)_4$, and $Fe(II)(bpy)_2(furan)_2$. The closest training point and its distance is indicated below each complex.

utility for chemical space exploration will be limited. A second concern is the manner in which optimization is carried out in continuous, data-driven representations versus discrete representations needed for characterization, for example, by simulation.

We develop strategies for balancing exploration in inorganic chemical space with model confidence in a manner that makes leads obtained from a ML model amenable to straightforward validation by first-principles simulation, a necessary step toward automated, adaptive learning. We circumvent the secondary question of continuous optimization[61−65] by employing widely used[66−69] genetic algorithms (GAs) on discrete ligand pools in combination with our ANN to discover unconventional spin-crossover complexes.

Spin-crossover complexes (SCOs) are defined by near-zero free-energy differences of high (H) and low (L) spin states ($\Delta G_{H\text{-}L}$), with changes in spin in response to light/heat due to entropic differences.[70] This behavior makes SCOs compelling for potential applications, for example, spintronics and sensing of light[71,72] or small molecules.[73−75] Conventional[76−79] Fe(II)/nitrogen SCOs are well-studied,[19,54,55] and design rules for these complexes have been recently suggested.[80,81] In this work, we use GAs on a larger (i.e., several thousand) compound space to reveal both expected and unconventional SCOs as judged through adiabatic electronic energy differences (i.e., $|\Delta E_{H\text{-}L}| < 5$ kcal/mol; see Computational Details).

We now will explore GA-driven strategies for discovering octahedral spin-crossover complexes from a chemical space composed of metals in the original ANN (i.e., M(II/III), where M = Cr, Mn, Fe, or Co) with 32 unique ligands with varied denticity (i.e., 16 monodentate, 14 bidentate, and 2 tetradentate), direct connecting atom identity (CA, i.e., C, N, and O), and size (i.e., from 2 atoms in CO to 52 atoms in cyanoaceticporphyrin) (Supporting Information Table S1 and Figure S1). Taking into account ligand compatibility and the symmetry required by the ANN[51] (i.e., one gene each for axial and equatorial ligand identity), these combinations produce a compound space of 5664 (i.e., 708 ligand combinations × 8 metals) transition-metal complexes ranging from 13 to 151 atoms in size (Supporting Information Table S2). Of the 32 ligands, 14 were in the original set of ligands used to train the ANN, but only 113 of 5664 compounds (2%) in the design space have been previously studied.[51]

For spin-crossover complex discovery, our target is to minimize the spin-state splitting (i.e., $\Delta E_{H\text{-}L}$) obtained (e.g., with DFT or an ANN), using H-L definitions as in previous work[51] (see Computational Details and Supporting Information Table S3). At each generation in the GA, compound spin-splitting fitness (F_s) is evaluated as

$$F_s = e^{-\left(\frac{\Delta E_{H\text{-}L}}{\Delta w_{H\text{-}L}}\right)^2} \quad (1)$$

where $\Delta w_{H\text{-}L}$ controls the decrease in fitness with increasing magnitude of $\Delta E_{H\text{-}L}$, chosen to be 15 kcal/mol to preserve $F_s \approx 1.0$ for $|\Delta E_{H\text{-}L}| \leq 5$ kcal/mol (Supporting Information Figure S2). For the GA, we follow a similar strategy to ref 22: Starting from a pool of 20 randomly selected complexes, 21 generations are carried out with fitness evaluation, which includes five crossovers and random mutation probability (p_{mut}) of 0.15 (i.e., of the metal or ligand genes, details in Supporting Information Text S1). Differences from standard choices[22,66] are the reduced number of generations and a higher mutation probability to increase diversity,[67] both motivated by the modest compound space. We introduce a diversity control mode to further increase diversity (i.e., percentage of complexes with distinct gene combinations in the total pool) by raising p_{mut} by 0.5 when the diversity of a generation falls below 25% of the pool and reduce p_{mut} to the 0.15 default once diversity reaches at least 25%.

Evaluation of compound fitness during GA optimization with a trained ANN motivates consideration of uncertainty in model predictions. Beyond sometimes overconfident credible intervals[82] obtained from dropout, we have identified[51] large (i.e., > 1.0) Euclidean norm of the distance (d) in normalized MCDL-25 descriptor space to training data to be a useful indicator of low ANN accuracy (Supporting Information Text S1). MCDL-25 emphasizes the direct metal−ligand environment: Preserving an oxygen connecting atom but replacing water ligands with larger furan ligands (i.e., changing the $\Delta\chi$ and topology from a truncated Kier index[83]) produces moderate distances (i.e., d = 0.5, see Figure 1), whereas a changed CA in otherwise comparably structured ligands (i.e., imidazole vs furan) produces large distances (i.e., d = 1.5). Differences in oxidation state (e.g., $Fe(II)(bpy)_3$ vs $Fe(III)(bpy)_3$: d = 0.1) are closer in descriptor space than different metals (e.g., $Fe(II)(furan)_6$ vs $Mn(II)(furan)_6$: d > 1.0) (see Figure 1). Large distances in descriptor space can arise from substantial differences in all ligands, even when metal, oxidation state, and direct CAs (i.e., all proximal features[56]) are unchanged. $Fe(II)(bpy)_2(furan)_2$ is distant (i.e., d = 1.0) from the closest ANN training point,[51] $Fe(II)(NCS)_4(H_2O)_2$, due to differing denticity, $\Delta\chi$, and truncated Kier index.[83] Remote changes more than three bonds away from the metal−ligand bond[56,84] are neglected in the nearsighted descriptor set, so distinct compounds can be identical in MCDL-25 (see Figure 1 and Supporting Information Table S1).

Thus using observations about the relationship between chemical differences and descriptor distances, we define our target discovery region for ANN scoring as $0.3 \leq d \leq 1.0$ to

DOI: 10.1021/acs.jpclett.8b00170
J. Phys. Chem. Lett. 2018, 9, 1064−1071

avoid both "discovery" of complexes too similar to training data and high-promise but very low-confidence complexes. We introduce a modified fitness function ($F_{s,d}$)

$$F_{s,d} = e^{-\left(\frac{\Delta E_{H\text{-}L}}{\Delta w_{H\text{-}L}}\right)^2} e^{-\left(\frac{d}{d_{opt}}\right)^2} \quad (2)$$

where in addition to a spin-splitting fitness term, a penalty scaled by d_{opt} is set to discourage sampling compounds with a very large distance to the training data. To encourage compound discovery, we introduce a distance control mode that adapts the fitness function from eq 1 to eq 2 only if the average d of all complexes is large (d_{av} > 0.6, selected by trial and error) after a generation has been selected for fitness and reverts to eq 1 if d_{av} falls below 0.6 in a subsequent generation. We have selected d_{opt} = 1.0 in eq 2 by trial and error to avoid overpenalizing discovery.

We compare four modes of spin crossover complex GA optimization using an ANN for scoring: (i) distance control in the fitness function, (ii) mutation-rate enhancement to encourage diversity, (iii) both distance and diversity controls, and (iv) a standard spin-splitting fitness GA. A 21-generation GA run requires a little over 5 min to complete (limited by complex assembly with partial force field optimization for optional follow-up DFT study), whereas fitness evaluation with DFT single points at guessed geometries[85] would require 4 days (Supporting Information Text S2). All molecules are built, scored, and evolved using an automated design extension to our molSimplify toolkit,[85] which is freely available online[86] (Supporting Information Text S3). As expected, the standard GA rapidly (i.e., within five generations) approaches a mean pool fitness of 1.0 through a dramatic drop in diversity corresponding to roughly one lead compound at the end of the GA run (Figure 2). Introducing diversity control improves the number of retained compounds, but diversity control or standard GA runs converge to high-distance/low-confidence leads ($d_{av} \approx 1.0$, Figure 2). Adjusting fitness evaluation from eq 1 to eq 2 with distance control reduces d_{av} to ∼0.5 (Figure 2). Introducing a distance control comes at the cost of slightly reducing the mean spin-splitting-only fitness of the retained ligands to ∼0.8, a modest increase in $|\Delta E_{H\text{-}L}|$ due to the exponential fitness function, and, interestingly, increases the pool diversity (Figure 2). Finally, combining both controls preserves the good features of both strategies: Diversity of leads at the end of a GA run is highest overall, and mean distance to training set is unchanged from distance control (Figure 2). Incorporating diversity or both controls increases the number of distinct complexes sampled in the GA runs by 50% (150 vs 100) over other modes and localizes retained hits to a narrow area of target distance and spin-splitting (Supporting Information Figures S3 and S4).

Over 50 repeats, roughly half (∼2800) of the compound space is sampled by the standard GA, and the slight reduction (∼2650) in compounds sampled with distance control is compensated by combination with diversity control (∼3300) (Supporting Information Table S4 and Figure S5). We evaluated the full feasible design space with the ANN in a little over 7 h on a standard desktop machine to identify the fraction of leads (i.e., $|\Delta E_{H\text{-}L}| < 5$ kcal/mol and $0.3 \leq d \leq 1.0$) missed during these GA optimizations (Supporting Information Text S2). Roughly 8% (474 complexes) of the constructed design space corresponds to our definition of lead compounds (Figure 3a and Supporting Information Table S4). Our recommended control strategy (both) recovers nearly 80% of the lead compounds, a substantial improvement over a standard GA or distance control. Most missed compounds are at larger (d > 0.5) distances (Figure 3b and Supporting Information Table S4 and Figure S6).

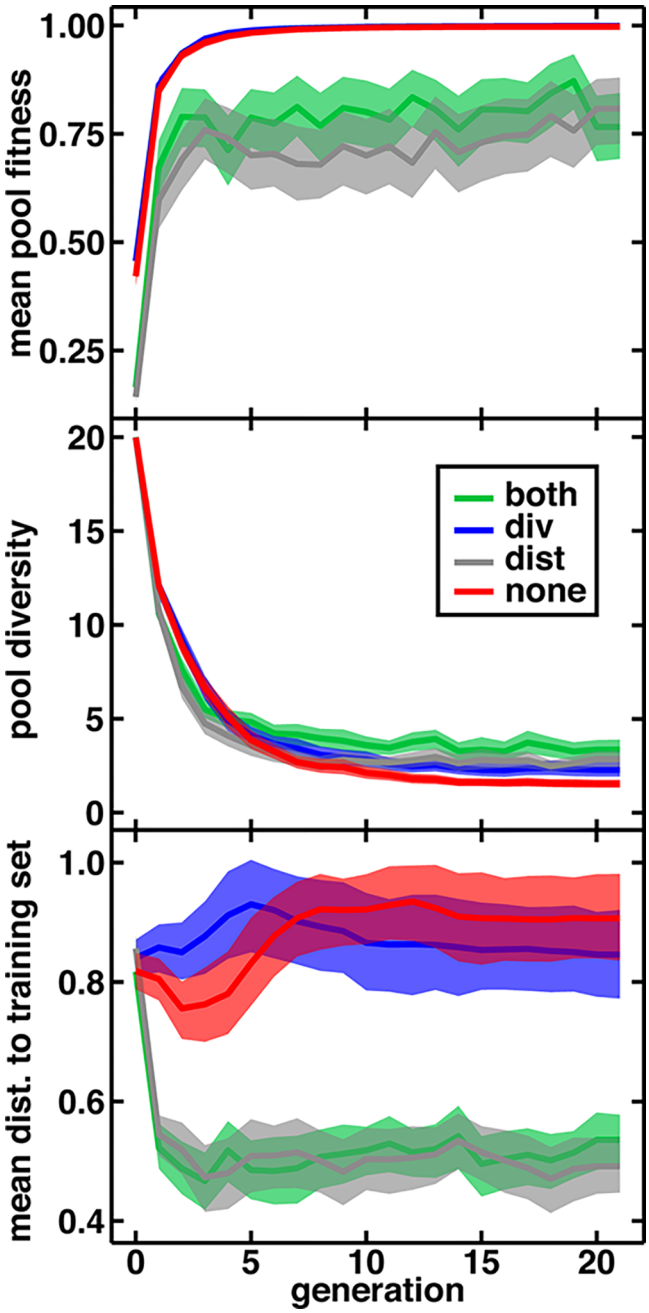


Figure 2. ANN GA runs with diversity and distance control (both, green), diversity control only (div, blue), distance only (dist, gray), and no controls on (none, red), as described in the main text: mean splitting-only pool fitness (top), diversity of the pool (middle), and mean distance to training data (bottom) with one standard deviation over 50 runs shown in translucent shading.

Dimensionality reduction[87] of the full compound space in continuous descriptors[56] similar to MCDL-25 reveals why it is challenging to ensure that a GA samples all compound leads (Figure 3a and Supporting Information Text S4). Although families of related complexes are reasonably well-clustered in this representation (i.e., most Fe(II)-substituted-bpy and nearly

DOI: 10.1021/acs.jpclett.8b00170
J. Phys. Chem. Lett. 2018, 9, 1064−1071

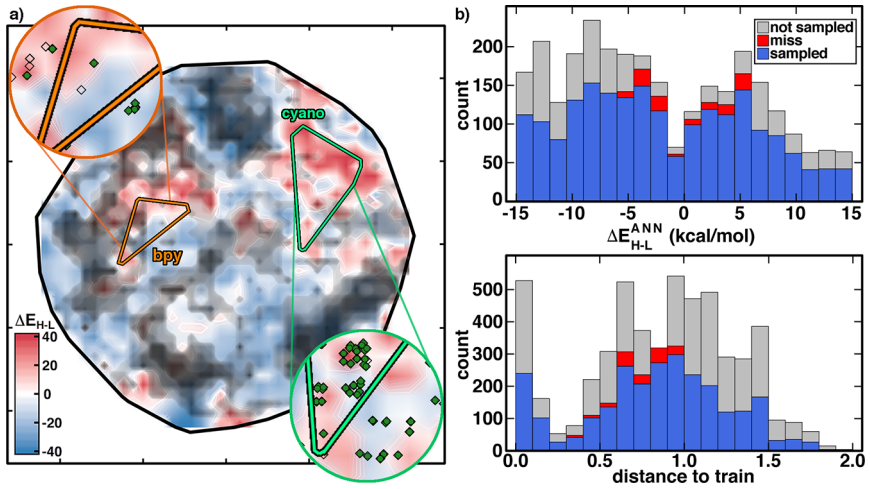


Figure 3. (a) t-SNE plot of the full compound space colored by $\Delta E_{H\text{-}L}$ (in kcal/mol as indicated in inset color bar) with increasingly high distance-to-train regions indicated in darker shades of gray. The convex hulls of two families of Fe(II) with substituted-bpy ligands and substituted-cyano ligands are indicated by orange and bright-green triangles, respectively. Insets show zooms to each of these regions with discrete hits in empty diamonds with sampled hits as filled dark-green circles. (b) 1D histograms of the ANN-predicted $\Delta E_{H\text{-}L}$ (top) and data distance to training data (bottom) using both controls in a stacked bar graph consisting of all sampled points (blue), nonsampled, nonhits (gray), and nonsampled hits (red).

all Fe(II)-substituted-cyano complexes are sampled in small regions), variation of properties in this space is quite rough (Figure 3a). Narrow target regions that correspond to SCOs are surrounded by nonleads, and several of these promising compound regions are in areas where the ANN confidence is low (black shading in Figure 3a).

Both distance- and diversity-controlled GA exploration provides a promising approach to reveal a large fraction of theoretical leads in compound space with an ANN. An additional concern is whether our distance control ensures reasonable reliability of the ANN-based fitness scoring. We quantify the ANN prediction accuracy over a randomly selected 51-complex subset (i.e., roughly 15%) of the 372 identified leads by fully geometry optimizing the high-spin and low-spin states (see Computational Details and Supporting Information Table S5). Overall performance on these newly generated complexes is good, with mean unsigned error (MUE) of 4.5 kcal/mol, 40% (80%) of all compounds predicted at or below 1× (2×) baseline error of the ANN[51] on a set-aside test set (Figure 4). Around 2/3 of ANN spin-crossover leads are validated (i.e., $|\Delta E_{H\text{-}L,DFT}| \leq 5$ kcal/mol) by DFT geometry optimization (Figure 4). Inclusion of solvent and thermodynamic corrections, which were omitted from ANN training or our fitness function, reduces this fraction only slightly to around 1/2 of candidates (22 of 49, see Supporting Information Table S5). Improvement upon this performance would likely require ANN training directly on $\Delta G_{H\text{-}L}$ rather than shifting the fitness function because inclusion of solvent and thermodynamic corrections does not produce a systematic shift of $\Delta G_{H\text{-}L}$ with respect to $\Delta E_{H\text{-}L}$. Unconventional, promising complexes (i.e., non-Fe(II)/N, with $\Delta G_{H\text{-}L} \approx 1.5$ kcal/mol) identified by the ANN and confirmed with DFT $\Delta G_{H\text{-}L}$ include Mn(II)-$(CNCH_3)_2(CO)_4$ and Fe(III)$(CO)_2(NCS)_4$. Conventional[76−79] complexes (e.g., Fe(II)$(phen)(en)_2$ and Fe(III)-$(NCS)_2(mebpy)_2$) are also captured (Supporting Information Table S5).

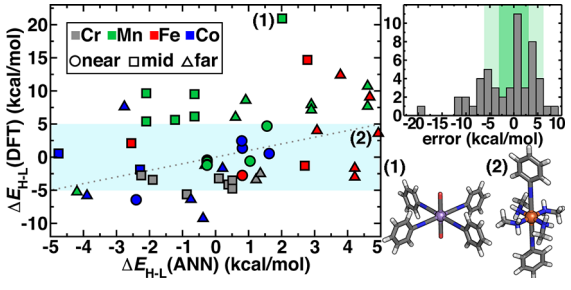


Figure 4. (left) B3LYP (DFT) geometry optimization $\Delta E_{H\text{-}L}$ versus ANN prediction distinguished by metal identity (Cr, gray; Mn, green; Fe, red; and Co, blue) and distance from closest training point (near, circles: <0.5, mid, squares: 0.5 to 0.75, and far, triangles: > 0.75). A parity line is shown (gray, dotted), and the $|\Delta E_{H\text{-}L}| < 5$ kcal/mol is shown in light blue. (right, top) Error histogram of ANN predictions with baseline error (medium green) and 2× baseline error (light green) regions shown. (right, bottom) Representative complexes corresponding to labels at left.

Categorizing distance to training data into near ($d < 0.5$), mid ($0.5 < d < 0.75$), and far ($0.75 < d < 1.0$) complexes reveals excellent prediction accuracy in the near subset (MUE = 1.5 kcal/mol) and nonmonotonically worsening performance for mid (MUE = 6.2 kcal/mol) and far (MUE = 4.7 kcal/mol) subsets (Figure 4 and Supporting Information Table S5 and Figure S7). Good performance is obtained for far iron complexes, such as Fe(II)$(CNPh)_2(NH_2CH_3)_4$ (d = 0.79 and error: 1.3 kcal/mol, see (2) in Figure 4). A systematic underprediction of Mn complexes is apparent, with MUE for all Mn of 5.1 versus 4.1 kcal/mol for remaining metals, despite comparable average distances over the two subsets (d_{av} = 0.65 vs 0.63). The most notable example is Mn(II)$(CO)_2(CNPh)_4$ (d = 0.51 and error: −18.9 kcal/mol, see (1) in Figure 4). The closest training points[51] are homoleptic Mn(II)$(CNCH_3)_6$ ($\Delta E_{H\text{-}L}$ = 10 kcal/mol) and Mn(II)$(CO)_6$ ($\Delta E_{H\text{-}L}$ = −6.6 kcal/mol), explaining why the ANN might predict the Mn(II)$(CO)_2(CNPh)_4$ complex to have near degenerate spin states (i.e., by averaging these two compounds), even though the ANN can predict nonadditive effects. The origin of this unexpected deviation is indicated by large (i.e., > 2.5 Å) Mn−CO distances in the DFT-optimized high-spin complex compared with low-spin complexes (1.9 Å), suggesting electronic structure differences in this sampled heteroleptic

DOI: 10.1021/acs.jpclett.8b00170
J. Phys. Chem. Lett. 2018, 9, 1064−1071

compound absent from the homoleptic-focused training data (Supporting Information). These observations at mid/far distances motivate adaptive retraining (i.e., to incorporate more heteroleptic combinations of weak and strong field ligands) for improved accuracy in using ANNs for discovery.

From a theoretical compound space of which only a fraction were likely spin-crossover complexes, the ANN-GA results are enriched in the number of DFT-level spin crossover complexes by around an order of magnitude (Supporting Information Text S2). The best estimates of a full geometry-optimization-driven GA run walltime are around 10−30 days, even with parallel evaluation of each generation (Supporting Information Text S2 and Figures S8 and S9). We considered alternatively using guessed[16,51,85] H and L geometries to evaluate single point (SP) $\Delta E_{H\text{-}L}$ with DFT, requiring around 4 days for a GA run (Supporting Information Text S2 and Figures S8 and S10). Imbalanced effects in bond length prediction errors[51] on spin-state ordering mean that the DFT-SP-GA performs worse than the ANN with MUE of 11 kcal/mol and only 30% of compounds remain spin-crossover complexes after geometry optimization (Supporting Information Table S6).

A final consideration in computational discovery of spin crossover complexes is the strong dependence of spin-state ordering on functional,[53,55,88−93] with few exceptions,[94] especially on admixture of Hartree−Fock (HF) exchange[53,55,88−92] due to differences in delocalization error between spin states.[95] Our ANN was trained on a range of HF exchange values, making it possible to identify SCOs in a functional-dependent manner. Rerunning the ANN GA with both controls at reduced 15% exchange (i.e., B3LYP*[88,92] vs 20% in B3LYP thus far) yields new candidates with weaker field ligands (e.g., $Mn(III)(NH_2CH_3)_4(4\text{-}CNPyr)_2$ and $Fe(II)(ox)_2(CN)_2$ $\Delta E_{H\text{-}L} \approx 0-2$ kcal/mol), in line with our prior observations[51,53] (Supporting Information Figure S11). Of leads predicted by the ANN, exchange sensitivity is predicted by the ANN to be lowest for Mn(III)/en ligand complexes or Co(II) complexes, and this kind of functional invariance could be a useful metric in future multiobjective optimization.

In conclusion, we have demonstrated an ML-driven strategy for accelerating SCO discovery with an ANN. By pairing our trained ML model with a strategy for controlling the novelty of leads in the GA, we discover complexes sufficiently distinct from training data but for which the ML model can still be suitably employed to make predictions. Using this approach, we have explored a space of >5500 candidate materials generated from eight possible metal/oxidation state combinations and 32 possible ligands. Of over 51 representative spin-crossover complexes distinct from ANN training points, average unsigned errors (4.5 kcal/mol) are close to the ANN's baseline 3.1 kcal/mol error. Two thirds of the discovered compounds, including unconventional complexes, are still considered spin-crossover candidates after full DFT geometry optimization. The largest errors can be avoided in future work by applying an even more conservative distance control, using a series of independently trained ANNs, or enriching the data set with more heteroleptic compounds. This strategy demonstrates the power of ML for accelerating materials discovery through prescreening vast chemical space. In future work, we will identify ways to exploit (instead of avoid) high-promise, low-confidence compounds for adaptive retraining of ML models during discovery. We expect this suite of ML models, discovery algorithms, and simulation automation software to be valuable for the optimization of key properties in inorganic chemistry.

COMPUTATIONAL DETAILS

Single-point energies and geometry optimizations were carried out with TeraChem[3,96] at the B3LYP[97−99] level of theory with LANL2DZ effective core potential[100] for all transition metals, bromine, and iodine and the 6-31G* basis for the remaining atoms, as employed during ANN training.[51] Basis set dependence is observed to be small (Supporting Information Table S7). Although their inclusion has been motivated,[101] vibrational or solvent contributions, which often have compensating effects,[16,101] are neglected during fitness scoring by DFT or with the trained ANN (Supporting Information Table S5). On representative molecules, vibrational enthalpy and entropy corrections were obtained through calculation of the gas-phase Hessian of each spin state. Solvent corrections were obtained from differences in solvation free energy on the gas-phase geometries using COSMO[86,102] (ε = 78.39 and a cavity constructed from 1.2× Bondi radii[103]).

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.8b00170.

Information and structures of ligands used in the GA pool; calculation of allowed ligand combinations; high-spin and low-spin multiplicity definitions; response of fitness function; GA parameters and algorithm; timing information for DFT and ANN calculations; mAD syntax; sampling of complexes with varied control modes; description of dimensionality reduction approach; ANN prediction errors; error versus distance metric for ANN predictions; detailed timing estimates for DFT single-point energies, geometry optimizations, and DFT-SP-GAs; errors of DFT-SP-GA hits; results of modified exchange ANN GA; and basis set dependence (PDF)

Summary data of all GA runs (ZIP)

DFT-optimized geometries of DFT-SP and ANN GA hits (ZIP)

AUTHOR INFORMATION

Corresponding Author

*E-mail: hjkulik@mit.edu. Tel: 617-253-4584.

ORCID

Jon Paul Janet: 0000-0001-7825-4797

Heather J. Kulik: 0000-0001-9342-0191

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge support by the Office of Naval Research under grant number N00014-17-1-2956, the Department of Energy under grant number DE-SC0018096, the National Science Foundation under grant number CBET-1704266, and MIT Energy Initiative seed grants (2014, 2017). J.P.J. was supported in part by an MIT-SUTD Graduate fellowship. H.J.K. holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund. This work was carried out in part using computational resources from the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. This work used the XStream computational

DOI: 10.1021/acs.jpclett.8b00170
J. Phys. Chem. Lett. 2018, 9, 1064−1071

resource, supported by the National Science Foundation Major Research Instrumentation program (ACI-1429830). We thank Adam H. Steeves for providing a critical reading of the manuscript.

REFERENCES

(1) Ufimtsev, I. S.; Martínez, T. J. Quantum Chemistry on Graphical Processing Units. 1. Strategies for Two-Electron Integral Evaluation. *J. Chem. Theory Comput.* **2008**, *4*, 222−231.
(2) Ufimtsev, I. S.; Martínez, T. J. Quantum Chemistry on Graphical Processing Units. 2. Direct Self-Consistent-Field Implementation. *J. Chem. Theory Comput.* **2009**, *5*, 1004−1015.
(3) Ufimtsev, I. S.; Martínez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 2619−2628.
(4) Ochsenfeld, C.; Kussmann, J.; Lambrecht, D. S. Linear-Scaling Methods in Quantum Chemistry. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Cundari, T. R., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, 2007; Vol. *23*, pp 1−82.
(5) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. Auxiliary Basis Sets for Main Row Atoms and Transition Metals and Their Use to Approximate Coulomb Potentials. *Theor. Chem. Acc.* **1997**, *97*, 119−124.
(6) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. Auxiliary Basis Sets to Approximate Coulomb Potentials. *Chem. Phys. Lett.* **1995**, *240*, 283−290.
(7) Nørskov, J. K.; Bligaard, T. The Catalyst Genome. *Angew. Chem., Int. Ed.* **2013**, *52*, 776−777.
(8) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; et al. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, *1*, 011002.
(9) Greeley, J.; Jaramillo, T. F.; Bonde, J.; Chorkendorff, I.; Nørskov, J. K. Computational High-Throughput Screening of Electrocatalytic Materials for Hydrogen Evolution. *Nat. Mater.* **2006**, *5*, 909−913.
(10) Nørskov, J. K.; Bligaard, T.; Rossmeisl, J.; Christensen, C. H. Towards the Computational Design of Solid Catalysts. *Nat. Chem.* **2009**, *1*, 37−46.
(11) Hautier, G.; Fischer, C. C.; Jain, A.; Mueller, T.; Ceder, G. Finding Nature's Missing Ternary Oxide Compounds Using Machine Learning and Density Functional Theory. *Chem. Mater.* **2010**, *22*, 3762−3767.
(12) Jain, A.; Hautier, G.; Moore, C. J.; Ping Ong, S.; Fischer, C. C.; Mueller, T.; Persson, K. A.; Ceder, G. A High-Throughput Infrastructure for Density Functional Theory Calculations. *Comput. Mater. Sci.* **2011**, *50*, 2295−2310.
(13) Hautier, G.; Miglio, A.; Ceder, G.; Rignanese, G.-M.; Gonze, X. Identification and Design Principles of Low Hole Effective Mass P-Type Transparent Conducting Oxides. *Nat. Commun.* **2013**, *4*, 2292.
(14) Gani, T. Z. H.; Ioannidis, E. I.; Kulik, H. J. Computational Discovery of Hydrogen Bond Design Rules for Electrochemical Ion Separation. *Chem. Mater.* **2016**, *28*, 6207−6218.
(15) Kim, J. Y.; Steeves, A. H.; Kulik, H. J. Harnessing Organic Ligand Libraries for First-Principles Inorganic Discovery: Indium Phosphide Quantum Dot Precursor Design Strategies. *Chem. Mater.* **2017**, *29*, 3632−3643.
(16) Janet, J. P.; Gani, T. Z. H.; Steeves, A. H.; Ioannidis, E. I.; Kulik, H. J. Leveraging Cheminformatics Strategies for Inorganic Discovery: Application to Redox Potential Design. *Ind. Eng. Chem. Res.* **2017**, *56*, 4898−4910.
(17) Curtarolo, S.; Hart, G. L.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O. The High-Throughput Highway to Computational Materials Design. *Nat. Mater.* **2013**, *12*, 191−201.
(18) Jensen, P. B.; Bialy, A.; Blanchard, D.; Lysgaard, S.; Reumert, A. K.; Quaade, U. J.; Vegge, T. Accelerated DFT-Based Design of Materials for Ammonia Storage. *Chem. Mater.* **2015**, *27*, 4552−4561.
(19) Bowman, D. N.; Bondarev, A.; Mukherjee, S.; Jakubikova, E. Tuning the Electronic Structure of Fe(II) Polypyridines via Donor Atom and Ligand Scaffold Modifications: A Computational Study. *Inorg. Chem.* **2015**, *54*, 8786−8793.
(20) Eckert, H.; Bajorath, J. Molecular Similarity Analysis in Virtual Screening: Foundations, Limitations and Novel Approaches. *Drug Discovery Today* **2007**, *12*, 225−233.
(21) Hachmann, J.; Olivares-Amaya, R.; Atahan-Evrenk, S.; Amador-Bedolla, C.; Sánchez-Carrera, R. S.; Gold-Parker, A.; Vogt, L.; Brockway, A. M.; Aspuru-Guzik, A. The Harvard Clean Energy Project: Large-Scale Computational Screening and Design of Organic Photovoltaics on the World Community Grid. *J. Phys. Chem. Lett.* **2011**, *2*, 2241−2251.
(22) Shu, Y.; Levine, B. G. Simulated Evolution of Fluorophores for Light Emitting Diodes. *J. Chem. Phys.* **2015**, *142*, 104104.
(23) Virshup, A. M.; Contreras-García, J.; Wipf, P.; Yang, W.; Beratan, D. N. Stochastic Voyages into Uncharted Chemical Space Produce a Representative Library of All Possible Drug-Like Compounds. *J. Am. Chem. Soc.* **2013**, *135*, 7296−7303.
(24) Kirkpatrick, P.; Ellis, C. Chemical Space. *Nature* **2004**, *432*, 823−823.
(25) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Quantum Chemistry Structures and Properties of 134 Kilo Molecules. *Sci. Data* **2014**, *1*, 140022.
(26) Collins, C. R.; Gordon, G. J.; von Lilienfeld, O. A.; Yaron, D. J. Constant Size Molecular Descriptors for Use with Machine Learning. 2017, arXiv:1701.06649. arXiv.org e-Print archive. https://arxiv.org/abs/1701.06649 (accessed February 8, 2018).
(27) Huang, B.; von Lilienfeld, O. A. Communication: Understanding Molecular Representations in Machine Learning: The Role of Uniqueness and Target Similarity. *J. Chem. Phys.* **2016**, *145*, 161102.
(28) Yao, K.; Herr, J. E.; Brown, S. N.; Parkhill, J. Intrinsic Bond Energies from a Bonds-in-Molecules Neural Network. *J. Phys. Chem. Lett.* **2017**, *8*, 2689−2694.
(29) Faber, F. A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S. S.; Dahl, G. E.; Vinyals, O.; Kearnes, S.; Riley, P. F.; von Lilienfeld, O. A. Prediction Errors of Molecular Machine Learning Models Lower Than Hybrid DFT Error. *J. Chem. Theory Comput.* **2017**, *13*, 5255−5264.
(30) Meredig, B.; Agrawal, A.; Kirklin, S.; Saal, J. E.; Doak, J. W.; Thompson, A.; Zhang, K.; Choudhary, A.; Wolverton, C. Combinatorial Screening for New Materials in Unconstrained Composition Space with Machine Learning. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 094104.
(31) Ma, X.; Li, Z.; Achenie, L. E. K.; Xin, H. Machine-Learning-Augmented Chemisorption Model for CO2 Electroreduction Catalyst Screening. *J. Phys. Chem. Lett.* **2015**, *6*, 3528−3533.
(32) Li, Z.; Ma, X.; Xin, H. Feature Engineering of Machine-Learning Chemisorption Models for Catalyst Design. *Catal. Today* **2017**, *280*, 232−238.
(33) Huo, H.; Rupp, M. Unified Representation for Machine Learning of Molecules and Crystals. 2017, arXiv:1704.06439v3. arXiv.org e-Print archive. https://arxiv.org/abs/1704.06439 (accessed February 8, 2018).
(34) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Delta-Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087−96.
(35) Huang, S.-D.; Shang, C.; Zhang, X.-J.; Liu, Z.-P. Material Discovery by Combining Stochastic Surface Walking Global Optimization with a Neural Network. *Chem. Sci.* **2017**, *8*, 6327−6337.
(36) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.
(37) Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett.* **2015**, *6*, 2326−2331.

DOI: 10.1021/acs.jpclett.8b00170
J. Phys. Chem. Lett. 2018, 9, 1064−1071

(38) Montavon, G.; Hansen, K.; Fazli, S.; Rupp, M.; Biegler, F.; Ziehe, A.; Tkatchenko, A.; von Lilienfeld, O. A.; Muller, K.-R. In *Learning Invariant Representations of Molecules for Atomization Energy Prediction*, Proceedings of the Neural Information Processing Systems Conference, Lake Tahoe, NV, December 3−8, 2012; Pereira, F.; Burges, C. J. C.; Bottou, L.; Weinberger, K. Q., Eds.; Curran Associates, Inc.: New York, 2012; pp 440−448.
(39) Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Müller, K.-R. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **2013**, *9*, 3404−3419.
(40) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(41) Benson, S. W.; Cruickshank, F.; Golden, D.; Haugen, G. R.; O'neal, H.; Rodgers, A.; Shaw, R.; Walsh, R. Additivity Rules for the Estimation of Thermochemical Properties. *Chem. Rev.* **1969**, *69*, 279−324.
(42) Maggiora, G.; Vogt, M.; Stumpfe, D.; Bajorath, J. Molecular Similarity in Medicinal Chemistry: Miniperspective. *J. Med. Chem.* **2014**, *57*, 3186−3204.
(43) Kubinyi, H. QSAR and 3D QSAR in Drug Design. Part 1: Methodology. *Drug Discovery Today* **1997**, *2*, 457−467.
(44) Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **2018**, DOI: 10.1021/acscentsci.7b00572.
(45) Hernández-Lobato, J. M.; Requeima, J.; Pyzer-Knapp, E. O.; Aspuru-Guzik, A. Parallel and Distributed Thompson Sampling for Large-Scale Accelerated Exploration of Chemical Space. 2017, arXiv:1706.01825. arXiv.org e-Print archive. https://arxiv.org/abs/1706.01825 (accessed February 8, 2018).
(46) Deeth, R. J. The Ligand Field Molecular Mechanics Model and the Stereoelectronic Effects of d and S Electrons. *Coord. Chem. Rev.* **2001**, *212*, 11−34.
(47) Tortorella, S.; Marotta, G.; Cruciani, G.; De Angelis, F. Quantitative Structure-Property Relationship Modeling of Ruthenium Sensitizers for Solar Cells Applications: Novel Tools for Designing Promising Candidates. *RSC Adv.* **2015**, *5*, 23865−23873.
(48) Cruz, V. L.; Martinez, S.; Ramos, J.; Martinez-Salazar, J. 3D-QSAR as a Tool for Understanding and Improving Single-Site Polymerization Catalysts. A Review. *Organometallics* **2014**, *33*, 2944−2959.
(49) Fey, N.; Orpen, A. G.; Harvey, J. N. Building Ligand Knowledge Bases for Organometallic Chemistry: Computational Description of Phosphorus (III)-Donor Ligands and the Metal−Phosphorus Bond. *Coord. Chem. Rev.* **2009**, *253*, 704−722.
(50) Le, T.; Epa, V. C.; Burden, F. R.; Winkler, D. A. Quantitative Structure−Property Relationship Modeling of Diverse Materials Properties. *Chem. Rev.* **2012**, *112*, 2889−2919.
(51) Janet, J. P.; Kulik, H. J. Predicting Electronic Structure Properties of Transition Metal Complexes with Neural Networks. *Chem. Sci.* **2017**, *8*, 5137−5152.
(52) Hastie, T.; Tibshirani, R.; Friedman, J. H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, 2009; p *xxii*, 745 p.
(53) Ioannidis, E. I.; Kulik, H. J. Towards Quantifying the Role of Exact Exchange in Predictions of Transition Metal Complex Properties. *J. Chem. Phys.* **2015**, *143*, 034104.
(54) 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.
(55) Bowman, D. N.; Jakubikova, E. Low-Spin Versus High-Spin Ground State in Pseudo-Octahedral Iron Complexes. *Inorg. Chem.* **2012**, *51*, 6011−6019.
(56) 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.
(57) Gómez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Duvenaud, D.; Maclaurin, D.; Blood-Forsythe, M. A.; Chae, H. S.; Einzinger, M.; Ha, D.-G.; Wu, T.; et al. Design of Efficient Molecular Organic Light-Emitting Diodes by a High-Throughput Virtual Screening and Experimental Approach. *Nat. Mater.* **2016**, *15*, 1120.
(58) Seko, A.; Hayashi, H.; Nakayama, K.; Takahashi, A.; Tanaka, I. Representation of Compounds for Machine-Learning Prediction of Physical Properties. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2017**, *95*, 144110.
(59) Kim, C.; Pilania, G.; Ramprasad, R. From Organized High-Throughput Data to Phenomenological Theory Using Machine Learning: The Example of Dielectric Breakdown. *Chem. Mater.* **2016**, *28*, 1304−1311.
(60) Pilania, G.; Wang, C.; Jiang, X.; Rajasekaran, S.; Ramprasad, R. Accelerating Materials Property Predictions Using Machine Learning. *Sci. Rep.* **2013**, *3*, 2810.
(61) Keinan, S.; Hu, X.; Beratan, D. N.; Yang, W. Designing Molecules with Optimal Properties Using the Linear Combination of Atomic Potentials Approach in an Am1 Semiempirical Framework. *J. Phys. Chem. A* **2007**, *111*, 176−181.
(62) Keinan, S.; Therien, M. J.; Beratan, D. N.; Yang, W. Molecular Design of Porphyrin-Based Nonlinear Optical Materials. *J. Phys. Chem. A* **2008**, *112*, 12203−12207.
(63) Wang, M.; Hu, X.; Beratan, D. N.; Yang, W. Designing Molecules by Optimizing Potentials. *J. Am. Chem. Soc.* **2006**, *128*, 3228−3232.
(64) Weymuth, T.; Reiher, M. Gradient-Driven Molecule Construction: An Inverse Approach Applied to the Design of Small-Molecule Fixating Catalysts. *Int. J. Quantum Chem.* **2014**, *114*, 838−850.
(65) Qin, J.; Khaira, G. S.; Su, Y.; Garner, G. P.; Miskin, M.; Jaeger, H. M.; de Pablo, J. J. Evolutionary Pattern Design for Copolymer Directed Self-Assembly. *Soft Matter* **2013**, *9*, 11467−11472.
(66) Leardi, R. Genetic Algorithms in Chemistry. *J. Chromatogr. A* **2007**, *1158*, 226−233.
(67) Leardi, R. Genetic Algorithms in Chemometrics and Chemistry: A Review. *J. Chemom.* **2001**, *15*, 559−569.
(68) Chu, Y.; Heyndrickx, W.; Occhipinti, G.; Jensen, V. R.; Alsberg, B. K. An Evolutionary Algorithm for De Novo Optimization of Functional Transition Metal Compounds. *J. Am. Chem. Soc.* **2012**, *134*, 8885−8895.
(69) Venkatraman, V.; Abburu, S.; Alsberg, B. K. Artificial Evolution of Coumarin Dyes for Dye Sensitized Solar Cells. *Phys. Chem. Chem. Phys.* **2015**, *17*, 27672−27682.
(70) Bousseksou, A.; Molnár, G.; Matouzenko, G. Switching of Molecular Spin States in Inorganic Complexes by Temperature, Pressure, Magnetic Field and Light: Towards Molecular Devices. *Eur. J. Inorg. Chem.* **2004**, *2004*, 4353−4369.
(71) Decurtins, S.; Gütlich, P.; Köhler, C.; Spiering, H.; Hauser, A. Light-Induced Excited Spin State Trapping in a Transition-Metal Complex: The Hexa-1-Propyltetrazole-Iron (II) Tetrafluoroborate Spin-Crossover System. *Chem. Phys. Lett.* **1984**, *105*, 1−4.
(72) Hauser, A. Light-Induced Spin Crossover and the High-Spin→Low-Spin Relaxation. In *Spin Crossover in Transition Metal Compounds II*; Springer: 2004; pp 155−198.
(73) Reed, D. A.; Xiao, D. J.; Gonzalez, M. I.; Darago, L. E.; Herm, Z. R.; Grandjean, F.; Long, J. R. Reversible CO Scavenging via Adsorbate-Dependent Spin State Transitions in an Iron (II)−Triazolate Metal−Organic Framework. *J. Am. Chem. Soc.* **2016**, *138*, 5594−5602.
(74) Groizard, T.; Papior, N.; Le Guennic, B.; Robert, V.; Kepenekian, M. Enhanced Cooperativity in Supported Spin-Crossover Metal−Organic Frameworks. *J. Phys. Chem. Lett.* **2017**, *8*, 3415−3420.
(75) Neville, S. M.; Halder, G. J.; Chapman, K. W.; Duriska, M. B.; Moubaraki, B.; Murray, K. S.; Kepert, C. J. Guest Tunable Structure and Spin Crossover Properties in a Nanoporous Coordination Framework Material. *J. Am. Chem. Soc.* **2009**, *131*, 12106−12108.

DOI: 10.1021/acs.jpclett.8b00170
J. Phys. Chem. Lett. 2018, 9, 1064−1071

(76) Gütlich, P. *Spin Crossover in Iron (II)-Complexes* **1981**, *44*, 83−195.
(77) Gütlich, P.; Hauser, A. Thermal and Light-Induced Spin Crossover in Iron (II) Complexes. *Coord. Chem. Rev.* **1990**, *97*, 1−22.
(78) Halcrow, M. A. Structure: Function Relationships in Molecular Spin-Crossover Complexes. *Chem. Soc. Rev.* **2011**, *40*, 4119−4142.
(79) Bousseksou, A.; Molnár, G.; Salmon, L.; Nicolazzi, W. Molecular Spin Crossover Phenomenon: Recent Achievements and Prospects. *Chem. Soc. Rev.* **2011**, *40*, 3313−3335.
(80) Phan, H.; Hrudka, J. J.; Igimbayeva, D.; Lawson Daku, L. v. 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.
(81) Rodríguez-Jiménez, S.; Yang, M.; Stewart, I.; Garden, A. L.; Brooker, S. A Simple Method of Predicting Spin State in Solution. *J. Am. Chem. Soc.* **2017**, *139*, 18392−18396.
(82) Gal, Y.; Ghahramani, Z. In *Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning*, International Conference on Machine Learning, 2016; pp 1050−1059.
(83) Kier, L. B. A Shape Index from Molecular Graphs. *Quant. Struct.-Act. Relat.* **1985**, *4*, 109−116.
(84) Fias, S.; Heidar-Zadeh, F.; Geerlings, P.; Ayers, P. W. Chemical Transferability of Functional Groups Follows from the Nearsightedness of Electronic Matter. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 11633−11638.
(85) Ioannidis, E. I.; Gani, T. Z. H.; Kulik, H. J. Molsimplify: A Toolkit for Automating Discovery in Inorganic Chemistry. *J. Comput. Chem.* **2016**, *37*, 2106−2117.
(86) Klamt, A.; Schuurmann, G. Cosmo: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and Its Gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, *2*, 799−805.
(87) van der Maaten, L.; Hinton, G. Visualizing Data Using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579−2605.
(88) Reiher, M.; Salomon, O.; Artur Hess, B. Reparameterization of Hybrid Functionals Based on Energy Differences of States of Different Multiplicity. *Theor. Chem. Acc.* **2001**, *107*, 48−55.
(89) Reiher, M. Theoretical Study of the Fe (Phen) 2 (NCS) 2 Spin-Crossover Complex with Reparametrized Density Functionals. *Inorg. Chem.* **2002**, *41*, 6928−6935.
(90) Ganzenmüller, G.; Berkaïne, N.; Fouqueau, A.; Casida, M. E.; Reiher, M. Comparison of Density Functionals for Differences between the High- (T2g5) and Low- (A1g1) Spin States of Iron(II) Compounds. IV. Results for the Ferrous Complexes [Fe(L)('NHS4')]. *J. Chem. Phys.* **2005**, *122*, 234321.
(91) Droghetti, A.; Alfè, D.; Sanvito, S. Assessment of Density Functional Theory for Iron (II) Molecules across the Spin-Crossover Transition. *J. Chem. Phys.* **2012**, *137*, 124303.
(92) Salomon, O.; Reiher, M.; Hess, B. A. Assertion and Validation of the Performance of the B3LYP* Functional for the First Transition Metal Row and the G2 Test Set. *J. Chem. Phys.* **2002**, *117*, 4729−4737.
(93) Ioannidis, E. I.; Kulik, H. J. Ligand-Field-Dependent Behavior of Meta-GGA Exchange in Transition-Metal Complex Spin-State Ordering. *J. Phys. Chem. A* **2017**, *121*, 874−884.
(94) Wilbraham, L.; Verma, P.; Truhlar, D. G.; Gagliardi, L.; Ciofini, I. Multiconfiguration Pair-Density Functional Theory Predicts Spin-State Ordering in Iron Complexes with the Same Accuracy as Complete Active Space Second-Order Perturbation Theory at a Significantly Reduced Computational Cost. *J. Phys. Chem. Lett.* **2017**, *8*, 2026−2030.
(95) Gani, T. Z. H.; Kulik, H. J. Unifying Exchange Sensitivity in Transition Metal Spin-State Ordering and Catalysis through Bond Valence Metrics. *J. Chem. Theory Comput.* **2017**, *13*, 5443−5457.
(96) *TeraChem v 1.9*; PetaChem, LLC, 2015. http://www.petachem.com. (accessed February 8, 2018).
(97) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623−11627.
(98) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(99) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.
(100) Hay, P. J.; Wadt, W. R. Ab Initio Effective Core Potentials for Molecular Calculations. Potentials for the Transition Metal Atoms Sc to Hg. *J. Chem. Phys.* **1985**, *82*, 270−283.
(101) Mortensen, S. R.; Kepp, K. P. Spin Propensities of Octahedral Complexes from Density Functional Theory. *J. Phys. Chem. A* **2015**, *119*, 4041−4050.
(102) Liu, F.; Luehr, N.; Kulik, H. J.; Martínez, T. J. Quantum Chemistry for Solvated Molecules on Graphical Processing Units Using Polarizable Continuum Models. *J. Chem. Theory Comput.* **2015**, *11*, 3131−3144.
(103) Bondi, A. Van Der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441−451.

DOI: 10.1021/acs.jpclett.8b00170
J. Phys. Chem. Lett. 2018, 9, 1064−1071