

Cite this: *Chem. Sci.*, 2017, 8, 5137

# Predicting electronic structure properties of transition metal complexes with neural networks†

Jon Paul Janet  and Heather J. Kulik \*

High-throughput computational screening has emerged as a critical component of materials discovery. Direct density functional theory (DFT) simulation of inorganic materials and molecular transition metal complexes is often used to describe subtle trends in inorganic bonding and spin-state ordering, but these calculations are computationally costly and properties are sensitive to the exchange–correlation functional employed. To begin to overcome these challenges, we trained artificial neural networks (ANNs) to predict quantum-mechanically-derived properties, including spin-state ordering, sensitivity to Hartree–Fock exchange, and spin-state specific bond lengths in transition metal complexes. Our ANN is trained on a small set of inorganic-chemistry-appropriate empirical inputs that are both maximally transferable and do not require precise three-dimensional structural information for prediction. Using these descriptors, our ANN predicts spin-state splittings of single-site transition metal complexes (*i.e.*, Cr–Ni) at arbitrary amounts of Hartree–Fock exchange to within 3 kcal mol<sup>−1</sup> accuracy of DFT calculations. Our exchange-sensitivity ANN enables improved predictions on a diverse test set of experimentally-characterized transition metal complexes by extrapolation from semi-local DFT to hybrid DFT. The ANN also outperforms other machine learning models (*i.e.*, support vector regression and kernel ridge regression), demonstrating particularly improved performance in transferability, as measured by prediction errors on the diverse test set. We establish the value of new uncertainty quantification tools to estimate ANN prediction uncertainty in computational chemistry, and we provide additional heuristics for identification of when a compound of interest is likely to be poorly predicted by the ANN. The ANNs developed in this work provide a strategy for screening transition metal complexes both with direct ANN prediction and with improved structure generation for validation with first principles simulation.

Received 19th March 2017

Accepted 9th May 2017

DOI: 10.1039/c7sc01247k

rsc.li/chemical-science

## 1. Introduction

High-throughput computational screening has become a leading component of the workflow for identifying new molecules,<sup>1,2</sup> catalysts,<sup>3</sup> and materials.<sup>4</sup> First-principles simulation remains critical to many screening and discovery studies, but relatively high computational cost of direct simulation limits exploration of chemical space to a small fraction of feasible compounds.<sup>5,6</sup> In order to accelerate discovery, lower levels of theory, including machine-learning models, have emerged as alternate approaches for efficient evaluation of new candidate materials.<sup>7</sup> Artificial neural networks (ANNs) have recently found wide application in the computational chemistry community.<sup>8–10</sup> Machine learning approaches were initially appreciated for their flexibility to fit potential energy surfaces and thus force field models.<sup>10–17</sup> Broader applications have recently been explored, including in exchange–correlation functional development,<sup>8,18</sup> general solutions to the

Schrödinger equation,<sup>19</sup> orbital free density functional theory,<sup>20,21</sup> many body expansions,<sup>22</sup> acceleration of dynamics,<sup>23–25</sup> band-gap prediction,<sup>26,27</sup> and molecular<sup>1,2</sup> or heterogeneous catalyst<sup>28</sup> and materials<sup>29–32</sup> discovery, to name a few.

Essential challenges for ANNs to replace direct calculation by first-principles methods include the appropriate determination of broadly applicable descriptors that enable the use of the ANN flexibly beyond molecules in the training set, *e.g.* for larger molecules or for those with diverse chemistry. Indeed, the most successful applications of ANNs at this time beyond proof-of-concept demonstrations have been in the development of force fields for well-defined compositions, *e.g.* of water.<sup>33,34</sup> Within organic chemistry, structural descriptors such as a Coulomb matrix<sup>35</sup> or local descriptions of the chemical environment and bonding<sup>36,37</sup> have been useful to enable predictions of energetics as long as a relatively narrow range of compositions is considered (*e.g.*, C, H, N, O compounds). These observations are consistent with previous successes in cheminformatics for evaluating molecular similarity,<sup>38</sup> force field development,<sup>39</sup> quantitative structure–activity relationships,<sup>40</sup> and group additivity<sup>41</sup> theories. For transition metal complexes, few force fields have been established that can capture a full range of inorganic

Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA. E-mail: [hjkulik@mit.edu](mailto:hjkulik@mit.edu); Tel: +1-617-253-4584

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7sc01247k



chemical bonding,<sup>42</sup> and the spin-state- and coordination-environment-dependence of bonding<sup>43</sup> suggests that more careful development of descriptors is required to broadly predict properties of open-shell transition metal complexes. Similarly, descriptors that worked well for organic molecules have been demonstrated to not be suitable in inorganic crystalline materials.<sup>44</sup> It is well-known<sup>45–47</sup> that there is a strong relationship between sensitivity of electronic properties (e.g., spin-state splitting) and the direct ligand–atom and ligand field strength<sup>48,49</sup> in transition-metal complexes. Since ligands with the same direct metal-bonding atom can have substantially different ligand-field strengths (e.g., C for both weaker field CH<sub>3</sub>CN *versus* strong-field CO), whereas distant substitutions (e.g., tetraphenylporphyrin *vs.* base porphine) will have a limited effect, a transition-metal complex descriptor set that carefully balances metal-proximal and metal-distant descriptors is needed.

Within transition metal chemistry and correlated, inorganic materials, a second concern arises for the development of ANN predictions of first-principles properties. Although efficient correlated wavefunction theory methods (e.g., MP2) may be straightforwardly applied to small organic molecules, such methods are not appropriate for transition metal complexes where best practices remain an open question.<sup>50</sup> Although promising avenues for ANNs include the mapping of lower-level theory results, e.g. from semi-empirical theory,<sup>51</sup> to a higher-level one, as has been demonstrated on atomization energies<sup>52</sup> and more recently reaction barriers,<sup>53</sup> suitable levels of theory for extrapolation are less clear in transition metal chemistry.

Additionally, uncertainty remains about the amount of exact (Hartree–Fock, HF) exchange to include in study of transition metal complexes, with recommendations ranging from no exchange, despite disproportionate delocalization errors in approximate DFT on transition metal complexes,<sup>48,54,55</sup> to alternately low<sup>56–58</sup> or high<sup>59</sup> amounts of exact exchange in a system-dependent manner. Indeed, there has been much interest recently in quantifying uncertainty with respect to functional choice in energetic predictions,<sup>60–62</sup> including through evaluation of sensitivity of predictions with respect to inclusion of exact exchange.<sup>45,59</sup> Spin-state splitting is particularly sensitive to exchange fraction,<sup>45–47</sup> making it a representative quantity for which it is useful to obtain both a direct value and its sensitivity to varying the exchange fraction. Thus, a machine-learning model that predicts spin-state ordering across exchange values will be useful for translating literature predictions or providing sensitivity measures on computed data.

Overall, a demonstration of ANNs in inorganic chemistry, e.g. for efficient discovery of new spin-crossover complexes,<sup>63,64</sup> for dye-sensitizers in solar cells,<sup>65</sup> or for identification of reactivity of open-shell catalysts<sup>66</sup> *via* rapid evaluation of spin-state ordering should satisfy two criteria: (i) contain flexible descriptors that balance metal-proximal and metal-distant features and (ii) be able to predict spin-state ordering across exchange–correlation mixing. In this work, we make progress toward both of these aims, harnessing cheminformatics-inspired transition metal-complex structure generation tools<sup>67</sup>

and established structure–functional sensitivity relationships in transition metal complexes<sup>45,59</sup> to train ANNs for transition metal complex property prediction.

The outline of the rest of this work is as follows. In Section 2 (Methods), we review the computational details of data set generation, we discuss our variable selection procedure, and we review details of the artificial neural network trained. In Section 3, we provide the Results and discussion on the trained neural networks for spin-state ordering, spin-state exchange sensitivity, and bond-length prediction on both training-set-representative complexes and diverse experimental complexes. Finally, in Section 4, we provide our Conclusions.

## 2. Methods

### 2.1 Test set construction and simulation details

**Data set construction.** Our training set consists of octahedral complexes of first-row transition metals in common oxidation states: Cr<sup>2+/3+</sup>, Mn<sup>2+/3+</sup>, Fe<sup>2+/3+</sup>, Co<sup>2+/3+</sup>, and Ni<sup>2+</sup>. High-spin (H) and low-spin (L) multiplicities were selected for each metal from the ground, high-spin state of the isolated atom and the higher-energy, lowest-spin state within 5 eV that had a consistent d-orbital occupation for both states, as obtained from the National Institute of Standards and Technology atomic spectra database.<sup>68</sup> The selected H–L states were: triplet-singlet for Ni<sup>2+</sup>, quartet-doublet for Co<sup>2+</sup> and Cr<sup>3+</sup>, quintet-singlet for Fe<sup>2+</sup> and Co<sup>3+</sup>, quintet-triplet for Cr<sup>2+</sup> and Mn<sup>3+</sup> (due to the fact that there is no data available for Mn<sup>3+</sup> singlets<sup>68</sup>), and sextet-doublet for Mn<sup>2+</sup> and Fe<sup>3+</sup>.

A set of common ligands in inorganic chemistry was chosen for variability in denticity, rigidity, and size (nine monodentate, six bidentate, and one tetradentate in Fig. 1 and ESI Table S1†). These ligands span the spectrochemical series from weak-field chloride (1, Cl<sup>−</sup>) to strong-field carbonyl (6, CO) along with representative intermediate-field ligands and connecting atoms, including S (2, SCN<sup>−</sup>), N (e.g., 9, NH<sub>3</sub>), and O (e.g., 14, acac). All possible homoleptic structures with all metals/oxidation states were generated from ten of these ligands (90 molecules) using the molSimplify toolkit<sup>67</sup> (ESI Table S2†). Additional heteroleptic complexes (114 molecules) were generated using molSimplify with one mono- or bidentate axial ligand type (L<sub>ax</sub>) and an equatorial ligand type (L<sub>eq</sub>) of compatible denticity (ligands shown in Fig. 1, schematic shown in Fig. 2, geometries provided in the ESI†). We also selected 35 molecules from the Cambridge Structural Database<sup>69</sup> (ESI Table S3†).

**First-principles geometry optimizations.** DFT gas-phase geometry optimizations were carried out using TeraChem.<sup>70,71</sup> DFT calculations employ the B3LYP hybrid functional<sup>72–74</sup> with 20% Hartree–Fock (HF) exchange ( $a_{\text{HF}} = 0.20$ ) and a variant<sup>45</sup> ( $a_{\text{HF}} = 0.00$  to 0.30 in 0.05 increments) that holds the semi-local DFT portion of exchange in a constant ratio. We calculate and predict spin-state splitting sensitivities HF exchange,  $\frac{\partial \Delta E_{\text{H-L}}}{\partial a_{\text{HF}}}$ , as approximated from linear fits, in units of kcal per mol per HFX, where 1 HFX corresponds to varying from 0% to 100% HF exchange. B3LYP<sup>72–74</sup> is chosen here due to its widespread use



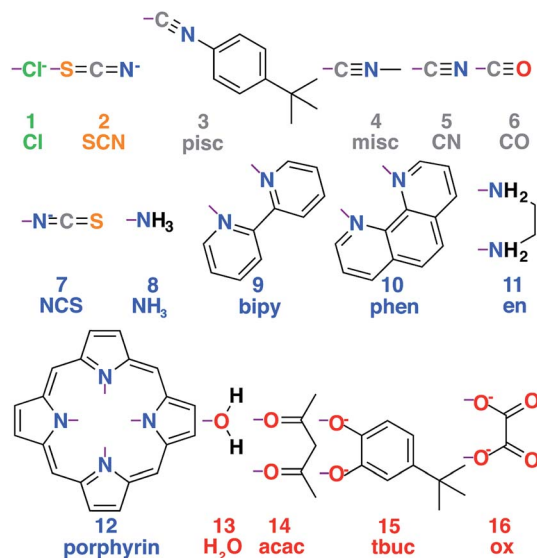


Fig. 1 Set of ligands used to generate the transition metal complex data set. Ligands are numbered 1–16 and colored according to the atom type that coordinates with the metal, with chlorine in green, carbon in gray, sulfur in orange, nitrogen in blue, and oxygen in red. Purple lines indicate the bonds formed to metal-coordinating atoms in the ligand complexes. Abbreviations for each ligand used in the text are also shown. Full chemical names are provided in ESI Table S3.†

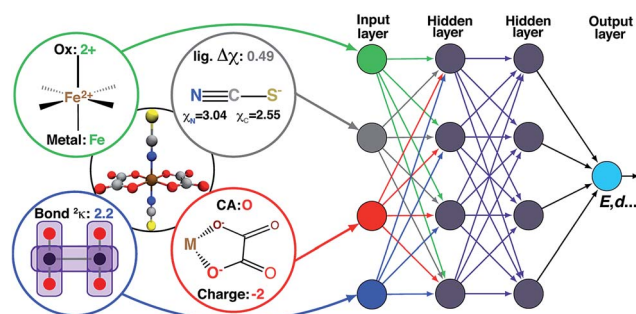


Fig. 2 Schematic diagram of descriptors (left) as inputs to the ANN (right), along with hidden layers, and output (e.g., spin-state splittings). The additive bias term in each node is omitted.

and our prior experience<sup>45</sup> with tuning it to study HF exchange sensitivity, where we observed<sup>45</sup> similar behavior with other GGA hybrids, e.g. PBE0, as long as the same HF exchange fraction was compared.

The composite basis set used consists of the LANL2DZ effective core potential<sup>75</sup> for transition metals and the 6-31G\* basis for the remaining atoms. All calculations are spin-unrestricted with virtual and open-shell orbitals level-shifted<sup>76</sup> by 1.0 and 0.1 eV, respectively, to aid self-consistent field (SCF) convergence to an unrestricted solution.

For all training and test case geometry optimizations, default tolerances of  $10^{-6}$  hartree for SCF energy changes between steps and a maximum gradient of  $4.5 \times 10^{-4}$  hartree per bohr were employed, as implemented in the DL-FIND interface<sup>77</sup> with TeraChem (ESI Table S4†). Entropic and solvent effects that enable comparison to experimental spin-state splittings have

been omitted, and we instead evaluate the DFT adiabatic electronic spin state splitting, as in previous work because our goal is to predict DFT properties and sensitivity to functional choice.<sup>45,78</sup> In high-throughput screening efforts ongoing in our lab, entropic and solvent effects that influence catalytic and redox properties will be considered explicitly.

For each molecular structure (90 homoleptic, 114 heteroleptic) 14 geometry optimizations were carried out at 7 exchange fractions (from 0.00 to 0.30) and in high- or low-spin, for a theoretical maximum of 2856 geometry optimizations. In practice, 166 structures were excluded due to (i) large spin contamination, as defined by an expectation value of  $\langle \hat{S}^2 \rangle$  that deviated more than  $1 \mu_B$  from the exact value ( $<1\%$ , 26 of 2856, see ESI Table S5†), (ii) dissociation in one or both spin states, especially of negatively charged ligands, leading to loss of octahedral coordination (4%, 126 of 2856, see ESI Table S6†), or (iii) challenges associated with obtaining a stable minimized geometry ( $<1\%$ , 14 of 2856, see ESI Table S2†). Eliminating these cases produced a final data set of 2690 geometry optimizations (structures and energetics provided in ESI, as outlined in ESI Text S1†). Although these excluded cases are a fraction of our original data set, they highlight considerations for application of the ANN in high-throughput screening: highly negatively charged complexes should be avoided, and single point DFT calculations should be used to confirm that a high-fitness complex does not suffer from large  $\langle \hat{S}^2 \rangle$  deviations.

## 2.2 Descriptor selection

High-throughput screening of transition-metal complex properties with direct prediction from an ANN requires mapping of an empirical feature space that represents the complex,  $\chi$ , to quantum-mechanical predictions. This feature space should be balanced to avoid (i) too few descriptors with insufficient predictive capability or (ii) too many descriptors that lead to over-fitting of the ANN. Molecular descriptors<sup>79</sup> that have been used for parameterizing chemical space include: atomic composition, electronegativity,<sup>37</sup> formal charges, and representations of the geometric structure. This last class of descriptors may be divided into those that depend either on 3D structural information<sup>13,20,80–82</sup> or on graph-theoretic connectivity maps<sup>83</sup> (e.g., the Randić,<sup>84</sup> Wiener shape,<sup>85</sup> or Kier<sup>86</sup> indices). Graph-theoretic methods are preferable to 3D structural information to avoid sensitivity to translation/rotation or molecule size,<sup>87</sup> though we note that subsystem descriptors<sup>13,82,88</sup> and element-specific pairwise potentials<sup>81,87</sup> have been employed successfully to overcome some challenges. A secondary reason to avoid use of 3D structural information is the implicit requirement of equilibrium geometries obtained from a geometry optimization, which are readily achieved with semi-empirical methods on small organic molecules<sup>83</sup> but would be prohibitive and error-prone for transition metal complexes.

To evaluate candidate descriptor sets, we use L1-regularized, least absolute shrinkage and selection operator (LASSO) linear least-squares regression,<sup>89</sup> as implemented in the glmnet<sup>90</sup> package in R3.2.5.<sup>91</sup> LASSO is used to reduce over-fitting, force



the coefficients of the least-powerful indicators to zero, and avoid monotonic decrease of model error as feature space dimensionality increases. Given observed input-output pairs  $(\mathbf{x}_i, \mathbf{y}_i)$  for  $i = 1, \dots, n$  with  $\mathbf{x} \in \chi \subset \mathbb{R}_i^m$  and  $\lambda \in \mathbb{R}$ , the output is modeled as:

$$\tilde{\mathbf{y}} = \beta^T \mathbf{X} + \beta_0 \mathbf{1} \quad (1)$$

for  $\beta, \beta_0 \in \mathbb{R}^m \times \mathbb{R}$ , where:

$$\{\beta, \beta_0\} = \operatorname{argmin}_{\{\beta, \beta_0\}} \left( \|\mathbf{y} - \beta^T \mathbf{X} - \beta_0 \mathbf{1}\|_2^2 + \lambda \left( \sum_{i=1}^m |\beta_i| \right) \right) \quad (2)$$

The parameter  $\lambda$  is selected by ten-fold cross-validation with values typically between  $10^{-1}$  and  $10^{-6}$ . Our descriptors include both continuous variables that are normalized and discrete variables that are described by zero-one binary coding (ESI Table S7†). Metal identity is best described by a set of discrete variables: 4 binary variables are chosen to represent Cr, Mn, Fe, and Ni, and Co corresponds to the case where all 4 variables are zero. This leads to a higher number of overall variables than for continuous descriptors (see Table 1).

Based on our previous studies of transition metal complexes,<sup>45,49</sup> we expect that spin-state ordering is predominantly determined by the immediate chemical environment around the metal center, potentially enabling predictive descriptors that are widely transferable across a range of molecule sizes. We compare 7 descriptor sets on the data and select the subset of descriptors that give the best simultaneous predictive performance for spin-state splitting,  $\Delta E_{\text{H-L}}$ , and its sensitivity with respect to HF exchange variation,  $\frac{\partial \Delta E_{\text{H-L}}}{\partial a_{\text{HF}}}$ , as indicated by the prediction root mean squared error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{i,\text{pred.}} - y_{i,\text{actual}})^2} \quad (3)$$

When two variable sets perform comparably, we select the variable set that will enable broader application of the ANN. All

**Table 1** Comparison of variable sets by root-mean-squared errors (RMSE) after regularization in  $\Delta E_{\text{H-L}}$  and  $\frac{\partial \Delta E_{\text{H-L}}}{\partial a_{\text{HF}}}$  prediction along with number of discrete variables (with all binary levels of the discrete variables counted in parentheses) and the number of continuous variables

Set	RMSE ( $\Delta E_{\text{H-L}}$ ) (kcal mol <sup>-1</sup> )	RMSE $\frac{\partial \Delta E_{\text{H-L}}}{\partial a_{\text{HF}}}$ (kcal per mol per HFX)	Discrete variables	Continuous variables
a	14.6	20.6	3 (37)	6
b	15.1	21.7	3 (15)	8
c	15.2	21.2	3 (15)	11
d	15.1	21.3	3 (15)	10
e	14.9	21.1	3 (15)	12
f	15.1	23.5	3 (15)	10
g	14.9	21.3	3 (15)	12

Property	Variable set						
	a	b	c	d	e	f	g
<b>Complex-based</b>							
Metal identity							
Oxidation state							
$a_{\text{HF}}$							
sum( $\Delta\chi$ )							
min( $\Delta\chi$ )							
max( $\Delta\chi$ )							
<b>Ligand-based</b>							
Identity							
Connection atom							
Charge							
Denticity							
Number of atoms							
Bond order							
Truncated Kier index							

**Fig. 3** Summary of variables chosen for each set a through g. Employed variables are indicated in shaded gray and grouped by whether they are assessed on the whole complex (complex-based) or on each individual axial or equatorial ligand (ligand-based).  $\Delta\chi$  is the difference in Pauling electronegativity between the ligand connecting atom and all atoms bonded to it, and the sum, maximum or minimum values are obtained over all ligands.

sets include the metal identity as a discrete variable and metal oxidation state, ligand formal charge, and ligand denticity as continuous variables (Fig. 3, some descriptors shown in Fig. 2). Set a represents our most specific model, where we explicitly code the full axial or equatorial ligand identity as a discrete variable, limiting the application of the model but producing one of the lowest RMSEs for  $\Delta E_{\text{H-L}}$  and  $\frac{\partial \Delta E_{\text{H-L}}}{\partial a_{\text{HF}}}$  (Table 1). Elimination of ligand identity in favor of ligand connecting atom elemental identity and total number of atoms in set b increases  $\Delta E_{\text{H-L}}$  RMSE slightly and decreases  $\frac{\partial \Delta E_{\text{H-L}}}{\partial a_{\text{HF}}}$  RMSE (see Table 1).

The shift from set a to b increases the model applicability but at the cost of omitting subtler ligand effects. For instance, ethylenediamine (**11**, en) and phenanthroline (**10**, phen) have the same ligand charge/denticity and direct ligand atom (N), making them equivalent in set b except for the larger size of phen. System size alone is not expected to be a good predictor of field strength (e.g., the small CO is one of the strongest field ligands). In set c, we introduce properties that depend on the empirical pairwise Pauling electronegativity difference ( $\Delta\chi$ ) between the ligand connecting atom (LC) and any  $i$ th atom connected (CA) to it:

$$\Delta\chi_{\text{LC},i} = \chi_{\text{LC}} - \chi_i \quad (4)$$

These whole-complex differences include the maximum, max( $\Delta\chi$ ), and minimum, min( $\Delta\chi$ ), as well as sum:





$$\text{sum}(\Delta\chi) = \sum_{j \in \text{lig.}} \sum_{i \in \text{CA}} \Delta\chi_{\text{LC},i}^j \quad (5)$$

which is taken over the direct ligand atom and all atoms bonded to it for all ligands (lig.) in the complex. These additional set c descriptors reduce  $\Delta E_{\text{H-L}}$  RMSE slightly and decrease the  $\frac{\partial \Delta E_{\text{H-L}}}{\partial a_{\text{HF}}}$  RMSE to its lowest value (see Table 1). In set d, we eliminate  $\min(\Delta\chi)$ , expecting it to be redundant with the max and sum, at the cost of a small increase in both RMSEs.

Finally, in sets e–g, we replace ligand size (*i.e.*, number of atoms) with general descriptors to enable prediction on molecules larger than those in any training set. For example, tetraphenylporphyrin will have comparable electronic properties to unfunctionalized porphyrin (**12**), despite a substantial size increase. In set e, we introduce the maximum bond order of the ligand connecting atom to any of its nearest neighbors, a measure of the rigidity of the ligand environment, which is zero if the ligand is atomic (see ESI Table S1†). In set f, we eliminate the number of atoms and bond order metric, replacing them with a broader measure of the ligand geometry adjacent to the metal. After trial and error, we have selected the truncated Kier shape index,<sup>86</sup>  $^2K$ , which is defined by the inverse ratio of the square of number unique paths of length two ( $^2P$ ) in the molecular graph of heavy atoms to the theoretical maximum and minimum for a linear alkane with the same number of atoms:

$$^2K = \frac{2^2 P_{\text{max}}^2 P_{\text{min}}}{(^2P)^2} \quad (6)$$

and set to zero for any molecules that do not have paths of length two. The truncation means that only the ligand atoms within three bonds of the connecting atom are included in the graph. The set f MSEs are comparable to or a slight increase from sets with molecule size, but they beneficially eliminate system size dependence. In set g, we reintroduce the bond order metric as well, providing the lowest MSEs except for set a or c,

both of which are much less transferable than set g. Thus, the comparable performance of set g to a full ligand descriptor (set a) supports our hypothesis that a combination of metal-centric and ligand-centric in a heuristic descriptor set can be predictive and transferable. This final feature space is 15-dimensional with five per-complex descriptors and five per-ligand descriptors for each equatorial or axial ligand (see Table 2 for ranges of values and descriptions). A comparison of all errors and weights of variables across the seven data sets is provided in ESI Tables S7–S21 and Fig. S1.†

### 2.3 Training and uncertainty quantification of ML models

ANNs enable complex mapping of inputs to outputs<sup>92</sup> beyond multiple linear regression and support the use of both discrete (*i.e.*, binary choices such as metal identity) and continuous (*e.g.*, the % of HF exchange) variables. Here, we apply an ANN with an input layer, two intermediate hidden layers, and an output layer (Fig. 2). The network topology was determined by trial and error, with additional hidden layers yielding no improved performance. All analysis is conducted in R3.2.5,<sup>91</sup> using the h2o<sup>93</sup> package with tanh non-linearity and linear output. Network weights and full training and test data are provided in the ESI.†

As with many ML models, ANNs are sensitive to over-fitting due to the number of weights to be trained.<sup>94</sup> We address overfitting using dropout,<sup>95,96</sup> wherein robustness of the fit is improved by zeroing out nodes in the network with an equal probability,  $p_{\text{drop}}$ , at each stage of training (5% for spin-state splitting, 15% for HF exchange sensitivity, and 30% for bond lengths, selected by trial and error). Dropout has been shown to address overfitting when training feedforward ANNs on small datasets,<sup>96</sup> with larger values of  $p_{\text{drop}}$  giving more aggressive regularization that worsens training errors but improves test errors. We use L2 weight regularization with a fixed penalty weight  $\lambda$ , as is applied in standard ridge regression, with an effective loss function for training:

$$\{\mathbf{W}\} = \text{argmin}_{\{\mathbf{W}\}} \left( \sum_{i=1}^N (y_i - \tilde{y}(\mathbf{x}_i))^2 + \lambda \sum_{l=1}^L (\|\mathbf{W}_l\|_2^2 + \|\mathbf{b}_l\|_2^2) \right) \quad (7)$$

here,  $\mathbf{W}_l$  refers to the weights from layer  $l$  to  $l+1$ ,  $\mathbf{b}_l$  are biases at layer  $l$ ,  $\tilde{y}(\mathbf{x}_i)$  is the ANN prediction for the input–output pair  $(\mathbf{x}_i, y_i)$ , and the sums run over  $N$  training pairs and  $L$  layers.

During network training, we randomize the order of data points and partition the first 60% as training data and the last 40% for testing. Dropout networks, consisting of two hidden layers of 50 nodes each, are trained on the data set for varying values of  $\lambda$  ranging from  $10^{-1}$  to  $10^{-6}$  using 10-fold cross validation. For each  $\lambda$ , the training data is partitioned into ten groups, and a network is trained on each combination of nine of the groups and scored based on eqn (7) on the left-out group. The parameter with the lowest average prediction error is used to select the best regularization parameter:  $5 \times 10^{-4}$  for spin-state splitting,  $10^{-2}$  for HF exchange sensitivity, and  $10^{-3}$  for bond lengths. We varied and optimized<sup>97</sup> the learning rate between 0.05 and 1.5, and optimal rates were selected as 1.0 (bond lengths) and 1.5

**Table 2** Optimal (set g) input space descriptors and their range in the training set.  $\Delta\chi$  is the difference in Pauling electronegativity between the ligand connecting atom and all atoms bonded to it. When training the ANNs, a continuous descriptor corresponds to a single input node, whereas discrete descriptors correspond to one node per level

Symbol	Type	Descriptor	Values or range
<b>Whole-complex descriptors</b>			
M	Discrete	Metal identity	Cr, Mn, Fe, Co, Ni
O	Continuous	Oxidation state	2 to 3
me	Continuous	Max. $\Delta\chi$ over all ligands	−0.89 to 1.20
se	Continuous	Sum of $\Delta\chi$ over all ligands	−5.30 to 7.20
$a_{\text{HF}}$	Continuous	HF exchange fraction	0.00 to 0.30
<b>Ligand-specific descriptors</b>			
L	Discrete	Ligand connection atom	Cl, S, C, N, or O
C	Continuous	Ligand charge	0 to −2
k	Continuous	Truncated Kier index	0.00 to 6.95
b	Continuous	Ligand bond order	0 to 3
D	Continuous	Ligand denticity	1 to 4



(spin-state splitting or HF exchange sensitivity). We use batch optimization for training (batch size = 20) for 2000 epochs. The training algorithm minimizes eqn (7) over the training data using stochastic gradient descent.<sup>97–100</sup>

It has not been possible to estimate ANN model uncertainty<sup>95,101</sup> with the possible exception of bootstrapping<sup>102</sup> by training the ANN on numerous subsamples of available training data. Model uncertainty will be due to either high-sensitivity to descriptor changes or test molecule distance in chemical space to training data (see also Section 3). Recent work<sup>94</sup> showed that minimization of the loss function in eqn (7) is equivalent to approximate variational optimization of a Gaussian process (GP), making previously suggested ANN sampling for different dropout realizations<sup>95</sup> a rigorously justified<sup>94</sup> model uncertainty estimate.

We sample  $J$  distinct networks (in this work,  $J = 100$ ) with different nodes dropped at the optimized weights and average over the predictions:

$$\tilde{y}(\mathbf{x}_n) = \frac{1}{J} \sum_{j=1}^J \tilde{y}_j(\mathbf{x}_n) \quad (8)$$

The ANN predictive variance is:<sup>94</sup>

$$\text{var}(\tilde{y}(\mathbf{x}_n)) \approx \tau^{-1} \mathbf{I} + \frac{1}{J} \sum_{j=1}^J (\mathbf{y}_j(\mathbf{x}_n)^T \mathbf{y}_j(\mathbf{x}_n) - \tilde{y}(\mathbf{x}_n)^T \tilde{y}(\mathbf{x}_n)) \quad (9)$$

here,  $\tau$  is

$$\tau = \frac{(1 - p_{\text{drop}})l^2}{2N\lambda} \quad (10)$$

where  $N$  is the number of training data points, and  $l$  is a model hyperparameter for the GP that affects the estimation of predictive variance but does not enter into the ANN training. The contribution of  $\tau$  in eqn (9) is a baseline variance inherent in the data, whereas the second term represents the variability of the GP itself. We obtain  $\tau$  values of 0.6 for spin-state splitting, 0.07 for HF exchange sensitivity, and 10 000 for bond lengths (see Section 3). We choose  $l$  by maximizing the log predictive likelihood of the corresponding GP based on the training data (details are provided in the ESI Text S2†).

We selected an ANN based on the successful demonstrations<sup>11,14,103</sup> of ANN-based models for predicting quantum chemical properties but also provide a comparison to two other common machine learning models:<sup>89</sup> kernel ridge regression (KRR) and a support vector regression model (SVR), both using a square-exponential kernel. We used the R package kernlab<sup>104</sup> and selected hyperparameters (the width of the kernel, and the magnitude of the regularization parameters which are given in the ESI Table S22†) using a grid search and ten-fold cross-validation using the R package CVST.<sup>105</sup> We also compared training on our descriptor set to a KRR model with a kernel based on the L1 distance between sorted Coulomb matrix representations,<sup>87</sup> as demonstrated previously.<sup>52,103</sup>

## 3. Results and discussion

### 3.1 Overview of data set spin-state energetics

Analysis of the qualitative and quantitative features of the spin-state splitting data set motivates the training of an ANN to move beyond ligand field arguments. We visualize qualitative ground states (*i.e.*, high-spin or low-spin) for the homoleptic subset of the data using a recursive binary tree (Fig. 4, descriptor definitions provided in Table 2), as previously outlined<sup>106</sup> and implemented in the open source rpart package<sup>107</sup> for R3.2.5.<sup>91</sup> A recursive binary tree is a list of “branches” of the data ordered by statistical significance that gives the most homogeneous final “leaves” (here, with at least 10 data points) after a given number of permitted divisions (here, 6). Using descriptor set  $g$ , the data are partitioned into branches by testing which descriptors provide the “best” division to produce majority high- or low-spin states in leaves based on the concept of information impurity<sup>107</sup> and pruning to remove statistically insignificant branches. The resulting electronic structure spectrochemical “tree” simultaneously addresses metal-specific strengths of ligands and exchange–correlation sensitivity. As expected, strong field direct carbon ligands (no Cl, N, O or S in Fig. 4) provide the root division of the tree, producing low-spin ground states for 92% of all Mn, Fe, and Co complexes (far right box on the third tier in Fig. 4). Next level divisions include the M(II) oxidation state for  $a_{\text{HF}} > 0.05$  that are predominantly (96%) high-spin. Spin-state ordering is well-known<sup>45,59</sup> to be sensitive to HF exchange, and the tree reveals Mn<sup>3+</sup> with nitrogen ligands to have the strongest  $a_{\text{HF}}$  dependence, since they are 69% high-spin for  $a_{\text{HF}} > 0.1$  but 90% low-spin for  $a_{\text{HF}} \leq 0.1$ . Extension of the recursive binary tree to heteroleptic compounds produces a second-level division based on  $\text{sum}(\Delta\chi)$ , validating the relevance of the identified electronegativity descriptors for predicting heteroleptic spin-state ordering (ESI Fig. S2†).

Quantitatively, the maximum  $\Delta E_{\text{H-L}}$  in the data set is 90.7 kcal mol<sup>−1</sup> for the strong-field Co(III)(misc)<sub>6</sub> complex at  $a_{\text{HF}} = 0.00$ , and the minimum value is −54.2 kcal mol<sup>−1</sup> for the weak-field Mn(II)(NCS<sup>−</sup>)<sub>6</sub> at  $a_{\text{HF}} = 0.30$ . These extrema are consistent with (i) the ordering of metals in the spectrochemical series<sup>43</sup> and (ii) the uniform effect of stabilizing high-spin states with increasing HF exchange. By comparing compound trends in the data set, we are able to identify whether additivity in ligand field effects, which has been leveraged previously in heuristic DFT correction models,<sup>108–110</sup> is a universally good assumption. For the Fe(III)(Cl<sup>−</sup>)<sub>6−*n*}(pisc)<sub>*n*</sub> complexes (denoted 1-1 through 3-3 in Fig. 5), increasing  $n$  from 0 to 2 through the addition of two axial pisc ligands increases the spin-state splitting by 15.1 kcal mol<sup>−1</sup> per replaced chloride. Transitioning to a complex with all equatorial pisc ligands ( $n = 4$ ) increases the spin-state splitting by only 10.4 kcal mol<sup>−1</sup> per additional ligand, and the homoleptic structure pisc ( $n = 6$ ) only adds 7.5 kcal mol<sup>−1</sup> per additional ligand beyond the  $n = 4$  case. An additive model cannot precisely reproduce diminishing ligand effects. As a stronger example for the need for nonlinear models such as an ANN, replacing two axial ligands from the strong-field Mn(II)(CO)<sub>6</sub> complex with the weaker-field NCS<sup>−</sup> (6-6 and 6-7 in ESI Fig. S3†)</sub>



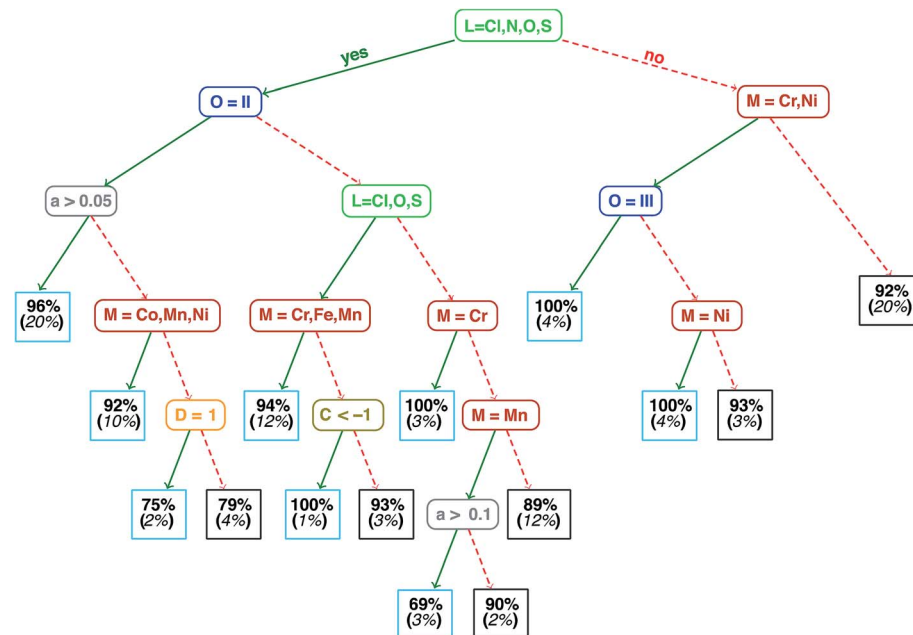


Fig. 4 Binary ground state classification tree for homoleptic compounds. M indicates metal identity, L ligand connection atom, O oxidation state, a the fraction of HF exchange, C the charge, and D the ligand denticity. Each leaf node indicates the percent of elements in that leaf (light blue boxes for high-spin and dark gray boxes for low-spin) in bold font and percentage of total homoleptic population in the node (italic font, in parentheses).

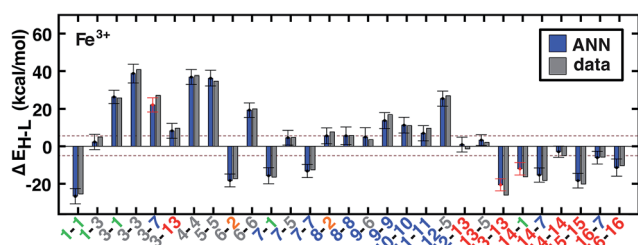


Fig. 5 ANN model predicted (ANN, blue bars) and computed (data, gray bars) spin-state splittings,  $\Delta E_{H-L}$ , for the B3LYP functional ( $a_{HF} = 0.20$ ) in  $\text{kcal mol}^{-1}$ . Complexes are labeled by equatorial and then axial ligands according to the numbering indicated in Fig. 1 and color-coded by direct ligand atom (green for chlorine, gray for carbon, blue for nitrogen, red for oxygen, and orange for sulfur). The error bars represent an estimated  $\pm 1$  standard deviation credible interval from the mean prediction, and error bars that do not encompass the computed value are highlighted in red. Brown dashed lines indicate a  $\pm 5 \text{ kcal mol}^{-1}$  range around zero  $\Delta E_{H-L}$ , corresponding to near-degenerate spin states.

alters  $\Delta E_{H-L}$  by  $< 1 \text{ kcal mol}^{-1}$ , as strong-field ligands (e.g., CO,  $\text{CN}^-$ ) have an overriding effect on spin-state splitting.

### 3.2 Spin-state splittings from an ANN

Motivated by non-linear effects in ligand additivity, we trained an ANN using a heuristic descriptor set (see Section 2.2) to predict qualitative spin-state ordering and quantitative spin-state splitting. The ANN predicts the correct ground state in 98% of the test cases (528 of 538) and 96% of training cases (777 of 807). All of the misclassifications are for cases in which DFT

$\Delta E_{H-L}$  is  $< \pm 5 \text{ kcal mol}^{-1}$  (ESI Table S23†). The ANN spin-state prediction errors are not sensitive to HF exchange mixing, and thus our trained ANN is able to predict ground states of transition metal complexes from the pure GGA limit to hybrids with moderate exchange.

We assess quantitative performance with root mean squared errors (RMSE) of the ANN (eqn (3)), overall and by metal (Fig. 6, ESI Table S24 and Fig. S3–S6†). The comparable RMSE of 3.0 and 3.1  $\text{kcal mol}^{-1}$  for the training and test data, respectively, indicate an appropriate degree of regularization. The ANN predicts DFT spin-state splittings within 1  $\text{kcal mol}^{-1}$  (i.e., “chemical accuracy”) for 31% (168 of 538) of the test data and within 3  $\text{kcal mol}^{-1}$  (i.e., “transition metal chemical accuracy<sup>111</sup>”) for 72% (389 of 538) of the test data. Only a small subset of 49 (4) test molecules have errors above 5 (10)  $\text{kcal mol}^{-1}$ , and correspond to strong-field Co and Cr complexes, e.g.,  $\text{Cr(II)(NCS)}_2(\text{pisc})_4$  (ESI Fig. S5†). The model is equivalently predictive for homoleptic and heteroleptic compounds at 2.2 and 2.3  $\text{kcal mol}^{-1}$  average unsigned error respectively.

The training and test RMSEs broken down by metal reveal comparable performance across the periodic table (Fig. 6). Slightly higher test RMSEs (maximum unsigned errors) for Co and Fe complexes at 3.8 (15.7) and 3.3 (13.0)  $\text{kcal mol}^{-1}$ , respectively, are due to the train/test partition and more variable ligand dependence of spin-state ordering in these complexes (Fig. 6 and ESI Table S24†). When the ANN performs poorly, the errors are due to both under- and over-estimation of  $\Delta E_{H-L}$  for both strong- and weak-field ligands, regardless of HF exchange fraction: e.g.,  $\Delta E_{H-L}$  for  $\text{Co(III)(CN)}_6$  at  $a_{HF} = 0.00$  and  $\text{Co(III)(en)}_3$  at  $a_{HF} = 0.20$  are overestimated by 14 and 9  $\text{kcal mol}^{-1}$ , respectively, but  $\Delta E_{H-L}$  for  $\text{Fe(III)(Cl)}_6$  at  $a_{HF} = 0.10$  and



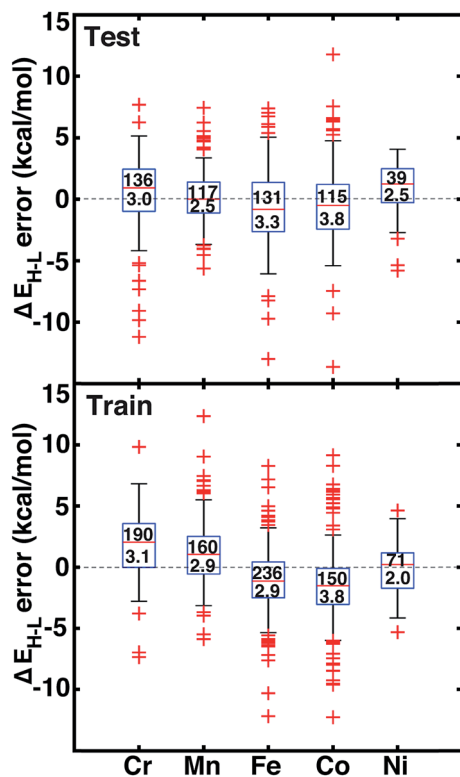


Fig. 6 Error boxplots for  $\Delta E_{H-L}$  in  $\text{kcal mol}^{-1}$  using the ANN for test (top) and training (bottom) data partitioned by metal identity. The top number inside the box indicates the number of cases in each set, and the bottom number indicates the RMSE in  $\text{kcal mol}^{-1}$ . The range for both graphs is from  $15 \text{ kcal mol}^{-1}$  to  $-15 \text{ kcal mol}^{-1}$ .

$\text{Co(II)}(\text{H}_2\text{O})_2(\text{CN}^-)_4$  at  $a_{\text{HF}} = 0.30$  and are underestimated by 9 and 7  $\text{kcal mol}^{-1}$ , respectively.

Quantified uncertainty estimates correspond to a baseline standard deviation in the model of approximately  $1.5 \text{ kcal mol}^{-1}$  ( $\sqrt{\tau^{-1}}$ ) and a mean total estimated standard deviation across the training and test cases of 3.8 and 3.9  $\text{kcal mol}^{-1}$ , respectively (see Section 2.3 and error bars on Fig. 5). These credible intervals are not rigorously confidence intervals but can highlight when prediction uncertainty is high: a  $\pm 1$  ( $\pm 2$ ) standard deviation (std. dev.) interval on ANN predictions captures 83% (98%) of computed values for test set (see ESI Fig. S7†). Highest std. dev. values of around 5  $\text{kcal mol}^{-1}$  are observed for Fe(II) and Mn(II) complexes and the lowest are around 3 for Cr and Co complexes (see ESI†). A single std. dev. around the ANN prediction contains the calculated  $\Delta E_{H-L}$  for 26 of 29 Fe(III) complexes at  $a_{\text{HF}} = 0.20$  but misses heteroleptic oxygen coordinating complexes, 13-13 and 14-1, and underestimates the effect of C/N ligands in 3-7 (Fig. 5). The model performs consistently across different ligand sizes, from porphyrin Fe(III) complexes (12-13, 12-5) to Fe(II)(NH<sub>3</sub>)<sub>6</sub> and Fe(II)(CO)<sub>6</sub> (6-6 and 8-8). For ligand-specific effects, the ANN performs well, reversing splitting magnitude as equatorial and axial ligands are swapped (e.g., 1-3 versus 3-1).

Review of other metals/oxidation states reveals comparable performance for cases where the high-spin state is always favored (e.g., Mn(II), Cr(III), or Ni(II)), low-spin state is always

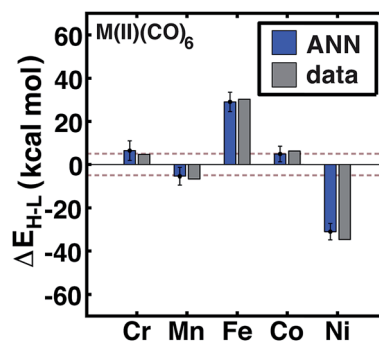


Fig. 7 ANN model predicted (ANN, blue bars) and computed (data, gray bars) spin-state splittings,  $\Delta E_{H-L}$ , with the B3LYP functional ( $a_{\text{HF}} = 0.20$ ) in  $\text{kcal mol}^{-1}$  on  $\text{M(II)}(\text{CO})_6$  complexes, where  $\text{M} = \text{Cr}, \text{Mn}, \text{Fe}, \text{Co}$ , or  $\text{Ni}$ . The error bars represent an estimated  $\pm 1$  standard deviation credible interval from the mean prediction, and brown dashed lines indicate a  $\pm 5 \text{ kcal mol}^{-1}$  range around zero  $\Delta E_{H-L}$ , corresponding to near-degenerate spin states.

favored (e.g., Cr(III)), and those where ligands have strong influence over the favored spin state (e.g., Fe(II) and Cr(II)) (see ESI Fig. S3–S6†). For instance, metal-specific effects examined through comparison of  $\text{M(II)}(\text{CO})_6$  complexes (Fig. 7) reveal good ANN performance both for where the strong-field ligand strongly favors the low-spin state (i.e., Fe and Ni) and where the spin-states are nearly degenerate (i.e., Cr, Mn, Co). The trends outlined here for 20% HF exchange hold at other exchange mixing values (ESI Table S23†). Thus, our ANN trained on a modest data set with heuristic descriptors predicts spin-state splitting within a few  $\text{kcal mol}^{-1}$  of the DFT result.

Comparing our results to KRR, SVR, and LASSO regression reinforces the choice of an ANN (Table 3 and ESI Fig. S8†). The ANN outperforms KRR with either our descriptor set or the sorted Coulomb matrix descriptor both on the full data set or at fixed HF exchange (ESI Text S3†). The ANN also performs slightly better than SVR on test data with our descriptors. Linear LASSO regression was employed for feature selection (Section 2.2) but is outperformed by all other methods (Table 3). We will revisit the performance of these models on a more diverse molecule test set in Section 3.5 to assess the question of transferability.

### 3.3 Predicting exchange sensitivity with an ANN

Spin-state splittings exhibit high sensitivity to exchange<sup>45,59</sup> with linear behavior that we previously identified<sup>45</sup> to be strongly dependent on direct ligand identity and field strength when we compared a set of Fe complexes. Over the current data set, computed exchange sensitivities are indeed linear, ranging from  $-174 \text{ kcal per mol per HFX}$  for strong-field Fe(II)(CO)<sub>6</sub> to  $-13 \text{ kcal per mol per HFX}$  for weak-field Cr(III)(en)<sub>2</sub>(NH<sub>3</sub>)<sub>2</sub>. Cr(III) is the least exchange-sensitive metal in our test set, whereas Fe(II) and Mn(II) are the most sensitive (ESI Table S25 and Fig. S9†).

We therefore generalize previous observations<sup>45</sup> in an ANN that predicts HF exchange sensitivity of spin-state ordering,  $\frac{\partial \Delta E_{H-L}}{\partial a_{\text{HF}}}$ , using the same descriptors as for direct spin-state





**Table 3** Train/test data and CSD test set RMSEs and max UEs for  $\Delta E_{H-L}$  in kcal mol<sup>-1</sup> for different machine learning methods and descriptor sets compared: KRR, kernel ridge regression, using square-exponential kernel for descriptor set g and the L1 matrix distance<sup>52</sup> for the sorted Coulomb matrix descriptor; SVR, support vector regression using square-exponential kernel; ANN, artificial neural network. Results are also given for the KRR/Coulomb case, restricted to B3LYP only since the Coulomb matrix does not naturally account for varying HF exchange

Model	Descriptor	Training		Test		CSD	
		RMSE	Max UE	RMSE	Max UE	RMSE	Max UE
LASSO	Set g	16.1	89.7	15.7	93.5	19.2	72.5
KRR	Set g	1.6	8.5	3.9	17.0	38.3	88.4
SVR	Set g	2.1	20.9	3.6	20.4	20.3	64.8
ANN	Set g	3.0	12.3	3.1	15.6	13.1	30.4
KRR	Sorted Coulomb	4.3	41.5	30.8	103.7	54.5	123.9
KRR, B3LYP only	Sorted Coulomb	17.2	58.0	28.1	69.5	46.7	118.7

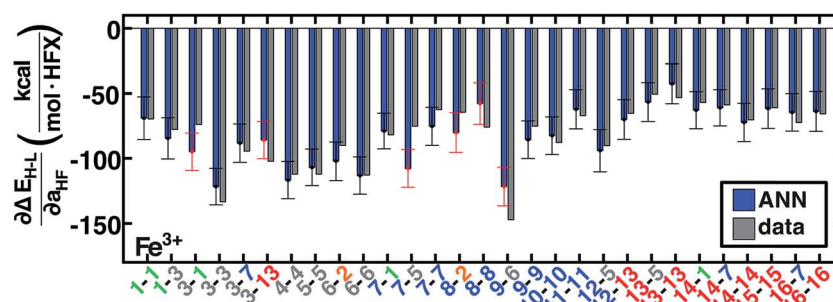
**Table 4** Test set RMSEs for  $\frac{\partial \Delta E_{H-L}}{\partial a_{HF}}$  in kcal per mol per HFX separated by metal and oxidation state along with minimum and maximum unsigned test errors (UE). The number of test cases is indicated in parentheses

Species	RMSE	Min. UE	Max. UE
Cr(II)	21 (14)	4	45
Cr(III)	17 (8)	2	37
Mn(II)	24 (6)	3	40
Mn(III)	38 (8)	4	92
Fe(II)	18 (9)	2	41
Fe(III)	15 (12)	<1	32
Co(II)	17 (8)	<1	26
Co(III)	20 (8)	<1	46
Ni(II)	9 (4)	1	15

splitting, excluding only  $a_{HF}$ . The smaller size of this data set (1/7 the size of the  $\Delta E_{H-L}$  data set) leads to overfitting, with lower RMSE values of 13 kcal per mol per HFX for the training data versus 22 kcal per mol per HFX for the test set (Table 4, ESI Fig. S10 and Table S26†). Although results are reported in units of HFX (from 0 to 100% exchange), for typical 20% variation in exchange, a 20 kcal per mol per HFX sensitivity error only corresponds to a 4 kcal mol<sup>-1</sup> energy difference. Both maximum

unsigned errors (UE) and RMSEs are largest for Mn(III) and Cr(II) complexes, with the largest case producing an 92 kcal per mol per HFX underprediction for Mn(III)(H<sub>2</sub>O)<sub>4</sub>(pisc)<sub>2</sub>. Overall, the ANN prediction errors are less than less than 20 (40) kcal per mol per HFX for 65% (95%) of the test data. The ANN provides a valuable strategy for predicting exchange sensitivity, reproducing non-monotonic and nonconvex ligand sensitivity in heteroleptic compounds: a Fe(III) complex with ox, 16, and NCS<sup>-</sup>, 7, ligands is more sensitive to HFX than the respective homoleptic complexes (Fig. 8, other metals in ESI Fig. S11–S14†).

Uncertainty intervals of ANN predictions for HFX sensitivity yield a narrow range from 14 kcal per mol per HFX to 17 kcal per mol per HFX. For the 29 Fe(III) complexes studied, 23 (80%) of the ANN credible intervals span the computed exchange sensitivity (Fig. 8). Across the full metal and oxidation state data set, 70% (83%) of the computed data is contained by  $\pm 1$  ( $\pm 2$ ) std. dev. intervals (Fig. 8 and ESI S15†). This performance can be further improved by extending the training data. Exchange-sensitivity provides value both for extrapolation of computed (see Section 3.6) or literature values obtained at an arbitrary exchange mixing and in identification of cases of high-sensitivity to DFT functional choice.



**Fig. 8** ANN model predicted (ANN, blue bars) and computed (data, gray bars) spin-state splitting sensitivities to HF exchange,  $\frac{\partial \Delta E_{H-L}}{\partial a_{HF}}$ , in kcal per mol per HFX, for Fe<sup>3+</sup> complexes. Complexes are labeled as equatorial and then axial ligands according to the numbering indicated in Fig. 1 and color-coded by direct ligand atom (green for chlorine, gray for carbon, blue for nitrogen, red for oxygen, and orange for sulfur). The error bars represent an estimated  $\pm 1$  standard deviation credible interval from the mean prediction, and error bars that do not encompass the computed value are highlighted in red.



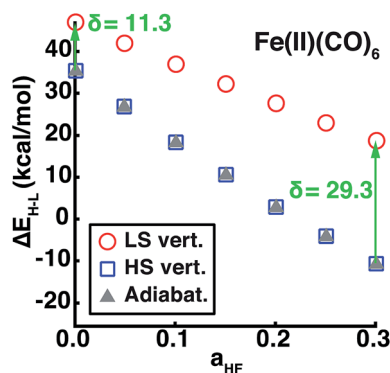


Fig. 9 The vertical or adiabatic spin-state splittings,  $\Delta E_{H-L}$ , in kcal mol<sup>-1</sup> as a function of HF exchange,  $a_{HF}$ , for Fe(II)(CO)<sub>6</sub>. Spin-state splittings evaluated at the HS or LS geometries are indicated by open blue squares and open red circles, respectively. The adiabatic spin-state splitting is shown as filled gray triangles. The HS vertical and adiabatic splittings overlap, whereas the LS vertical overestimates  $\Delta E_{H-L}$ , as indicated by the green arrow and annotated  $\delta$  in kcal mol<sup>-1</sup> for  $a_{HF} = 0.00$  and  $a_{HF} = 0.30$ .

### 3.4 Predicting equilibrium geometries with an ANN

Using our descriptor set, we trained an ANN on the minimum metal–ligand bond distances for both low-spin and high-spin geometries ( $\min(R_{LS/HS})$ ), which only differ from the exact metal–ligand bond length for distorted or heteroleptic compounds. This ANN for bond length prediction extends capabilities we have recently introduced for generating high-quality transition metal complex geometries<sup>67</sup> in order to enable spin-state dependent predictions without requiring extended geometry-optimization. Furthermore, comparison of adiabatic and vertical spin-state splittings computed either at the low- or high-spin optimized geometries reveals that the vertical splitting at the HS geometry is indistinguishable from the adiabatic splitting, but the LS geometry vertical splitting favors the LS state by 10–30 kcal mol<sup>-1</sup>, increasing with  $a_{HF}$  (Fig. 9). Thus, if the ANN bond length predictions are accurate, adiabatic spin-state splittings can be obtained from DFT single points at ANN-predicted HS-only or both LS/HS geometries.

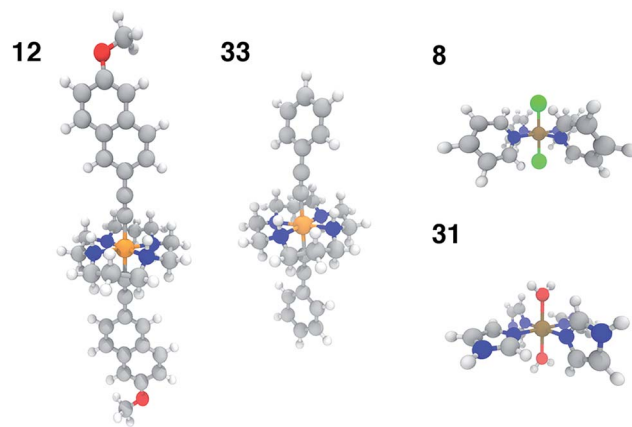


Fig. 11 Representative CSD test set molecules shown in ball and stick representation with carbon atoms in gray, nitrogen atoms in blue, oxygen in red, hydrogen in white, chlorine in green, chromium in orange, and iron in brown. Test molecules 12 (CSD ID: SUMLET) and 33 (CSD ID: YUJCIQ) are Cr(III) cyclams for which the ANN performs least well, and test molecules 8 (CSD ID: TPYFEC04) and 31 (CSD ID: BIP-GEN) are cases for which the ANN predicts  $\Delta E_{H-L}$  within 3 kcal mol<sup>-1</sup>.

Metal–ligand bond distances in the  $a_{HF} = 0.20$  data set vary from  $\min(R_{LS}) = 1.81$  Å (in Fe(II)(pisc)<sub>2</sub>(Cl<sup>-</sup>)<sub>4</sub>) to  $\min(R_{HS}) = 2.55$  Å (in Fe(III)(Cl<sup>-</sup>)<sub>6</sub>). The metal–ligand bond length ANN produces comparable RMSE across training (0.02 Å for LS and HS) and test (0.02 Å for LS and 0.03 Å for HS) data with comparable errors regardless of metal identity and oxidation- or spin-state (ESI Tables S27–S29 and Fig. S16–S27†). ANN bond length std. devs. range from 0.026 to 0.045 Å with a ~0.01 Å baseline contribution. For low-spin (high-spin) complexes, 79% (81%) and 96% (96%) of the calculated values fall within one and two std. dev. of ANN-predicted bond lengths, respectively (ESI Fig. S17 and S23†).

The ANN bond lengths fall outside of computed values for low-spin Fe(III) complexes by more than a full standard deviation in seven cases, *e.g.*, underestimating Fe–C distances in CN (7-5, 13-5) and pisc (3-7, 3-13) complexes (Fig. 10). However, it also reproduces subtle trends, *e.g.* replacing axial ligands in

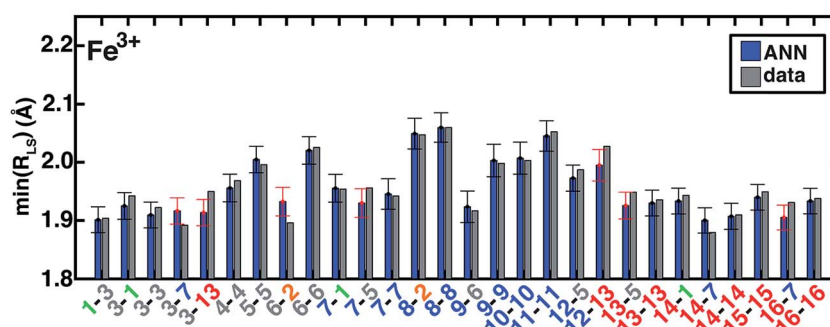


Fig. 10 ANN model predictions (ANN, blue bars) and computed (data, gray bars) minimum LS Fe<sup>3+</sup> bond lengths,  $\min(R_{LS})$ , in Å. Complexes are labeled as equatorial and then axial ligands according to the numbering indicated in Fig. 1 and color-coded by direct ligand atom (green for chlorine, gray for carbon, blue for nitrogen, and red for oxygen). The error bars represent an estimated  $\pm 1$  standard deviation credible interval around the mean prediction, and error bars that do not encompass the computed value are highlighted in red. Fe(III)(Cl)<sub>6</sub> (1-1) is excluded due to being off scale: it has a predicted/calculated bond length of 2.44/2.45 Å, and an error standard deviation of  $\pm 0.02$ .

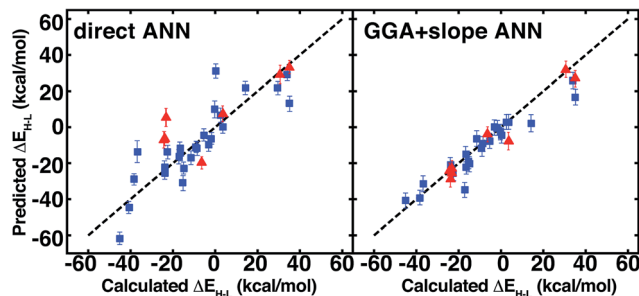


Fig. 12 ANN spin-state splitting energy,  $\Delta E_{H-L}$ , predicted values on CSD test structures vs. DFT-calculated values, both at  $a_{HF} = 0.20$  and in  $\text{kcal mol}^{-1}$ . Direct prediction (left) is compared to GGA calculations and extrapolation using the predicted slope from the ANN (right). Error bars represent a credible interval of one standard deviation from the model uncertainty analysis (either in direct ANN at left or slope ANN at right), and a parity line (black, dashed) is indicated. Cyclams are indicated by red triangles, as described in main text, and the remaining test cases are indicated by blue squares.

homoleptic LS  $\text{Fe(III)(pisc)}_6$  (3-3 in Fig. 10,  $\min(R_{LS}) = 1.92 \text{ \AA}$ ) with  $\text{Cl}^-$  increases the minimum bond distance to  $1.94 \text{ \AA}$  (3-1 in Fig. 10), but replacing equatorial pisc ligands instead with  $\text{Cl}^-$  (1-3 in Fig. 10) decreases the minimum bond distance to  $1.90 \text{ \AA}$ , a feature reproduced by the ANN. Non-additive bond length effects motivate the use of the ANN in initial geometry construction.<sup>67</sup> Indeed, when we use ANN-predicted metal-ligand bond lengths in structure generation instead of our previous strategy based on a discrete database of DFT bond lengths,<sup>67</sup> we reduce the metal-ligand component of the gradient by 54–90% (ESI Text S4, Fig. S28 and Table S30†). The ANN-predicted bond lengths and spin states are now available in molSimplify<sup>67</sup> as an improved tool for structure generation.

### 3.5 Expanding the test set with experimental transition metal complexes

In order to test the broad applicability of the trained ANNs, we selected 35 homoleptic and heteroleptic octahedral complexes from the Cambridge Structural Database<sup>69</sup> (CSD) with a range of metals (Cr to Ni) and direct ligand atom types (N, C, O, S, Cl) (ESI Table S30†). The CSD test cases span a broader range of compounds than the training set, containing (i) larger macrocycles, *e.g.* substituted porphyrins (tests 9, 25), clathrochelates (test 16), phthalocyanines (tests 4, 7), and cyclams (tests 5, 12, 14, 17, 24, 29, and 33, 12 and 33 shown in Fig. 11) and (ii) coordination combinations or functional groups, *e.g.*, OCN in test 30, absent from the training set. Indeed, large CSD test molecule sizes, *e.g.* up to 103 atoms in a single equatorial ligand, further motivates our relatively size-independent descriptor set over forms that do not scale well with molecule size.

The ANN predicts CSD test case spin-state splittings within  $5 \text{ kcal mol}^{-1}$  for 15 of the 35 complexes, an overall mean unsigned error of  $10 \text{ kcal mol}^{-1}$ , and RMSE of  $13 \text{ kcal mol}^{-1}$  (see ESI Table S31†). The large RMSE is due in part to poor performance on early-transition-metal cyclams (red symbols in left panel of Fig. 12) for which the ANN overestimates spin-state

splitting by at about  $30 \text{ kcal mol}^{-1}$  (Cr-cyclams, tests 12 and 33 in Fig. 11). The ANN predicts spin-state splittings within around  $3 \text{ kcal mol}^{-1}$  for several non-macrocyclic complexes that are better represented in the training data (*e.g.*, test cases 8 and 31 in Fig. 11). The correct ground state is assigned in 90% of CSD test cases (96% after excluding cyclams); the only incorrect, non-cyclam spin state assignment is a spin-crossover complex, test 25 (calculated  $\Delta E_{H-L} = -0.2 \text{ kcal mol}^{-1}$ ). Compared to other machine learning models (KRR and SVR), the ANN is more transferable to dissimilar CSD structures (Table 3), outperforming the next-best model, SVR, by 30%. The relative success of the ANN on the CSD data is partially attributable to the use of dropout regularization, which has been shown<sup>96</sup> to improve robustness.

The observation of good performance with reasonable similarity between CSD structures and the training data but poor performance when the CSD structure is not well-represented motivates a quantitative estimate of test compound similarity to training data. We first computed overall molecular similarity metrics (*e.g.*, FP2 fingerprint *via* Tanimoto,<sup>38,112</sup> as implemented in OpenBabel<sup>113,114</sup>) but found limited correlation ( $R^2 = 0.1$ ) to prediction error (see ESI Fig. S29 and Text S5†). Comparing the Euclidean and uncentered Pearson distances in descriptor space between the CSD test cases and the closest training data descriptors provides improved correlation to prediction error of  $R^2 = 0.3$  and  $R^2 = 0.2$ , respectively (ESI Fig. S30†). Large errors (*i.e.*,  $>15 \text{ kcal mol}^{-1}$ ) are only observed at a Euclidean norm difference exceeding 1.0 (half of the CSD data), providing an indication of lack of reliability in ANN prediction. This high distance to training data does not guarantee inaccurate prediction, *e.g.*, CSD test case 8, a  $\text{Fe(II)}$  tetrapyrroline complex, is predicted with fortuitously good  $\sim 2 \text{ kcal mol}^{-1}$  error but has a Euclidean norm difference  $>1.4$ . We have implemented the Euclidean norm metric alongside the ANN in our automated screening code<sup>67</sup> to detect complexes that are poorly represented in training data and advise retraining or direct calculation.

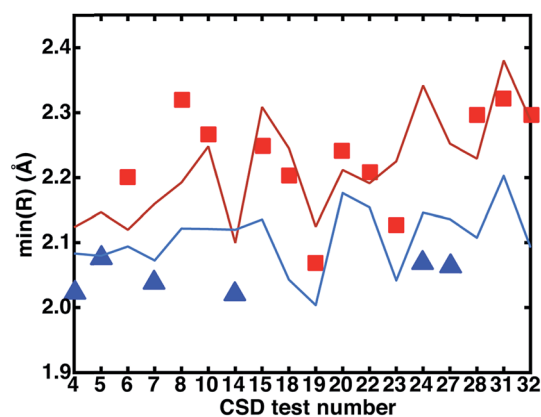


Fig. 13 Comparison of measured CSD bond distances in the crystal phase, represented by symbols (red squares for high-spin or blue triangles for low-spin based on DFT assignment at  $a_{HF} = 0.20$ ) with the ANN predicted HS (red line) and LS (blue line) bond distances. Only the CSD test cases where the difference between DFT LS and HS computed bond distances is  $\geq 0.05 \text{ \AA}$  are shown for clarity. For all of these cases, the ANN correctly predicts the DFT spin state.



ANN-predicted equilibrium metal–ligand bond lengths for both HS and LS CSD geometries produced RMSEs of 0.10 and 0.07 Å, respectively (ESI Tables S32 and S33†). Trends in bond length prediction error differ from those obtained for spin-state splitting. For instance, bond length errors are average in the cyclams even though spin-state splitting predictions were poor. The large Euclidean distance to training data heuristic ( $>1.0$ ) is observed for five of the seven large (*i.e.*,  $>0.1$  Å) HS bond distance errors (see ESI Texts S4 and S5, Fig. S31 and S32†). The highest HS prediction errors ( $>0.2$  Å) occur for tests 8 and 35, underestimating the Fe–N bond length by 0.2 Å (2.1 Å ANN *vs.* 2.3 Å calculated) in the former case. Despite poor geometric predictions, the ANN predicts test 8  $\Delta E_{\text{H-L}}$  to within 3 kcal mol<sup>−1</sup>, and this differing performance is due to the fact that predictions of these two outputs are independent. Interligand effects that are ignored by our descriptor set can restrict bond length extension, *e.g.* in test 16, where an O–H...O<sup>−</sup> interligand hydrogen-bond produces an unusually short 1.9 Å high-spin Fe–N bond distance (*vs.* ANN prediction of 2.1 Å). Future work will focus on incorporating extended metrics of rigidity to account for these effects.

We investigated the relationship between the experimental CSD bond distances and the ANN-predicted bond distances. If the experimentally measured bond distance lies close to one spin state's predicted bond length, then the complex may be expected to be in that spin state, assuming (i) the ANN provides a good prediction of the spin-state specific bond lengths and (ii) that the gas-phase optimized DFT and CSD bond distances are comparable. The majority of experimental bond lengths are near the extrema of the ANN predictions (subset with a calculated DFT LS–HS bond distance of at least 0.05 Å shown in Fig. 13 to improve readability, full set in ESI Fig. S33†). Nine of the twelve (9 of 9 in Fig. 13) experimental bond lengths that are on or above the predicted HS bond distance boundary have an HS ground state, eleven of the fifteen (6 of 6 in Fig. 13) experimental bond lengths that are on or below the predicted LS bond distance have an LS ground state, and remaining structures (3 in Fig. 13) reside at intermediate distances. Some discrepancies are due to differences between the gas phase geometries and those in the crystal environment (*e.g.*, test 27 in Fig. 13 and see ESI Tables S31–S33†). This bond-length-based spin-assignment thus provides a strategy for corroboration of direct spin-state prediction.

### 3.6 Extrapolating pure exchange–correlation functionals to hybrids with an ANN

Linear spin-state HF exchange sensitivity may be exploited to predict properties at one  $a_{\text{HF}}$  value from computed properties obtained at another, *e.g.*, to translate literature values or to accelerate periodic, plane-wave calculations where incorporation of HF exchange increases computational cost. We carry out comparison of the utility of this  $\Delta$ -ML-inspired<sup>52</sup> strategy on the 35 CSD test set to identify if prediction errors are improved, especially for molecules poorly-represented in the training set.

On the CSD molecules, extrapolating  $a_{\text{HF}} = 0.00$  spin-state ordering to  $a_{\text{HF}} = 0.20$  with the exchange-sensitivity ANN

reduces the maximum error to 23 kcal mol<sup>−1</sup> and decreases the mean unsigned error and RMSE to 5 kcal mol<sup>−1</sup> and 7 kcal mol<sup>−1</sup> (the right pane of Fig. 11 and ESI Table S34†). For the GGA + slope ANN approach, excluding the nine cyclams does not change the RMSE/MUE values, confirming good ANN exchange-sensitivity prediction even when spin-state splitting prediction is poor.

These reduced average errors are quite close to the uncertainty introduced by the slope prediction performance at around 4 kcal mol<sup>−1</sup> over a 20% exchange interval. Although this approach does eliminate the largest outliers and improve prediction across the CSD test set, it necessitates semi-local DFT geometry optimizations or a judicious bond length choice for vertically-approximated spin-state ordering. This approach also has limited benefit for cases well-represented in the training data set due to the sparser data set in the exchange sensitivity ANN. Indeed, over the original test set molecules, extrapolated ANN exchange sensitivities on top of calculated  $a_{\text{HF}} = 0.00$  splittings produce an RMSE of around 4 kcal mol<sup>−1</sup> comparable to or slightly worse than direct prediction (ESI Fig. S34†).

## 4. Conclusions

We have presented a series of ANN models trained using 2690 DFT geometry optimizations of octahedral transition metal complexes generated from a set of 16 candidate axial and equatorial ligands and transition metals (Cr–Ni) at varying fractions of HF exchange. From the unseen test cases of a 60–40% train-test partition, we demonstrated good accuracy on spin-state splitting predictions of around 3 kcal mol<sup>−1</sup> and metal–ligand bond distances around 0.02–0.03 Å. Our simple descriptor set, including: (i) the ligand connection atom, (ii) electronegativity and bonding of the coordinating ligand atom environment, (iii) ligand formal charge, (iv) ligand denticity, and (v) metal identity and oxidation state ensures transferability of the ANN. Importantly, the employed connectivity models are not 3D-structure-based, instead relying on a truncated graph-theoretic representation of the ligand, making the approach suitable for screening large numbers of complexes without precise structural information. Although we have trained ANNs to predict bond lengths and spin-state splitting, the data set and descriptors could be used to predict other quantities such as ionization potential, redox potential, or molecular orbital energies. Such efforts are currently underway in our lab.

A test of our ANN on diverse molecules obtained from an experimental database indicated good performance, with MUEs of 5 kcal mol<sup>−1</sup> for spin states for compounds within our proposed Euclidean distance reliability criteria and 10 kcal mol<sup>−1</sup> for the full set. In both diverse and representative cases, the ANN outperforms other machine learning models. Our ANN predictions of HF exchange sensitivity provide a tool for interpolating between exchange–correlation functionals or extrapolating from semi-local GGAs to a hybrid result, which we demonstrated on CSD cases, improving MUE to 5 kcal mol<sup>−1</sup> across the full 35 molecule set.

Natural extensions to this work include the development of the current ANN for extrapolation of GGA to hybrid functional





properties in condensed matter systems and generalizing the coordination definition to enable prediction of properties of unsaturated metals in catalytic cycles. Overall, we have demonstrated a relatively sparse feature space to be capable of predicting electronic structure properties of transition metal complexes, and we anticipate that this strategy may be used for both high-throughput screening with knowledge of functional choice sensitivity and in guiding assessment of sources of errors in approximate DFT.

## Acknowledgements

The authors acknowledge partial support by the National Science Foundation under grant number ECCS-1449291. 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-1053575. The authors thank Adam H. Steeves for providing a critical reading of the manuscript and Prof. Youssef Marzouk for helpful conversations.

## References

- 1 R. Gomez-Bombarelli, J. Aguilera-Iparraguirre, T. D. Hirzel, D. Duvenaud, D. Maclaurin, M. A. Blood-Forsythe, H. S. Chae, M. Einzinger, D. G. Ha, T. Wu, G. Markopoulos, S. Jeon, H. Kang, H. Miyazaki, M. Numata, S. Kim, W. Huang, S. I. Hong, M. Baldo, R. P. Adams and A. Aspuru-Guzik, Design of Efficient Molecular Organic Light-Emitting Diodes by a High-Throughput Virtual Screening and Experimental Approach, *Nat. Mater.*, 2016, **15**, 1120–1127.
- 2 E. O. Pyzer-Knapp, K. Li and A. Aspuru-Guzik, Learning from the Harvard Clean Energy Project: The Use of Neural Networks to Accelerate Materials Discovery, *Adv. Funct. Mater.*, 2015, **25**, 6495–6502.
- 3 J. K. Norskov and T. Bligaard, The Catalyst Genome, *Angew. Chem., Int. Ed. Engl.*, 2013, **52**, 776–777.
- 4 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner and G. Ceder, Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation, *APL Mater.*, 2013, **1**, 011002.
- 5 A. M. Virshup, J. Contreras-García, P. Wipf, W. Yang and D. N. Beratan, Stochastic Voyages into Uncharted Chemical Space Produce a Representative Library of All Possible Drug-Like Compounds, *J. Am. Chem. Soc.*, 2013, **135**, 7296–7303.
- 6 P. Kirkpatrick and C. Ellis, Chemical Space, *Nature*, 2004, **432**, 823–823.
- 7 B. Meredig, A. Agrawal, S. Kirklin, J. E. Saal, J. W. Doak, A. Thompson, K. Zhang, A. Choudhary and C. Wolverton, Combinatorial Screening for New Materials in Unconstrained Composition Space with Machine Learning, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2014, **89**, 094104.
- 8 L. Li, J. C. Snyder, I. M. Pelaschier, J. Huang, U.-N. Niranjan, P. Duncan, M. Rupp, K.-R. Müller and K. Burke, Understanding Machine-Learned Density Functionals, *Int. J. Quantum Chem.*, 2016, **116**, 819–833.
- 9 M. Rupp, Machine Learning for Quantum Mechanics in a Nutshell, *Int. J. Quantum Chem.*, 2015, **115**, 1058–1073.
- 10 J. Behler, Perspective: Machine Learning Potentials for Atomistic Simulations, *J. Chem. Phys.*, 2016, **145**, 170901.
- 11 J. Behler, Representing Potential Energy Surfaces by High-Dimensional Neural Network Potentials, *J. Phys.: Condens. Matter*, 2014, **26**, 183001.
- 12 S. Lorenz, A. Groß and M. Scheffler, Representing High-Dimensional Potential-Energy Surfaces for Reactions at Surfaces by Neural Networks, *Chem. Phys. Lett.*, 2004, **395**, 210–215.
- 13 N. Artrith, T. Morawietz and J. Behler, High-Dimensional Neural-Network Potentials for Multicomponent Systems: Applications to Zinc Oxide, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **83**, 153101.
- 14 J. Behler and M. Parrinello, Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces, *Phys. Rev. Lett.*, 2007, **98**, 146401.
- 15 F. V. Prudente and J. J. S. Neto, The Fitting of Potential Energy Surfaces Using Neural Networks. Application to the Study of the Photodissociation Processes, *Chem. Phys. Lett.*, 1998, **287**, 585–589.
- 16 L. Mones, N. Bernstein and G. Csanyi, Exploration, Sampling, and Reconstruction of Free Energy Surfaces with Gaussian Process Regression, *J. Chem. Theory Comput.*, 2016, **12**, 5100–5110.
- 17 J. S. Smith, O. Isayev and A. E. Roitberg, Ani-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost, *Chem. Sci.*, 2017, 3192–3203.
- 18 J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller and K. Burke, Finding Density Functionals with Machine Learning, *Phys. Rev. Lett.*, 2012, **108**, 253002.
- 19 K. Mills, M. Spanner and I. Tamblin, Deep Learning and the Schrödinger Equation, arXiv preprint arXiv:1702.01361, 2017.
- 20 K. Yao and J. Parkhill, Kinetic Energy of Hydrocarbons as a Function of Electron Density and Convolutional Neural Networks, *J. Chem. Theory Comput.*, 2016, **12**, 1139–1147.
- 21 J. C. Snyder, M. Rupp, K. Hansen, L. Blooston, K.-R. Müller and K. Burke, Orbital-Free Bond Breaking via Machine Learning, *J. Chem. Phys.*, 2013, **139**, 224104.
- 22 K. Yao, J. E. Herr and J. Parkhill, The Many-Body Expansion Combined with Neural Networks, *J. Chem. Phys.*, 2017, **146**, 014106.
- 23 F. Hase, S. Valleau, E. Pyzer-Knapp and A. Aspuru-Guzik, Machine Learning Exciton Dynamics, *Chem. Sci.*, 2016, **7**, 5139–5147.
- 24 Z. Li, J. R. Kermode and A. De Vita, Molecular Dynamics with on-the-Fly Machine Learning of Quantum-Mechanical Forces, *Phys. Rev. Lett.*, 2015, **114**, 096405.



- 25 V. Botu and R. Ramprasad, Adaptive Machine Learning Framework to Accelerate *Ab Initio* Molecular Dynamics, *Int. J. Quantum Chem.*, 2015, **115**, 1074–1083.
- 26 G. Pilania, J. Gubernatis and T. Lookman, Multi-Fidelity Machine Learning Models for Accurate Bandgap Predictions of Solids, *Comput. Mater. Sci.*, 2017, **129**, 156–163.
- 27 G. Pilania, A. Mannodi-Kanakkithodi, B. P. Uberuaga, R. Ramprasad, J. E. Gubernatis and T. Lookman, Machine Learning Bandgaps of Double Perovskites, *Sci. Rep.*, 2016, **6**, 19375.
- 28 X. Ma, Z. Li, L. E. K. Achenie and H. Xin, Machine-Learning-Augmented Chemisorption Model for CO<sub>2</sub> Electroreduction Catalyst Screening, *J. Phys. Chem. Lett.*, 2015, **6**, 3528–3533.
- 29 A. Mannodi-Kanakkithodi, G. Pilania, T. D. Huan, T. Lookman and R. Ramprasad, Machine Learning Strategy for Accelerated Design of Polymer Dielectrics, *Sci. Rep.*, 2016, **6**, 20952.
- 30 T. D. Huan, A. Mannodi-Kanakkithodi and R. Ramprasad, Accelerated Materials Property Predictions and Design Using Motif-Based Fingerprints, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2015, **92**, 014106.
- 31 G. Pilania, C. Wang, X. Jiang, S. Rajasekaran and R. Ramprasad, Accelerating Materials Property Predictions Using Machine Learning, *Sci. Rep.*, 2013, **3**, 2810.
- 32 J. Lee, A. Seko, K. Shitara and I. Tanaka, Prediction Model of Band-Gap for AX Binary Compounds by Combination of Density Functional Theory Calculations and Machine Learning Techniques, *Phys. Rev. B*, 2016, **93**, 115104.
- 33 T. Morawietz and J. Behler, A Density-Functional Theory-Based Neural Network Potential for Water Clusters Including van der Waals Corrections, *J. Phys. Chem. A*, 2013, **117**, 7356–7366.
- 34 T. Morawietz, A. Singraber, C. Dellago and J. Behler, How van der Waals Interactions Determine the Unique Properties of Water, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, 201602375.
- 35 M. Rupp, A. Tkatchenko, K.-R. Müller and O. A. von Lilienfeld, Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning, *Phys. Rev. Lett.*, 2012, **108**, 058301.
- 36 B. Huang and O. A. von Lilienfeld, Communication: Understanding Molecular Representations in Machine Learning: The Role of Uniqueness and Target Similarity, *J. Chem. Phys.*, 2016, **145**, 161102.
- 37 S. De, A. P. Bartk, G. Csányi and M. Ceriotti, Comparing Molecules and Solids across Structural and Alchemical Space, *Phys. Chem. Chem. Phys.*, 2016, **18**, 1–18.
- 38 G. Maggiora, M. Vogt, D. Stumpfe and J. Bajorath, Molecular Similarity in Medicinal Chemistry: Miniperspective, *J. Med. Chem.*, 2013, **57**, 3186–3204.
- 39 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, Development and Testing of a General Amber Force Field, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
- 40 H. Kubinyi, QSAR and 3D QSAR in Drug Design. Part 1: Methodology, *Drug Discovery Today*, 1997, **2**, 457–467.
- 41 S. W. Benson, F. Cruickshank, D. Golden, G. R. Haugen, H. O'neal, A. Rodgers, R. Shaw and R. Walsh, Additivity Rules for the Estimation of Thermochemical Properties, *Chem. Rev.*, 1969, **69**, 279–324.
- 42 R. J. Deeth, The Ligand Field Molecular Mechanics Model and the Stereoelectronic Effects of d and s Electrons, *Coord. Chem. Rev.*, 2001, **212**, 11–34.
- 43 D. F. Shriver and P. W. Atkins, *Inorganic Chemistry*, W. H. Freeman and Co., 3rd edn, 1999.
- 44 K. T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K. R. Müller and E. K. U. Gross, How to Represent Crystal Structures for Machine Learning: Towards Fast Prediction of Electronic Properties, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2014, **89**, 205118.
- 45 E. I. Ioannidis and H. J. Kulik, Towards Quantifying the Role of Exact Exchange in Predictions of Transition Metal Complex Properties, *J. Chem. Phys.*, 2015, **143**, 034104.
- 46 D. C. Ashley and E. Jakubikova, Ironing out the Photochemical and Spin-Crossover Behavior of Fe(II) Coordination Compounds with Computational Chemistry, *Coord. Chem. Rev.*, 2017, 97–111.
- 47 D. N. Bowman and E. Jakubikova, Low-Spin versus High-Spin Ground State in Pseudo-Octahedral Iron Complexes, *Inorg. Chem.*, 2012, **51**, 6011–6019.
- 48 T. Z. H. Gani and H. J. Kulik, Where Does the Density Localize? Convergent Behavior for Global Hybrids, Range Separation, and DFT+U, *J. Chem. Theory Comput.*, 2016, **12**, 5931–5945, DOI: 10.1021/acs.jctc.6b00937.
- 49 E. I. Ioannidis and H. J. Kulik, Ligand-Field-Dependent Behavior of Meta-GGA Exchange in Transition-Metal Complex Spin-State Ordering, *J. Phys. Chem. A*, 2017, **121**(4), 874–884, DOI: 10.1021/acs.jpca.6b11930.
- 50 W. Huang, D.-H. Xing, J.-B. Lu, B. Long, W. E. Schwarz and J. Li, How Much Can Density Functional Approximations (DFA) Fail? The Extreme Case of the FeO<sub>4</sub> Species, *J. Chem. Theory Comput.*, 2016, **12**, 1525–1533.
- 51 J. J. P. Stewart, Optimization of Parameters for Semiempirical Methods VI: More Modifications to the NDDO Approximations and Re-Optimization of Parameters, *J. Mol. Model.*, 2013, **19**, 1–32.
- 52 R. Ramakrishnan, P. O. Dral, M. Rupp and O. A. von Lilienfeld, Big Data Meets Quantum Chemistry Approximations: The Delta-Machine Learning Approach, *J. Chem. Theory Comput.*, 2015, **11**, 2087–2096.
- 53 L. Shen, J. Wu and W. Yang, Multiscale Quantum Mechanics/Molecular Mechanics Simulations with Neural Networks, *J. Chem. Theory Comput.*, 2016, **12**, 4934–4946.
- 54 H. J. Kulik, Perspective: Treating Electron over-Delocalization with the DFT+U Method, *J. Chem. Phys.*, 2015, **142**, 240901.
- 55 A. J. Cohen, P. Mori-Sanchez and W. Yang, Insights into Current Limitations of Density Functional Theory, *Science*, 2008, **321**, 792–794.
- 56 O. Salomon, M. Reiher and B. A. Hess, 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.



- 57 M. Reiher, Theoretical Study of the  $\text{Fe}(\text{Phen})_2(\text{NCS})_2$  Spin-Crossover Complex with Reparametrized Density Functionals, *Inorg. Chem.*, 2002, **41**, 6928–6935.
- 58 M. Reiher, O. Salomon and B. A. Hess, Reparameterization of Hybrid Functionals Based on Energy Differences of States of Different Multiplicity, *Theor. Chem. Acc.*, 2001, **107**, 48–55.
- 59 A. Droghetti, D. Alf   and S. Sanvito, Assessment of Density Functional Theory for Iron(II) Molecules across the Spin-Crossover Transition, *J. Chem. Phys.*, 2012, **137**, 124303.
- 60 J. E. Sutton, W. Guo, M. A. Katsoulakis and D. G. Vlachos, Effects of Correlated Parameters and Uncertainty in Electronic-Structure-Based Chemical Kinetic Modelling, *Nat. Chem.*, 2016, **8**, 331–337.
- 61 G. N. Simm and M. Reiher, Systematic Error Estimation for Chemical Reaction Energies, *J. Chem. Theory Comput.*, 2016, 2762–2773.
- 62 E. Walker, S. C. Ammal, G. A. Terejanu and A. Heyden, Uncertainty Quantification Framework Applied to the Water–Gas Shift Reaction over Pt-Based Catalysts, *J. Phys. Chem. C*, 2016, **120**, 10328–10339.
- 63 M. A. Halcrow, Structure: Function Relationships in Molecular Spin-Crossover Complexes, *Chem. Soc. Rev.*, 2011, **40**, 4119–4142.
- 64 J.-F. L  tard, P. Guionneau and L. Goux-Capes, Towards Spin Crossover Applications, in *Spin Crossover in Transition Metal Compounds III*, Springer, 2004, pp. 221–249.
- 65 C. A. Bignozzi, R. Argazzi, R. Boaretto, E. Busatto, S. Carli, F. Ronconi and S. Caramori, The Role of Transition Metal Complexes in Dye-Sensitized Solar Devices, *Coord. Chem. Rev.*, 2013, **257**, 1472–1492.
- 66 J. N. Harvey, R. Poli and K. M. Smith, Understanding the Reactivity of Transition Metal Complexes Involving Multiple Spin States, *Coord. Chem. Rev.*, 2003, **238**, 347–361.
- 67 E. I. Ioannidis, T. Z. H. Gani and H. J. Kulik, Molsimplify: A Toolkit for Automating Discovery in Inorganic Chemistry, *J. Comput. Chem.*, 2016, **37**, 2106–2117.
- 68 A. Kramida, Y. Ralchenko and J. Reader, NIST ASD Team NIST Atomic Spectra Database (Version 5.3), <http://physics.nist.gov/asd>, accessed March 14, 2017.
- 69 C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, The Cambridge Structural Database, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2016, **72**, 171–179.
- 70 I. S. Ufimtsev and T. J. Martinez, 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.
- 71 Petachem, <http://www.petachem.com>, accessed March 14, 2017.
- 72 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *Ab Initio* Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
- 73 A. D. Becke, Density-Functional Thermochemistry. III. The Role of Exact Exchange, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
- 74 C. Lee, W. Yang and R. G. Parr, 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.
- 75 P. J. Hay and W. R. Wadt, *Ab Initio* Effective Core Potentials for Molecular Calculations. Potentials for the Transition Metal Atoms Sc to Hg, *J. Chem. Phys.*, 1985, **82**, 270–283.
- 76 V. R. Saunders and I. H. Hillier, A “Level-Shifting” Method for Converging Closed Shell Hartree–Fock Wave Functions, *Int. J. Quantum Chem.*, 1973, **7**, 699–705.
- 77 J. K  stner, J. M. Carr, T. W. Keal, W. Thiel, A. Wander and P. Sherwood, DL-FIND: An Open-Source Geometry Optimizer for Atomistic Simulations, *J. Phys. Chem. A*, 2009, **113**, 11856–11865.
- 78 G. Ganzenmuller, N. Berkaine, A. Fouqueau, M. E. Casida and M. Reiher, Comparison of Density Functionals for Differences between the High- (5t2g) and Low- (1a1g) Spin States of Iron(II) Compounds. IV. Results for the Ferrous Complexes  $[\text{Fe}(\text{L})(\text{NHS}_4)]$ , *J. Chem. Phys.*, 2005, **122**, 234321.
- 79 A. Cereto-Massague, M. J. Ojeda, C. Valls, M. Mulero, S. Garcia-Vallve and G. Pujadas, Molecular Fingerprint Similarity Search in Virtual Screening, *Methods*, 2015, **71**, 58–63.
- 80 R. P. Sheridan, M. D. Miller, D. J. Underwood and S. K. Kearsley, Chemical Similarity Using Geometric Atom Pair Descriptors, *J. Chem. Inf. Model.*, 1996, **36**, 128–136.
- 81 K. Hansen, F. Biegler, R. Ramakrishnan and W. Pronobis, Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space, *J. Phys. Chem. Lett.*, 2015, **6**, 2326–2331.
- 82 M. Gastegger and P. Marquetand, High-Dimensional Neural Network Potentials for Organic Reactions and an Improved Training Algorithm, *J. Chem. Theory Comput.*, 2015, **11**, 2187–2198.
- 83 J. A. Hageman, J. A. Westerhuis, H. W. Frhauf and G. Rothenberg, Design and Assembly of Virtual Homogeneous Catalyst Libraries – Towards *in silico* Catalyst Optimisation, *Adv. Synth. Catal.*, 2006, **348**, 361–369.
- 84 M. Randic, On Molecular Branching, *J. Am. Chem. Soc.*, 1975, **97**, 57–77.
- 85 H. Wiener, Correlation of Heats of Isomerization, and Differences in Heats of Vaporization of Isomers, among the Paraffin Hydrocarbons, *J. Am. Chem. Soc.*, 1947, **69**, 2636–2638.
- 86 L. B. Kier, A Shape Index from Molecular Graphs, *Quant. Struct.-Act. Relat.*, 1985, **4**, 109–116.
- 87 G. Montavon, K. Hansen, S. Fazli and M. Rupp, in *Learning Invariant Representations of Molecules for Atomization Energy Prediction*, Advances in Neural Information Processing Systems, ed. F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger, Curran Associates, Inc., 2012, pp. 440–448.
- 88 M. Gastegger, C. Kauffmann, J. Behler and P. Marquetand, Comparing the Accuracy of High-Dimensional Neural Network Potentials and the Systematic Molecular



- Fragmentation Method: A Benchmark Study for All-Trans Alkanes, *J. Chem. Phys.*, 2016, **144**, 194110.
- 89 T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning*, Springer, New York, 2009, vol. 18, p. 764.
  - 90 J. Friedman, T. Hastie and R. Tibshirani, Regularization Paths for Generalized Linear Models via Coordinate Descent, *J. Stat. Software*, 2010, **33**, 1–22.
  - 91 R Core Development Team, *R: A Language and Environment for Statistical Computing*, 2016.
  - 92 H. Larochelle, Y. Bengio, J. Louradour and P. Lamblin, Exploring Strategies for Training Deep Neural Networks, *J. Mach. Learn. Res.*, 2009, **10**, 1–40.
  - 93 S. Aiello, T. Kraljevic and P. Maj, *H2O: R Interface for H2O*, 2015.
  - 94 Y. Gal and Z. Ghahramani, Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning, arXiv preprint arXiv:1506.02142, 2015.
  - 95 N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, Dropout: A Simple Way to Prevent Neural Networks from Overfitting, *J. Mach. Learn. Res.*, 2014, **15**, 1929–1958.
  - 96 G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever and R. R. Salakhutdinov, Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors, arXiv preprint arXiv:1207.0580, 2012, pp. 1–18.
  - 97 Y. Bengio, Practical Recommendations for Gradient-Based Training of Deep Architectures, in *Neural Networks: Tricks of the Trade*, ed. G. B. Orr, K. R. Muller and M. Gregoire, Springer, 2012, pp. 437–478.
  - 98 Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, Gradient-Based Learning Applied to Document Recognition, *Proc. IEEE*, 1998, **86**, 2278–2323.
  - 99 A. Candel, V. Parmar, E. LeDell and A. Arora, *Deep Learning with H2O*, H2O, 2015.
  - 100 F. Niu, B. Recht, C. Re and S. J. Wright, *Hogwild!: A Lock-Free Approach to Parallelizing Stochastic Gradient Descent*, Advances in Neural Information Processing Systems, 2011, p. 21.
  - 101 G. B. Kingston, M. F. Lambert and H. R. Maier, Bayesian Training of Artificial Neural Networks Used for Water Resources Modeling, *Water Resour. Res.*, 2005, **41**, 1–11.
  - 102 P. Secchi and E. Zio, Quantifying Uncertainties in the Estimation of Safety Parameters by Using Bootstrapped Artificial Neural Networks, *Ann. Nucl. Energy*, 2008, **35**, 2338–2350.
  - 103 K. Hansen, G. Montavon, F. Biegler, S. Fazli, M. Rupp, M. Scheffler, O. A. von Lilienfeld, A. Tkatchenko and K.-R. Müller, Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies, *J. Chem. Theory Comput.*, 2013, **9**, 3404–3419.
  - 104 A. Zeileis, K. Hornik, A. Smola and A. Karatzoglou, Kernlab – an S4 Package for Kernel Methods in R, *J. Stat. Software*, 2004, **11**, 1–20.
  - 105 T. Krueger, D. Panknin and M. Braun, Fast Cross-Validation via Sequential Testing, *J. Mach. Learn. Res.*, 2015, **16**, 1103–1155.
  - 106 L. Breiman, J. Friedman, R. A. Olshen and C. Stone, *Classification and Regression Trees*, Chapman and Hall, CRC, 1984, vol. 5, pp. 95–96.
  - 107 T. Therneau, B. Atkinson, B. Ripley, Rpart: Recursive Partitioning and Regression Trees, <https://cran.r-project.org/package=rpart>, accessed March 14, 2017.
  - 108 D. Coskun, S. V. Jerome and R. A. Friesner, Evaluation of the Performance of the B3LYP, PBE0, and M06 DFT Functionals, and DBLOC-Corrected Versions, in the Calculation of Redox Potentials and Spin Splittings for Transition Metal Containing Systems, *J. Chem. Theory Comput.*, 2016, **12**, 1121–1128.
  - 109 T. F. Hughes, J. N. Harvey and R. A. Friesner, A B3LYP-DBLOC Empirical Correction Scheme for Ligand Removal Enthalpies of Transition Metal Complexes: Parameterization against Experimental and CCSD(T)-F12 Heats of Formation, *Phys. Chem. Chem. Phys.*, 2012, **14**, 7724–7738.
  - 110 T. F. Hughes and R. A. Friesner, Correcting Systematic Errors in DFT Spin-Splitting Energetics for Transition Metal Complexes, *J. Chem. Theory Comput.*, 2011, **7**, 19–32.
  - 111 W. Jiang, N. J. Deyonker, J. J. Determan and A. K. Wilson, Toward Accurate Theoretical Thermochemistry of First Row Transition Metal Complexes, *J. Phys. Chem. A*, 2012, **116**, 870–885.
  - 112 D. Bajusz, A. Racz and K. Heberger, Why Is Tanimoto Index an Appropriate Choice for Fingerprint-Based Similarity Calculations?, *J. Cheminf.*, 2015, **7**, 20.
  - 113 N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, Open Babel: An Open Chemical Toolbox, *J. Cheminf.*, 2011, **3**, 33.
  - 114 The Open Babel Package Version 2.3.1, <http://openbabel.org>, accessed March 14, 2017.

