

Exchange Spin Coupling from Gaussian Process Regression

Marc Philipp Bahlke,^{*,†} Natnael Mogos,[†] Jonny Proppe,^{*,‡} and Carmen
Herrmann^{*,†}

*†Department of Chemistry, University of Hamburg, Martin-Luther-King-Platz 6, 20146
Hamburg, Germany*

*‡Institute of Physical Chemistry, Georg-August University, Tammannstr. 6, 37077
Göttingen, Germany*

E-mail: marc.philipp.bahlke@chemie.uni-hamburg.de; jonny.proppe@chemie.uni-goettingen.de;
carmen.herrmann@chemie.uni-hamburg.de

Abstract

Heisenberg exchange spin coupling between metal centers is essential for describing and understanding the electronic structure of many molecular catalysts, metalloenzymes, and molecular magnets for potential application in information technology. We explore the machine-learnability of exchange spin coupling, which has not been studied yet. We employ Gaussian process regression since it can potentially deal with small training sets (as likely associated with the rather complex molecular structures required for exploring spin coupling) and since it provides uncertainty estimates (“error bars”) along with predicted values. We compare a range of descriptors and kernels for 257 small dicopper complexes and find that a simple descriptor based on chemical intuition, consisting only of copper–bridge angles and copper–copper distances, clearly outperforms several more sophisticated descriptors when it comes to extrapolating towards larger experimentally relevant complexes. Exchange spin coupling is similarly easy to learn as the polarizability, while learning dipole moments is much harder. The strength of the sophisticated descriptors lies in their ability to linearize structure–property relationships, to the point that a simple linear ridge regression performs just as well as the kernel-based machine-learning model for our small dicopper data set. The superior extrapolation performance of the simple descriptor is unique to exchange spin coupling, reinforcing the crucial role of choosing a suitable descriptor, and highlighting the interesting question of the role of chemical intuition vs. systematic or automated selection of features for machine learning in chemistry and material science.

1 Introduction

Interactions between unpaired electrons localized on metal centers are an important feature of the active sites of many metalloenzymes¹, artificial catalysts², and molecular magnets with possible applications such as data storage, quantum computing or magnetic materials^{3–11}. Such interactions are usually not mediated directly by magnetic interactions due to the distance between the spin centers, but rather via a combination of Coulomb interaction and exchange interactions resulting from the Pauli exclusion principle. If the interactions are mediated by (formally) closed-shell ligands, they are referred to as superexchange^{12,13}. The low-energy spectrum of (super-)exchange-coupled systems with two local spins is often described via a Heisenberg–Dirac–van-Vleck Hamiltonian¹⁴,

$$\hat{H} = -2J\hat{\mathbf{S}}_1 \cdot \hat{\mathbf{S}}_2, \quad (1)$$

where $\hat{\mathbf{S}}_1$ and $\hat{\mathbf{S}}_2$ denote the local spin vectors and J is the Heisenberg exchange coupling constant. J is positive if the two spins are ferromagnetically (F) coupled, i.e., aligned, and negative if they are antiferromagnetically (AF) coupled, i.e. antialigned in the ground state. The magnitude of J expresses the strength of the interaction, i.e., the energy difference between the F and the AF coupled states^a.

Our goal is to predict J for transition metal complexes. This is most often done by combining Kohn–Sham (KS) density functional theory (DFT) with a broken-symmetry approach to model the AF state^{15–21}, or by employing multireference wave function methods^{18,22–26}. The latter is computationally demanding, and requires efforts towards avoiding artefacts resulting from too inflexible wave function expressions^{27–30}. DFT is feasible for many complexes of interest for the research areas mentioned above in full atomistic detail, but still computationally expensive for applications in tasks such as screening, and also for routine evaluations of spin coupling in

^aThere are different definitions of \hat{H} , in particular the factor of 2 and the minus sign may be missing, resulting in a scaling and sign change of J .

more complex structures as relevant in molecular spintronics or nanostructured materials³.

A possible solution to this problem are machine-learning (ML) approaches^{31,32}, which have seen an enormous gain in popularity and development in the past few years in the chemistry, materials science and solid-state physics communities^{33–53}, and which may save orders of magnitude in computing time compared to DFT, and often comparable or even better accuracy. ML has proven successful for spin-dependent molecular properties, in particular for spin-state energies in spin crossover complexes^{35,36,54–57}, magnetic moments and magnetic anisotropy^{58,59}, and also for properties closely related^{60,61,61–68} to exchange spin coupling such as charge transfer^{69–71} and excitation energy transfer^{71,72}. The capability of ML for exchange spin coupling has not been explored yet.

When making predictions with ML for molecular structures outside of the training set, it is desirable to have a measure for how reliable those predictions are, i.e., an error or uncertainty estimate. Such estimates have been implemented in chemistry-related ML frameworks⁷³, e.g. based on a latent space distance metric for artificial neural networks⁷⁴ or employing resampling techniques such as bootstrapping or subsampling⁷⁵.

Gaussian process regression (GPR)⁷⁶ provides a natural access to error bars / uncertainty estimates. Experience so far shows that it is suitable for rather small training sets, which is a favorable property given that reference data for J are relatively costly to generate. GPR has been used in chemistry for, e.g., fitting repulsive potentials in tight-binding DFT⁷⁷, for correcting empirical dispersion models⁷⁸, for evaluating work functions⁷⁹, for calculating vibrational Raman spectra⁸⁰, for transition-state⁸¹ and molecular-structure optimization^{82–85}, for fitting potential-energy surfaces^{86,87}, and for the error-controlled exploration of reaction networks^{88,89}. We will employ GPR here for asking to what extent it is possible to machine-learn J , as compared with other molecular properties.

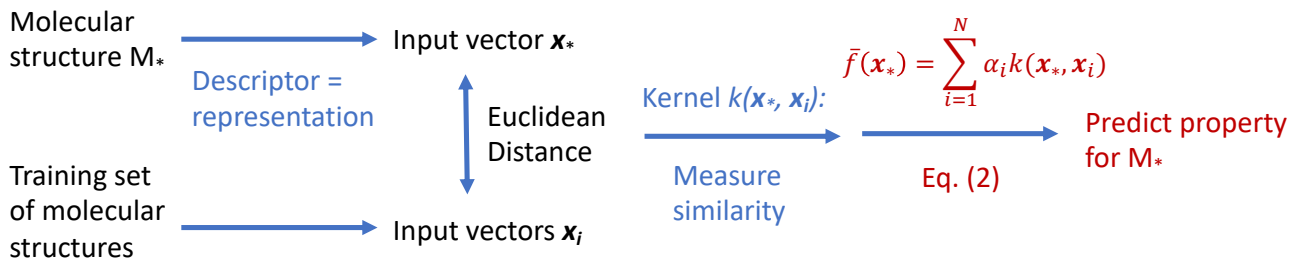


Figure 1: Overview of machine learning via Gaussian Process Regression, illustrating the central importance of for molecular descriptors / representations and kernels for GPR.

GPR is a kernel method in which structure–property relationships are inferred from similarities between molecules. For measuring such similarities, in GPR as in any ML approach in chemistry, it is important to choose a unique and information-conserving way of translating molecular structures into ML-accessible format (molecular representations / descriptors)^{34,49,55,90–97}. Also the choice of kernel will affect the quality of the ML model^{76,98,99}. We will study which combination of descriptor and kernel works best for evaluating J for a set of model molecular systems.

We focus on 257 dinuclear copper(II) complexes with halogenide, oxalate, acetate, methanolate and hydroxide ligands (Figure 2). These $[\text{Cu}_2]$ systems show distinct structure–property correlations¹⁰⁰ for J . They feature only one unpaired-spin electron per metal center, and they are small and structurally simple enough to allow for straightforward automated construction of all possible molecules^b. Polynuclear copper complexes can serve as functional and structural models of multicopper enzymes^{16,102,103}. Dinuclear copper complexes constitute important building blocks for materials such as metal-organic frameworks^{104–106}, and an important test case for electronic structure methods^{16,27–29,107}.

^bEmploying a more sophisticated approach, systematic enumeration of more complex transition metal complexes has also been demonstrated recently for mononuclear systems¹⁰¹.

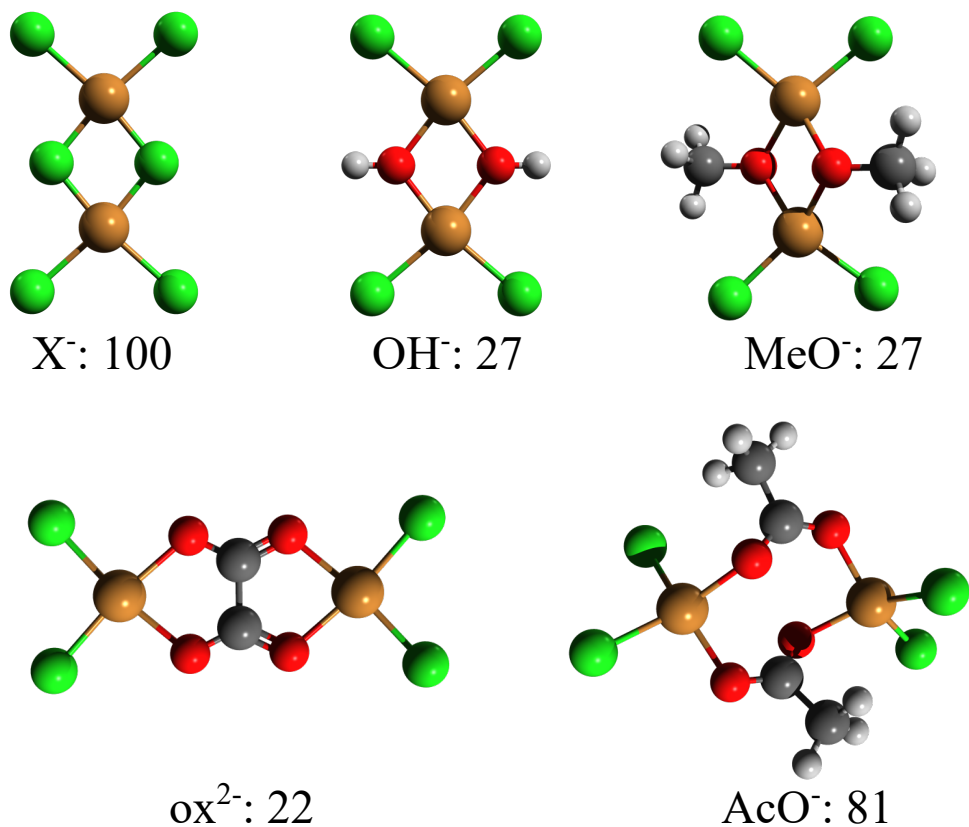


Figure 2: Composition of the $[\text{Cu}_2]$ data set used in this work. The structures contain all possible combinations of F, Br, and Cl (green atoms). In total, the data set consists of 257 structures.

For evaluating exchange-coupling constants J , we employ a Green’s function approach^{108–110}. This approach can estimate J from only the high-spin (ferromagnetically coupled) electronic structure, which is much easier to converge than the antiferromagnetically coupled one, thus lending itself to the robust generation of training data in ML.

Due to the lack of published data on modeling J via ML, we first analyze the performance of selected descriptors within GPR and compare it with linear ridge regression. While not being as flexible as GPR-derived models, linear ridge regression allows for better interpretation through the regression coefficients. It may guide researchers in answering the question of why a certain J is positive (F coupling) or negative (AF coupling). Such insights are hard to obtain from other ML methods such as kernel ridge regression, neural networks, or GPR. In this

context it is also studied how the prediction is affected by the kernel function, whose selection is required in GPR. We then compare the prediction performance for J with other properties such as the polarizability and the dipole moment. Finally, we study the prediction of J for experimental structures selected from the literature, to answer the question to what extent GPR is capable of extrapolating to structures far from the training set.

2 Theory

2.1 Gaussian Process Regression

GPR relies on the assumption that two function values, $f(\mathbf{x}_1)$ and $f(\mathbf{x}_2)$, are highly correlated if their input vectors (molecular representations/descriptors), \mathbf{x}_1 and \mathbf{x}_2 , are similar to each other, i.e., $f(\mathbf{x})$ is a smooth function. This is a reasonable assumption all over the natural sciences and beyond, in particular for the relation between molecular structures (which have to be translated to fixed-size vectors, see Section 2.3) and properties. The questions we seek to answer are how to measure similarity between molecular structures/representations and how fast similarity increases/decreases, i.e., how fast the correlation between $f(\mathbf{x}_1)$ and $f(\mathbf{x}_2)$ changes if \mathbf{x}_2 is changed.

In GPR, the property $f(\mathbf{x}_*)$ of a molecular structure of interest — represented by the input vector \mathbf{x}_* — is represented as a normal distribution. Its mean, the predictive function $\bar{f}(\mathbf{x}_*)$, constitutes its most likely value and can be expressed as a linear combination over kernels,

$$\bar{f}(\mathbf{x}_*) = \sum_{i=1}^N \alpha_i k(\mathbf{x}_*, \mathbf{x}_i), \quad (2)$$

where $k(\mathbf{x}_*, \mathbf{x}_i)$ is the kernel (or covariance function or similarity measure) with respect to the i -th structure of the training set (N training data points in total) and α_i is the corresponding weight. Similarly, the definition of the variance (“squared uncertainty”) of $f(\mathbf{x}_*)$, $\mathbb{V}[f(\mathbf{x}_*)]$,

involves a linear combination over kernels subtracted from a constant (the self-kernel),

$$\mathbb{V}[f(\mathbf{x}_*)] = k(\mathbf{x}_*, \mathbf{x}_*) - \sum_{i=1}^N \beta_i^* k(\mathbf{x}_*, \mathbf{x}_i), \quad (3)$$

with $0 \leq \sum_{i=1}^N \beta_i^* k(\mathbf{x}_*, \mathbf{x}_i) \leq k(\mathbf{x}_*, \mathbf{x}_*)$. Both $\{\alpha_i\}$ and $\{\beta_i^*\}$ are determined during regression (see Section A.2 for more details). As a result, with an increasing Euclidean distance between vector \mathbf{x}_* and the training vectors $\{\mathbf{x}_i\}$, the kernels $\{k(\mathbf{x}_*, \mathbf{x}_i)\}$ converge toward zero, finally leading to $\bar{f}(\mathbf{x}_*) = 0$ and $\mathbb{V}[f(\mathbf{x}_*)] = k(\mathbf{x}_*, \mathbf{x}_*)$. Hence, with the help of GPR, one can easily identify explorations into uncharted territory.

2.2 Kernels

Several ways of defining kernels are employed in ML⁷⁶, and there is at present no general recipe for how to choose the best performing one¹¹¹. One option is a Gaussian kernel, also called radial basis function (RBF) or squared exponential kernel,

$$k_{\text{RBF}}(\mathbf{x}_1, \mathbf{x}_2) = \sigma_f^2 \exp\left(-\frac{|\mathbf{x}_1 - \mathbf{x}_2|^2}{2l^2}\right), \quad (4)$$

where σ_f^2 (signal variance) and l (length scale) are hyperparameters to be optimized (see Section A.2). The kernel generates a rather smooth predictive function.

The RBF kernel can be defined as a kernel of Matérn(ν) type with $\nu \rightarrow \infty$. A Matérn(ν) kernel can be considered an interpolation between Gaussian (squared exponential) and Laplacian (exponential) kernels, with $\nu = 1/2$ as lower bound representing the Laplacian kernel,

$$k_{\nu=1/2}(\mathbf{x}_1, \mathbf{x}_2) = \sigma_f^2 \exp\left(-\frac{|\mathbf{x}_1 - \mathbf{x}_2|}{l}\right), \quad (5)$$

resulting in a rough predictive function.

We also consider a kernel in between the two, with $\nu = 3/2$,

$$k_{\nu=3/2}(\mathbf{x}_1, \mathbf{x}_2) = \sigma_f^2 \left(1 + \frac{\sqrt{3} |\mathbf{x}_1 - \mathbf{x}_2|}{l} \right) \exp \left(-\frac{\sqrt{3} |\mathbf{x}_1 - \mathbf{x}_2|}{l} \right). \quad (6)$$

We choose this family of kernels because it allows for generating a variety of correlation patterns, but this does not exclude the possibility that there may be other types of kernels equally (or even better) suited for our type of ML problem⁷⁶. We will ask below how ML performance depends on this choice of kernel, and to what extent this depends on the choice of descriptor and molecular property.

2.3 Molecular Representations

Molecular representations or descriptors (here, in fixed-size vectorial form) are well-established in organic chemistry¹¹². For transition-metal systems, there is far less experience. For the spin state energetics in spin-crossover compounds, revised two-dimensional autocorrelation functions combined with feature selection algorithms have proven as very successful, both with kernel ridge regression and with artificial neural networks⁵⁷. For learning magnetic anisotropy in single-ion molecular magnets, bispectrum components combined with ridge regression were shown to be a good choice⁵⁹. As a first step towards exploring the machine-learnability of J , we will focus on a selection of descriptors which have proven successful for a variety of properties and structures, and which are inexpensive to evaluate computationally: Smooth overlap of atomic positions (SOAP)¹¹³, many-body tensor representation (MBTR), and many-body interaction descriptors¹¹⁴ (F_{2B} and F_{3B}). In addition, we study how Cu–Cu distances and Cu–bridge–Cu angles perform as intuition-driven descriptors, which we call Ang+Dist throughout this work. Since it is known that small changes in bond distances (specifically between the central metal atoms) and angles can delicately affect J , and for the reason that such structural variations can not be encoded sufficiently in two-dimensional autocorrelation

functions⁵⁵, we decided to omit this descriptor from our study, despite the good performance for spin-crossover complexes⁵⁷.

2.3.1 Smooth Overlap of Atomic Positions (SOAP)

In the SOAP approach, the molecular structure is expressed in terms of the local electron density around a selected atom¹¹³,

$$\rho(\mathbf{r}) = \sum_{nml} c_{nml} g_n(\mathbf{r}) Y_{ml}(\mathbf{r}), \quad (7)$$

where g_n are radial basis functions, Y_{ml} are spherical harmonics, and c_{nml} coefficients. Rotationally invariant quantities can be constructed as a power spectrum or bispectrum. This descriptor, along with MBTR as introduced below, is presented in an appealing visual form in Ref. 97.

2.3.2 Many-Body Tensor Representation (MBTR)

In the many-body tensor representation, sums over broadened k -body terms m_k are sorted according to element types^{97,115}, where for $k = 1$, m_k corresponds to atomic charge numbers, for $k = 2$ to interatomic distances, and for $k = 3$ to angles between three atoms (or the cosine of that angle),

$$f_k(r, Z_1, Z_2, \dots, Z_k) = \sum_{i_1, i_2, \dots, i_k} \omega_k(i) \mathcal{N}(r|m_k(i)) \prod_{j=1}^k \delta_{Z_i Z_j}, \quad (8)$$

which is discretized in practice. $\mathcal{N}(r|m_k)$ is a normal distribution with mean m_k evaluated at r , ω_k a weighting function reducing the importance of terms involving far-apart atoms, and Z_i denotes an atomic charge number.

See Refs. 97, 116 for a visual representation and Refs. 115, 116 for further details on this descriptor.

2.3.3 Many-Body Interaction (F_{MB}) Descriptors

The F_{MB} approach¹¹⁴ to engineering descriptors is based on grouping interactions according to the element types of the atoms involved. The parameter M represents the type of interaction (e.g., $M = 2$ and $M = 3$ stand for pairwise and three-body interactions, respectively). Independent of the value of M , the set $\mathcal{A}_n = \{Z_i\}_{i=1}^n$ with $Z_i \neq Z_j \forall i, j \in \{1, 2, \dots, n\}$ is defined, which holds all atomic numbers (element types) that occur in a data set under consideration (n in total). In the following, we will focus exclusively on pairwise and three-body interactions. The set S_{2B} comprises all element pairs (Z_i, Z_j) with $Z_i \leq Z_j$ and $Z_i, Z_j \in \mathcal{A}_n$. By analogy, the set S_{3B} comprises all element triples (Z_i, Z_j, Z_k) with $Z_i \leq Z_j \leq Z_k$ and $Z_i, Z_j, Z_k \in \mathcal{A}_n$.

Given a pair $(Z_i, Z_j) \in S_{2B}$ and an order index $m \in \{1, 2, \dots, 15\}$, the corresponding element of the F_{2B} descriptor reads

$$F_{2B}^{i,j,m} = \sum_{\substack{a,b=1 \\ b>a}}^N \frac{\delta_{Z_i, \min(Z^{(a)}, Z^{(b)})} \cdot \delta_{Z_j, \max(Z^{(a)}, Z^{(b)})}}{r_{ab}^m} \quad (9)$$

Here, a and b refer to the a -th and b -th atoms, respectively, of an N -atom system separated by Euclidean distance r_{ab} . The corresponding atomic numbers are denoted $Z^{(a)}$ and $Z^{(b)}$, respectively. The Kronecker delta, δ , ensures that only atom pairs matching with the queried pair (Z_i, Z_j) are taken into account.

In the case of three-body interactions, all possible combinations of the order indices $m_1, m_2, m_3 \in \{1, 2, \dots, 6\}$ that fulfill the constraint $m_1 \neq m_2 \neq m_3 \neq m_1$ are considered. Given a triple $(Z_i, Z_j, Z_k) \in S_{3B}$ and the order indices m_1, m_2 , and m_3 , the corresponding element of the F_{3B} descriptor reads

$$F_{3B}^{i,j,k,m_1,m_2,m_3} = \sum_{\substack{a,b,c=1 \\ c>b>a}}^N \frac{\delta_{Z_i,\min(Z^{(a)},Z^{(b)},Z^{(c)})} \cdot \delta_{Z_j,\text{median}(Z^{(a)},Z^{(b)},Z^{(c)})} \cdot \delta_{Z_k,\max(Z^{(a)},Z^{(b)},Z^{(c)})}}{r_{ab}^{m_1} \cdot r_{ac}^{m_2} \cdot r_{bc}^{m_3}} \quad (10)$$

The definition of the F_{3B} descriptor is slightly different than the one presented in the original work by Pronobis *et al.*¹¹⁴ First, we do not multiply the terms of the sum by a prefactor, which is zero if at least one atom pair of a triple is separated by more than are threshold (approximately 1.1 times the average bond length of that element pair) and one otherwise. This choice increases CPU time but allows us to resolve weak three-body interactions. Second, we choose the constraint $Z_i \leq Z_j \leq Z_k$ instead of $Z_i \leq Z_k$ since three-body interactions are not encoded as bond angles. In addition, this choice leads to a significant reduction of dimensions and, hence, compensates for the increased CPU cost by dropping the prefactor.

2.3.4 Angle and distances (Ang+Dist)

Hay *et al.*¹⁰⁰ showed that there is a direct correlation between the Cu–ligand–Cu angle and J , and that exchange spin coupling is affected via direct exchange interaction through the distance between the spin centers. Motivated by this work, we study whether one can build a reasonable model using only Cu–bridge angles and Cu–Cu distances as features for J prediction.

In our case we define the Cu–bridge (θ) as the angle between a copper atom and its bridging atom (Figure 3), which is the closest atom of the bridge to a copper atom. Note that for bridging bidentate ligands, the bridging atoms can thus be different for both copper atoms (e.g., in case of AcO^- and ox^{2-} , see Figure 2). This can in principle also be extended to cases where there is only one bridging atom, or more than two. Together with the Cu–Cu distance, this leads in our case to a feature size of 5 dimensions in total.

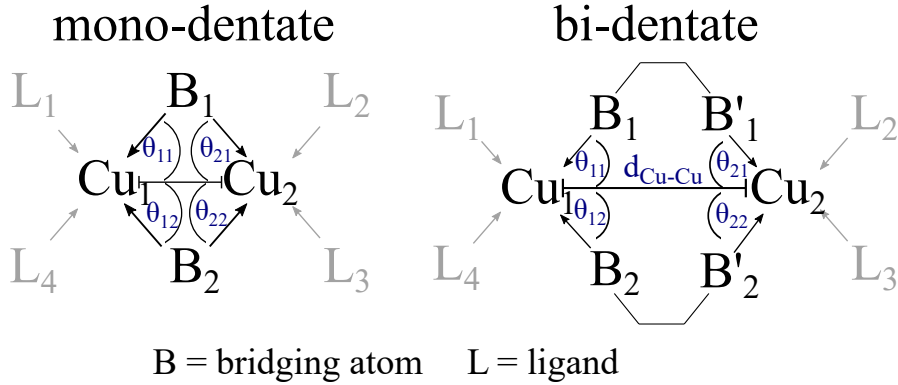


Figure 3: Schematic representation of a dicopper complex. θ is the angle between the Cu–Cu-axis and the bridging atom B. $d_{\text{Cu-Cu}}$ is the interatomic distance between both copper atoms.

2.4 Robust Evaluation of Exchange Coupling

When evaluating exchange coupling constants from DFT, often the antiferromagnetic state is modeled by a broken-symmetry determinant¹⁵, and J is evaluated from the energy difference between that determinant and the one describing the F coupled state. Since convergence of the self-consistent field algorithm is often tricky for the broken-symmetry determinant, this method does not lend itself to generating training data sets for machine learning. A more robust approach is based on the finding that for many spin-coupled systems, the total energy depends as a cosine function on the angle between the local spin vectors (Figure 4). In that case, it is possible to estimate the energy difference between ferromagnetically and antiferromagnetically coupled states from the curvature of the energy with respect to local spin rotation, allowing for estimating this energy difference from the electronic structure of one spin state only^{108,110,117–120}. We are employing a variant of this approach based on Green’s functions, which does not take into account orbital relaxation upon spin rotation, but which is very robust in its application¹¹⁰. For two coupled local spin-1/2 centers, J is related to the energy difference as $J = 1/2(E_{\text{AF}} - E_{\text{F}})$.

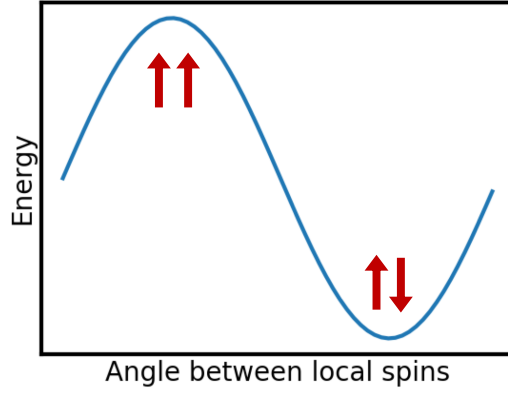


Figure 4: Total energy of a coupled two-spin system as a function of the angle between the two local spins, for the frequent case of a cosine dependence. The larger the energy difference between the ferromagnetically ($\uparrow\uparrow$) and antiferromagnetically ($\uparrow\downarrow$) states, the larger the curvature at either of the two extrema.

In the Green's function approach, J can be evaluated from the electronic structure of a Slater determinant describing the F coupled state as

$$\begin{aligned}
 J_G = & -\frac{1}{4S_A S_B} \sum_{\substack{i \in \text{occ} \\ k \in \text{virt}}} \sum_{\substack{\{\mu, \nu\} \in A \\ \{\mu', \nu'\} \in B}} C_{\nu i}^\alpha (F_{\mu\nu}^\alpha - F_{\mu\nu}^\beta) C_{\mu k}^{\beta*} \\
 & C_{\mu' i}^{\alpha*} \left(F_{\mu' \nu'}^\alpha - F_{\mu' \nu'}^\beta \right) C_{\nu' k}^\beta \frac{1}{\epsilon_i^\alpha - \epsilon_k^\beta} \\
 & -\frac{1}{4S_A S_B} \sum_{\substack{k \in \text{occ} \\ i \in \text{virt}}} \sum_{\substack{\{\mu, \nu\} \in A \\ \{\mu', \nu'\} \in B}} C_{\nu i}^\alpha (F_{\mu\nu}^\alpha - F_{\mu\nu}^\beta) C_{\mu k}^{\beta*} \\
 & C_{\mu' i}^{\alpha*} \left(F_{\mu' \nu'}^\alpha - F_{\mu' \nu'}^\beta \right) C_{\nu' k}^\beta \frac{1}{\epsilon_k^\beta - \epsilon_i^\alpha},
 \end{aligned} \tag{11}$$

where $F_{\mu' \nu'}^\sigma$ are the elements of the Fock matrices, $C_{\nu i}^\sigma$ the molecular orbital coefficients for a given spin σ in a Löwdin-transformed single-particle basis, and ϵ_i^σ the molecular orbital energies, and S_A and S_B refer to the local spin quantum numbers on the magnetic sites A and B .

Reducing the electronic structure to two electrons in two magnetic orbitals and employing configuration interaction¹⁰⁰ allows for a simplified qualitative understanding of J in terms of the energy splitting E^{mgap} between the two singly occupied molecular orbitals (SOMOs) in the

F-coupled state (triplet)^c,

$$J = k_{AB} - \frac{1}{2} \frac{(E^{\text{mgap}})^2}{\pi_{AA} - \pi_{AB}}, \quad (12)$$

with π_{AA} and π_{AB} are the Coulomb integrals resulting from repulsion between two electrons in the same and in two different magnetic orbitals ($\pi_{AA} \gg \pi_{AB}$), and k_{AB} the exchange integral between electrons in two magnetic orbitals k_{AB} . The (positive) exchange integral k_{AB} gives ferromagnetic contributions, while the second term gives antiferromagnetic contributions. Given this relation, we will compare machine-learning of E^{mgap} and J , expecting similar performance.

As a third quantity related to spin-dependent properties, we will explore machine learning of local unpaired electron densities on the copper atoms, which are defined as

$$N_A^s = \left(N_A^\alpha - N_A^\beta \right), \quad (13)$$

where N_A^σ denotes the number of electrons of spin σ assigned to a atom A .

2.5 Error Evaluation

We measure the performance of our models by the mean absolute error (MAE),

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \tilde{y}_i|. \quad (14)$$

Here, N is the number of samples (test data), y_i is the reference value of a given property (in this work J , α and μ) of test molecule i , and \tilde{y}_i is the predicted value of i . Because the MAE alone has to be interpreted with care as a performance estimator¹²¹, we use in addition the coefficient of determination,

^cThe magnetic orbitals can be defined as the local orbitals whose linear combination leads to the two SOMOs in the triplet.

$$R^2 = 1 - \frac{\sum_i (y_i - \tilde{y}_i)^2}{\sum_i (y_i - \bar{y})^2}, \quad (15)$$

where \bar{y} is the mean of all references y_i . In the case of perfect predictions and zero uncertainty in reference data, R^2 equals 1. If $R^2 = 0$, the prediction model performs just as well as the average (arithmetic) reference value. Negative R^2 -values indicate prediction models that perform even worse.

3 Methods

3.1 Data Set Construction and First-Principles Calculations

Exchange spin coupling depends systematically on bond angles and interatomic distances¹⁴. A machine-learning algorithm will need to have information on these angles and distances, either by implicitly learning them along with J from chemical compositions and connectivities, or by being explicitly provided with this information. In the following we will provide molecular structures optimized in the high-spin state with DFT (B3LYP). This is a good choice for the proof-of-principle of machine-learning J we are aiming at and also prevents possible drawbacks associated with experimental reference data (such as environment and temperature effects). In future practical applications, machine-learning J only makes sense if the benefit of replacing a conventional approach by ML is not marginal compared to the overall cost of obtaining molecular structures. Besides the implicit learning of molecular structures mentioned above, efficient ways of obtaining them could be the extraction from X-ray crystallographic or computational databases, or structural optimizations via cheap methods such as force-fields or tight-binding approaches¹²²⁻¹²⁶, or explicitly via ML models^{36,127}.

For evaluating the general feasibility and performance of machine-learning J with Gaussian process regression, we constructed a data set of $[\text{Cu}_2(\mu_2\text{-X})_2\text{X}_4]^{2-}$ and $[\text{Cu}_2(\mu_2\text{-Y})_2\text{X}_4]^{2-}$ complexes

inspired by Ref. 100 (Figure 2), comprising all possible combinations of halogens $X=\text{F,Br,Cl}$, and with bridges $Y=\text{acetate (AcO}^-)$, oxalate (ox^{2-}), methanolate (MeO^-) and hydroxide (OH^-). For these complexes, we generated input structures in an automated fashion and compared the eigenvalues of the Coulomb matrix to identify chemically similar structures, which were then removed. Afterwards, the structures were optimized employing Kohn–Sham density functional theory (KS-DFT) in the (ferromagnetically coupled) triplet state with BP86^{128,129} as the approximate exchange–correlation functional and def2-SVP¹³⁰ as the basis set, using the program package TURBOMOLE 7.1^{131,132}. Convergence criteria were set to 10^{-6} Hartree for the self-consistent-field algorithm and to 10^{-3} a.u. for the gradient in molecular structure optimizations. For $[\text{Cu}_2(\mu_2\text{-X})_2\text{X}_4]^{2-}$, planar and non-planar structures were generated (motivated by the fact that both configurations can be obtained experimentally and to gain further structural diversity). To ensure balance between different structure classes, we drew a random sample of 100 structures of this type to be included in the data set, which from now on we call the **[Cu₂]** set.

For molecular properties, single-point calculations in the ferromagnetically coupled (high-spin) state were performed employing KS-DFT with B3LYP^{133,134} as the approximate exchange–correlation functional and def-TZVP¹³⁵ as a basis set, using the program package GAUSSIAN09¹³⁶. Convergence criteria were set to 10^{-6} Hartree for the self-consistent-field algorithm. Exchange coupling constants were evaluated based on a Green’s function approach¹¹⁰ with the copper atoms defined as spin centers, with ideal local spin quantum numbers of $S_A = S_B = 1/2$, employing our program package ARTAIOS¹³⁷. Dipole moments μ and polarizabilities α were extracted from the GAUSSIAN output. Strictly speaking, properties of the antiferromagnetically coupled complexes should be evaluated in their antiferromagnetically coupled ground state. Since our goal is to prevent the calculation of the broken-symmetry determinants modeling those ground states, and since we are interested in a proof-of-principle, we are focusing on the properties in the high-spin state here. These DFT-generated (B3LYP/TZVP) properties were used as references.

The [Cu₂] set was then split into a 80% training set (**Cu-train**) and a 20 % test set (**Cu-test**), as documented in the Supporting Information. To this end, all molecules were represented using the F_{2B} descriptor and then grouped into 25 clusters (via k -means clustering), from which 80 % of the molecules within each cluster were randomly assigned to the training set. This procedure ensures that the entire data set is evenly divided.

3.2 Descriptors, Kernels, and Gaussian Process Regression

The molecular descriptors SOAP¹¹³ and MBTR^{97,115} were employed as implemented in the python module DESCRIBE¹¹⁶, with spherical Gaussian-type orbitals as radial basis functions (RBFs) in SOAP. The cutoff around the local region was set to 6.0 Å, the number of RBFs was set to 8 and the maximum degree of the spherical harmonics was set to 6. For density expansion, the standard deviations of the Gaussians were kept at the default value of 1.0. In our case, the SOAP output (the power spectrum) of DESCRIBE was used as feature vector, which has been averaged over all atom types. Other input settings were kept as default. For MBTR, the one-body term ($k = 1$) contains the atomic species of a molecule represented on a grid between 0 and 35 (100 points). Pairwise inverse distances were selected for the two-body term ($k = 2$), whereas for the three-body term ($k = 3$) the cosine of the angles between three atoms is evaluated. For both, $k = 2$ and $k = 3$, an exponential weighting function was applied with a scaling parameter $s = 0.5$ and a cutoff value of 0.001. For each k term, the output was normalized such that the length (norm) of each k -vector is one.

Many-body interaction descriptors¹¹⁴ (F_{MB}) were calculated with a Python script developed in our groups.

GPR and linear ridge regression models were build using the Python package scikit-learn¹³⁸. In all GPR cases, the same length-scale parameter was used for all feature dimensions. In this study, the Matérn (1/2), Matérn (3/2) and the radial basis function (RBF) kernel (=Matérn ($\nu \rightarrow \infty$)) are compared, and used as implemented in scikit-learn. Hyperparameters are op-

timized using the limited-memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm for bound constrained optimization as also implemented in scikit-learn. The optimization was restarted 15 times from different values, to avoid being trapped in a local minimum. The Python scripts developed for and employed in this work will be made publicly available (see Supporting Information).

4 Model Building for Regression-Based J Prediction

Whether GPR can make sufficient predictions for J , and whether it performs better compared to a simple linear ridge regression, shall be analyzed by comparing the errors (MAE and R^2) among the selected models under consideration here for predictions on **Cu-test**.

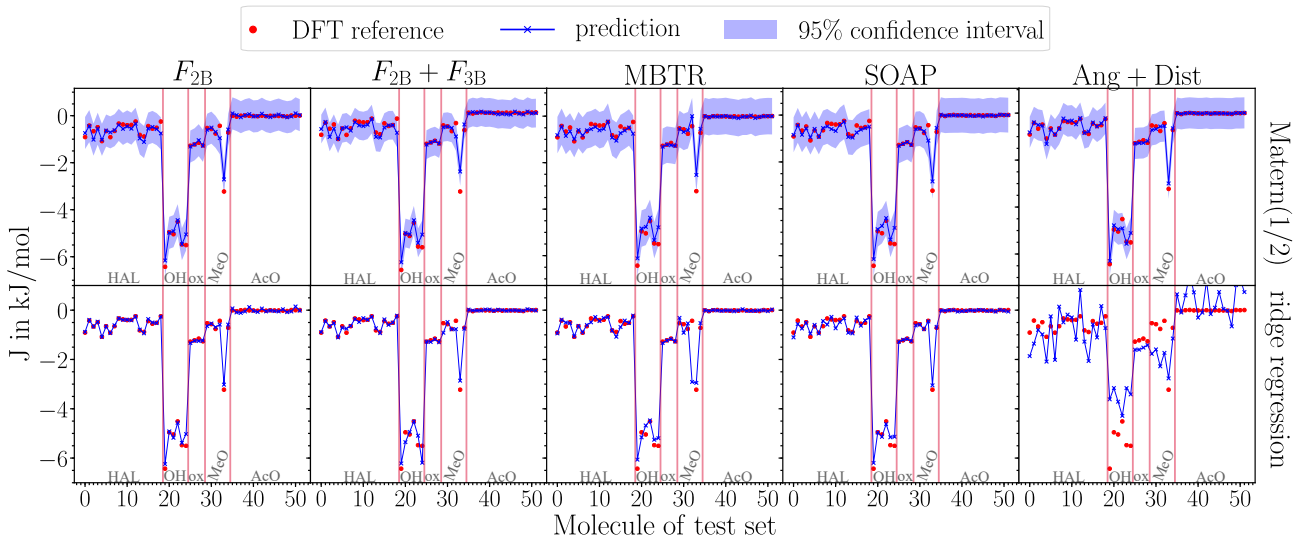


Figure 5: Reference (DFT) and predicted J values for **Cu-test**, using GPR and linear ridge regression for the underlying model building/training. For models built with GPR, we exemplarily show the Matérn(1/2) kernel function.

Fitting the models using **Cu-train** and subsequent prediction for **Cu-test** allows for an evaluation of the MAE according to the different types of Cu-Cu-bridges (consider Table 1, and, for selected models, Figure 5). Such an analysis can become important for judging whether there is more training data necessary for one bridging type or not, in the sense that outliers in

the MAE (for one class of molecules) are easier to identify compared to considering the MAE calculated over the entire test set.

Table 1: MAEs obtained from different GPR models and linear ridge regression for **Cu-test** ordered according to bridge types. In addition, MAE and R^2 calculated over the entire **Cu-test** set are provided. MAE Values in kJ/mol. Color code: yellow: $0.50 < R^2 < 0.90$, green: $R^2 > 0.90$.

	HAL	OH	ox	MeO	AcO	MAE	R^2
linear ridge reg.							
F_{2B}	0.02	0.10	0.06	0.11	0.06	0.06	1.00
$F_{2B} + F_{3B}$	0.05	0.22	0.03	0.24	0.02	0.08	0.99
MBTR	0.08	0.24	0.02	0.71	0.02	0.13	0.95
SOAP	0.12	0.15	0.01	0.06	0.02	0.08	0.99
Ang + Dist	0.62	1.70	0.36	1.11	0.56	0.75	0.65
Matérn(1/2)							
F_{2B}	0.12	0.10	0.05	0.25	0.05	0.11	0.99
$F_{2B} + F_{3B}$	0.12	0.14	0.02	0.32	0.04	0.12	0.98
MBTR	0.12	0.20	0.05	0.28	0.02	0.12	0.99
SOAP	0.11	0.19	0.02	0.24	0.02	0.11	0.99
Ang+Dist	0.07	0.18	0.06	0.17	0.01	0.07	1.00
Matérn(3/2)							
F_{2B}	0.10	0.29	0.02	0.19	0.02	0.10	0.98
$F_{2B} + F_{3B}$	0.04	0.14	0.08	0.27	0.02	0.08	0.99
MBTR	0.04	0.21	0.04	0.23	0.011	0.08	0.99
SOAP	0.06	0.12	0.02	0.10	0.01	0.07	0.99
Ang+Dist	0.13	0.62	0.03	0.18	0.01	0.15	0.97
RBF							
F_{2B}	0.12	0.23	0.04	0.15	0.10	0.12	0.99
$F_{2B} + F_{3B}$	0.06	0.17	0.11	0.38	0.03	0.10	0.99
MBTR	0.03	0.21	0.04	0.23	0.01	0.08	0.99
SOAP	0.06	0.19	0.02	0.10	0.01	0.06	0.99
Ang+Dist	0.35	1.81	0.08	0.29	0.01	0.41	0.72

From Figure 5 and Table 1, it can be seen that with linear ridge regression, one obtains reasonable predictions with all $R^2 > 0.9$ when $F_{2B} + F_{3B}$, SOAP or MBTR are used as descriptors. The observation that a linear model (such as employed in ridge regression) results in reasonable predictions for **Cu-test** may be caused by the highly engineered descriptors which encode structural features in a nonlinear fashion. In case of our intuition-driven descriptor (Ang+Dist),

which in its generation contains no further (non-linear) transformations, one gets rather poor predictions of J for **Cu-test**. This suggests that although J depends systematically on bond angles and distances between the spin-centers, there is no linear structure–property relationship between J and these structural parameters. Note, however, that the Ang+Dist descriptor will be identified below as the considerably best performing one for extrapolation to experimental structures outside of our dataset (Section 6).

Within GPR, all combinations of the descriptors SOAP, MBTR, F_{2B} and F_{3B} with all kernel functions under study result in R^2 values of 0.98 to 0.99 and MAEs of 0.06 kJ/mol to 0.12 kJ/mol, from which one would conclude that there is only little effort needed for selecting a proper GPR model for J prediction when these descriptors are used. For the Ang+Dist descriptor, the results depend significantly on the choice of kernel function. For RBF \rightarrow Matérn(3/2) \rightarrow Matérn(1/2) the MAE decreases from 0.41 kJ/mol to 0.07 kJ/mol (at the same time the R^2 increases from 0.72 to 1.00), suggesting that when Ang+Dist is selected as descriptor the best model can be obtained using Matérn(1/2).

The observations so far suggest that in principle a linear model is sufficient for J prediction, at least for the **Cu-test** molecules when a proper descriptor is selected to featurize the molecules. It was already shown in Ref. 114 that for a certain property, and a suitable choice of a descriptor, linear models show reasonable prediction quality compared to more sophisticated ML techniques, such as kernel-ridge-regression (KRR). In that work, the prediction of atomization energies using linear ridge regression results in comparable MAEs (and RMSEs) as with KRR, whereas the overall best results are obtained with the $F_{2B}+F_{3B}$ descriptor. In our case, the best linear model is obtained when ridge regression is used in combination with the F_{2B} descriptor, which performs slightly better than when F_{2B} and F_{3B} are used. This can probably come from the fact that the addition of F_{3B} significantly increases the feature space, which can lead to worse results when only a small number of samples (molecules in the training set) is available.

5 Comparison with Other Molecular Properties

In Section 4, it was shown that for the **[Cu₂]** set it is possible to train proper prediction models using GPR and linear ridge regression. In this part, we study whether this is also possible for properties intensively used in the literature to benchmark ML methods such as the dipole moment μ or the polarizability α of the **[Cu₂]** set, and how these models behave in comparison to J prediction. Such a comparison is important to ensure that our findings are not specific to exchange spin coupling. To this end, prediction models were built in analogy to J using α and μ as obtained from B3LYP/TZVP as references.

Considering the R^2 values (Table 2) as obtained from predictions for **Cu-test**, it becomes clear that for the polarizability all GPR models under study here, except for MBTR and Ang+Dist as descriptors, work similarly well (as also observed for J), as all values are close to 1. For linear ridge regression, it turns out that SOAP, F_{2B} and $F_{2B}+F_{3B}$ gives results that are as good as those obtained with GPR. Thus, with linear models and GPR, the predictions of the properties J and α behave very similarly, except that the descriptor (Dist+Ang) generated from chemical intuition for J behaves poorly for α prediction.

For the dipole moment, the prediction based on linear ridge regression works only when the F_{2B} descriptor is used. One obtains a MAE of 0.64 D with an R^2 value of 0.74. Interestingly, with all other descriptors under study, the R^2 value is close to zero or negative, suggesting that no satisfying prediction for **Cu-test** is possible. Using GPR, it turns out that all models which are not built with Ang+Dist as descriptor yield a positive R^2 , but significantly lower than 1. In the case of F_{2B} /RBF, the value of R^2 is as low as 0.03, which indicates that for μ prediction, one cannot rely on this model. The Ang+Dist descriptor is, again, not designed to make any predictions for μ , which is why it is not surprising that it fails for μ .

Table 2: MAE (R^2) for J , α and μ prediction for **Cu-test**, after training on **Cu-train**. Color code: red: $R^2 < 0.0$, orange: $0 < R^2 < 0.50$, yellow: $0.50 < R^2 < 0.90$, green: $R^2 > 0.90$.

	J [kJ/mol]	α [Bohr ³]	μ [D]
linear ridge reg.			
F_{2B}	0.06 (1.00)	1.26 (0.97)	0.64 (0.74)
$F_{2B} + F_{3B}$	0.08 (0.99)	1.34 (0.98)	1.07 (0.15)
MBTR	0.13 (0.95)	578 (-1·10 ⁴)	1971 (-7.6·10 ⁷)
SOAP	0.08 (0.99)	2.70 (0.96)	0.86 (0.48)
Ang+Dist	0.75 (0.65)	19.2 (0.49)	1.30 (0.03)
Matérn(1/2)			
F_{2B}	0.11 (0.99)	2.57 (0.96)	0.83 (0.53)
$F_{2B} + F_{3B}$	0.12 (0.98)	3.49 (0.95)	0.91 (0.50)
MBTR	0.11 (0.99)	6.76 (0.54)	1.02 (0.37)
SOAP	0.11 (0.99)	2.40 (0.96)	0.75 (0.68)
Ang+Dist	0.07 (1.00)	13.1 (0.70)	1.17 (0.22)
Matérn(3/2)			
F_{2B}	0.10 (0.98)	2.30 (0.97)	0.95 (0.37)
$F_{2B} + F_{3B}$	0.08 (0.99)	1.67 (0.98)	0.84 (0.54)
MBTR	0.08 (0.99)	5.95 (0.56)	0.99 (0.39)
SOAP	0.07 (0.99)	1.36 (0.97)	0.70 (0.72)
Ang+Dist	0.15 (0.97)	17.0 (0.51)	1.42 (-0.23)
RBF			
F_{2B}	0.12 (0.99)	2.50 (0.96)	1.12 (0.03)
$F_{2B} + F_{3B}$	0.10 (0.99)	1.38 (0.98)	0.84 (0.50)
MBTR	0.08 (0.99)	6.11 (0.56)	1.03 (0.37)
SOAP	0.06 (0.99)	1.64 (0.97)	0.80 (0.63)
Ang+Dist	0.41 (0.72)	28.2 (-0.28)	2.14 (-1.53)

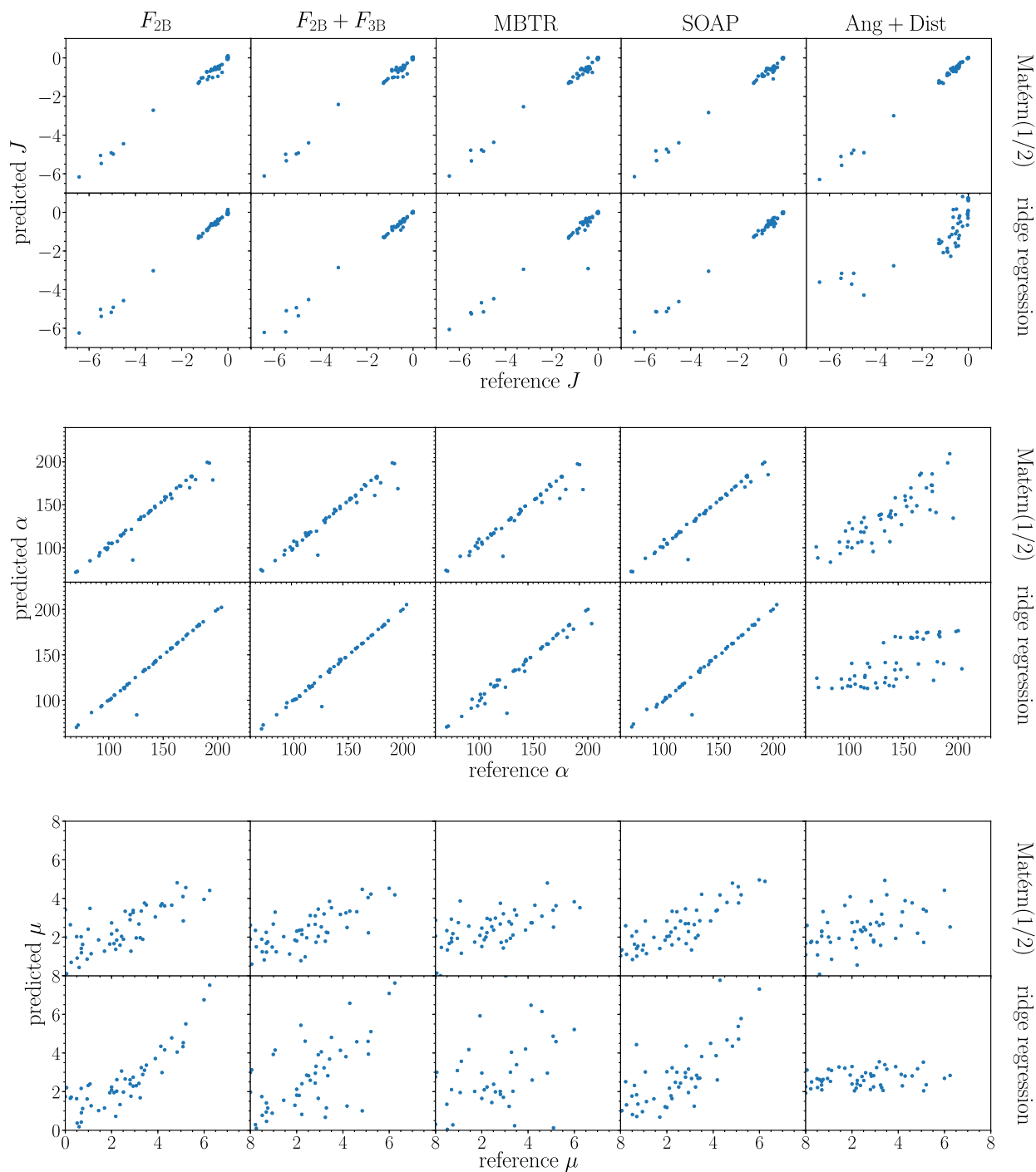


Figure 6: Reference versus predicted values for J , α and μ for the **Cu-test** set. Predictions were done with models trained from linear ridge regression and GPR. For the latter we only show the results obtained from the Matérn(1/2) kernel.

Furthermore, it is noticeable that for all kernel functions, the SOAP descriptor performs best for predicting μ , which when combined with Matérn(3/2) yields the overall best performance

(MAE: 0.70 kJ/mol, R^2 : 0.72) using GPR in our case. Yet, it is slightly worse than when linear ridge regression is used with F_{2B} as a descriptor (MAE: 0.64 kJ/mol, R^2 : 0.74), which is an interesting observation, because when a proper descriptor is selected, a linear model can perform similarly well as the more sophisticated GPR.

Among the different properties, the prediction of μ seems to work in general worse than for J and α , which is reflected in the lower R^2 values and when considering the predicted versus reference values (see Fig 6). This difference between α and μ is in good agreement with results from other data sets studied in the literature (e.g., see Figure 6 of Ref. 114). This suggests that our conclusions are not specific for our choice of data set.

To explore this aspect further, we applied the same strategy as for α and μ to train linear and GPR models for the quite popular data set of 6095 constitutional isomers of $C_7H_{10}O_2$ published by Ramakrishnan *et al.*¹³⁹. Here we also found that the prediction of α works better than for μ , and is less sensitive to the selected kernel function for the underlying GPR (though the RBF kernel seems to perform worse for this data set as compared to Matérn(1/2) and Matérn(3/2). Furthermore, it is also observed that for μ prediction with a suitable choice of the descriptor ($F_{2B}+F_{3B}$), the overall best result is obtained with linear ridge regression, which matches with the findings for the [Cu₂] set in the sense that a linear model can perform as well as GPR for μ prediction. A more detailed analysis of our data for the 6095 constitutional isomers of $C_7H_{10}O_2$ can be found in the Supporting Information.

6 Extrapolation to Experimental Structures

Our discussion so far would suggest that simple linear ridge regression is sufficient for predicting J , rather than requiring a more complex GPR model. This may be related to the unknown structures (**Cu-test**) being similar to those used during fitting (which is ensured when the data set is split in a non-biased way). In the following, it shall thus be studied how well the

different models work when extrapolating to experimentally observed structures that are not closely resembling any structures in **Cu-train**.

To this end, we selected the dinuclear Cu complexes $[\text{Cu}(\text{teen})\text{OH}]_2^{2+}$ (**teen**), $[\text{Cu}(\text{EAEP})\text{OH}]_2^{2+}$ (**EAEP**) and $[\text{Cu}(\text{bipy})\text{OH}]_2^{2+}$ (**bipy**), Clinoatacamite (**Clino**) and Tolbachite (**Tolba**) (Figure 7)^d, whose properties were investigated in experiment and theory in previous studies^{100,140–144}. For all experimental molecules, predictions were made in their DFT-optimized molecular structures with all trained models. Since we found that in general the choice of the kernel function is only important for the Ang+Dist descriptor, and since Matérn(1/2) performs best for this descriptor, we limit the discussion to this kernel (the results of the remaining kernel functions are provided in the Supporting Information).

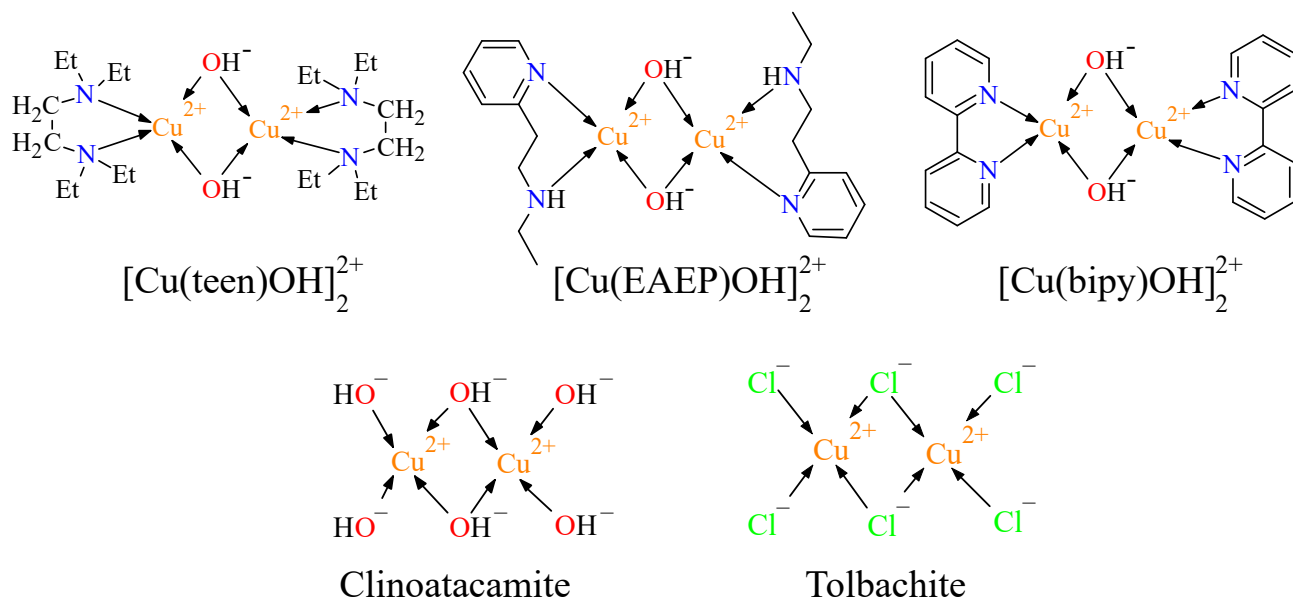


Figure 7: Schematic illustration of experimentally studied dinuclear copper complexes (or cut-outs from mineral structures) used in this study to evaluate the extrapolation performance of GPR and linear ridge regression. The DFT-optimized molecular structures can be found in the Supporting Information.

^dNote that we neglected the out-of-plane Cu ligands for **Clino** and **Tolba**, i.e. in the minerals the Cu atoms have an octahedral coordination.

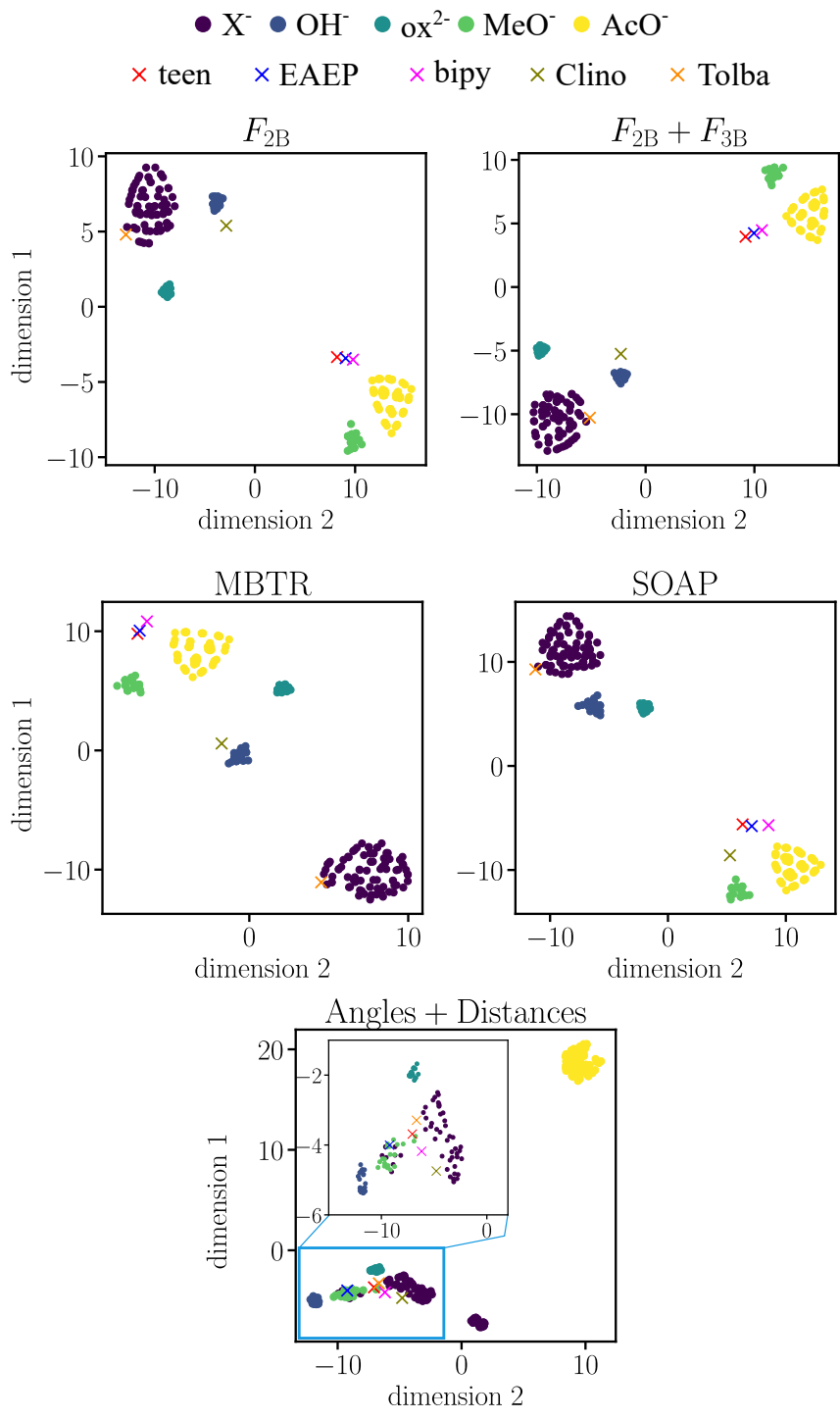


Figure 8: t-SNE distribution of the training data and the experimental structures. In all cases a perplexity value of 50 was used. Distributions with other perplexity values can be found in the Supporting Information.

Comparing the experimental J_s (Table 3) of **teen**, **EAEP** and **bipy** with the results obtained

from DFT using the Green’s function approach to calculate J , it becomes clear that reproducing experimental trends can be a challenging task. This is partly because of the missing crystal effects and counter-ions in our DFT simulations. Furthermore, errors due to the approximated nature of the exchange–correlation functional used in DFT are to be expected. Both are not affecting our goal in this work, the prediction of DFT reference values with a regression model.

The DFT references (Table 3) suggest that **teen** and **EAEP** are supposed to show the strongest antiferromagnetic coupling ($J = -1.21$ kJ/mol), followed by **bipy** and **Tolba** (both with $J = -0.74$ kJ/mol). **teen** has the weakest AF coupling with -0.46 kJ/mol, and **Cliño** is suggested to have almost no exchange spin coupling.

Table 3 summarizes the J values predicted for the experimental structures (in their DFT-optimized molecular structures) employing linear ridge regression and GPR (Matérn (1/2) only) with all descriptors under study. With ridge regression, one obtains no reasonable results, as in some cases the absolute values of J are significantly overestimated, and in some cases the model even predicts the incorrect sign (with SOAP and Ang+Dist being at least in the right ballpark, and with the good performance for **Tolba** being related to its closeness to the training set). This suggests that for extrapolation to structures clearly different from the training data, a linear model is not sufficient, although prediction for a test set chemically close to the training set looks promising (see Section 4).

Within GPR (Matérn(1/2)), the F_{2B} , $F_{2B}+F_{3B}$, MBTR and SOAP descriptors are not capable of predicting reasonable J s with sufficiently low errors for all molecules except for **Tolba**. This structure is clearly the most similar one to the **[Cu₂]** set, and thus it is not surprising to obtain good predictions and small errors for it. In fact, considering a t-SNE distribution¹⁴⁵ (see Figure 8), for all mentioned descriptors **Tolba** lies close to the halogene-bridged molecules in the training set. The remaining molecules are somewhat apart from the training data, which is probably the reason for the poor prediction and rather large errors.

The descriptor guided by chemical intuition (Ang+Dist) yields the best results using GPR

(Matérn(1/2)) when predictions on structures not similar to the test set are done. Except for **Clino**, the predicted J s are close to the reference values and, at the same time, have sufficiently low errors (within the error bar all predictions fit to the reference). The rather good predictions and low error bars possibly result from an appropriate distribution of the data using the Ang+Dist descriptor (see Figure 8). This distribution also suggests why the prediction for **Clino** works rather poorly with the Ang+Dist descriptor: data point is slightly further apart from the training data than the remaining experimental structures, and this is consistently observed for different perplexity values (see Supporting Information). Furthermore, it agrees with the fact that for **Clino**, none of the Euclidean distances to the different groups of Cu–Cu bridges (Figure 9)^e is as small as for the remaining experimental systems. The origin for this observation might be that **Clino** shows a distinct nonplanar structure (see Supporting Information), which is not captured within the **[Cu₂]** set. For future applications, such issues might be solved with active learning strategies^{78,146}. Overall, although Ang+Dist has only a dimension of 5 features, it captures the essential information correlated with J . Using artificial neuronal networks, it was shown that such sparse descriptors can also give accurate results for other properties of transition metal complexes¹⁴⁷.

^eWe employed the center of the cluster of a given Cu–Cu bridge to calculate the distances.

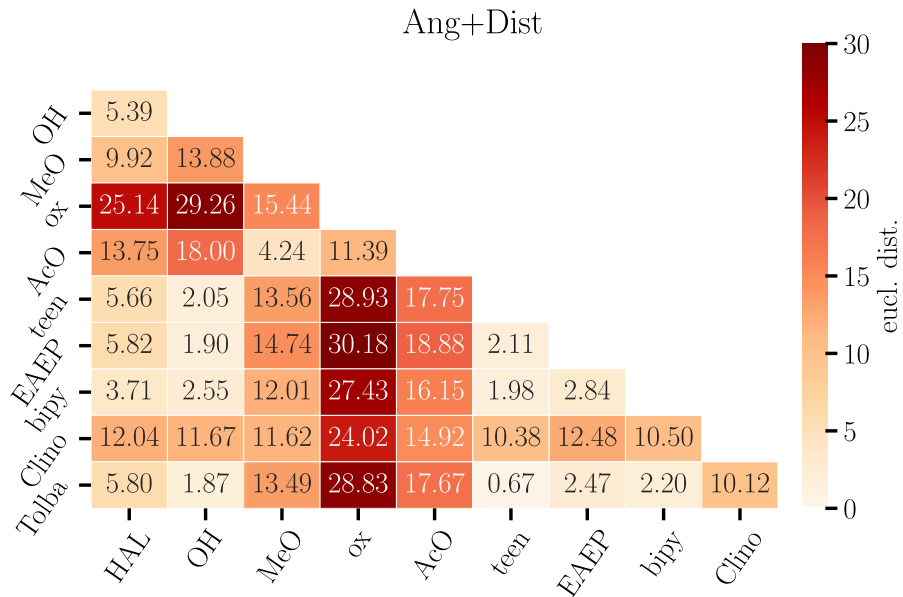


Figure 9: Euclidean distances between the experimental structures and the center of a given bridge type, calculated from the Ang+Dist descriptor.

To capture the qualitative trends correctly, i.e. that for **teen** a smaller AF coupling is predicted than for **bipy** and **Tolba**, probably requires further model tuning. Nevertheless, the results suggest that a descriptor guided by simple chemical intuition can lead to reasonable results. Such information (Cu–Cu distance and bridging angles) is likely also encoded in descriptors such as SOAP, MBTR or F_{2B} (F_{3B}), but might require more training data to make the model recognize these features.

Table 3: J for experimental structures (see Figure 7) in kJ/mol, as obtained from DFT and as obtained from predictions from linear ridge regression and GPR. In addition, for GPR the standard deviations are provided. Predictions were done on DFT (BP86/def2-SVP)-optimized structures.

	teen	EAEP	bipy	Clino	Tolba
exp. ¹⁰⁰	-1.23	-0.39	0.52	-	-
DFT					
B3LYP/TZVP	-0.46	-1.21	-0.74	0.00	-0.74
linear ridge reg.					
F_{2B}	2.86	137.79	270.96	47.94	-0.79
$F_{2B} + F_{3B}$	76.18	732.04	$2 \cdot 10^3$	-19.14	-0.73
MBTR	$1 \cdot 10^7$	$4 \cdot 10^6$	$2 \cdot 10^6$	$4 \cdot 10^6$	-0.52
SOAP	-3.46	-0.56	-1.42	-0.88	-0.74
Ang. + Dist.	-1.36	-2.10	-1.33	2.71	-1.25
Matérn(1/2)					
F_{2B}	-0.71 ± 2.38	-0.90 ± 2.03	-0.93 ± 2.35	-4.76 ± 1.93	-0.79 ± 0.74
$F_{2B} + F_{3B}$	-0.64 ± 4.88	0.00 ± 5.49	0.00 ± 5.49	-3.69 ± 2.01	-0.91 ± 0.83
MBTR	0.00 ± 5.08	0.00 ± 5.08	0.00 ± 5.08	0.00 ± 5.08	-0.78 ± 0.72
SOAP	-0.71 ± 3.72	-0.79 ± 3.33	-0.81 ± 3.01	-2.75 ± 0.83	-0.92 ± 0.51
Ang. + Dist.	-0.62 ± 0.50	-1.18 ± 0.24	-0.64 ± 0.46	-0.20 ± 1.38	-0.50 ± 0.44

Since extrapolation to structures not captured in a training data set is of practical importance, we have shown that this is in principle possible using GPR. Our results suggest that, at least in the case of few training data points, a descriptor based on chemical intuition outperforms other descriptors which contain more information (in terms of features) about the molecular systems under study. It is known that ML can suffer from too many features compared with the number of samples^{148,149}, which could be one reason for why the GPR models trained with the more sophisticated descriptors perform poorly when it comes to extrapolation. In further studies, we will address this issue.

7 Conclusion and Outlook

We have explored the machine-learnability of Heisenberg exchange spin coupling constants J for a series of 257 small doubly bridged dicopper complexes, employing Gaussian process regression and a selection of different kernel functions (Matérn(1/2), Matérn(3/2), radial basis functions) and molecular descriptors (F_{MB} , MBTR, SOAP, angles + distances). We find that the performance of GPR is nearly independent of the choice of kernel and descriptor, with MAEs on the order of 0.1 kJ/mol (J ranges from -6.50 to 0.03 kJ/mol), with the exception of the angles+distances descriptor yielding much larger errors. Interestingly, a simple linear ridge regression performs equally well, with the exception of the simplest descriptor consisting of only Cu-bridge-Cu angles and Cu-Cu distances, suggesting that any nonlinearities in structure-property relations are taken care of by the more sophisticated descriptors. However, when extrapolating to experimental dicopper complexes clearly outside of our training/test data set, only the angles + distances descriptor is capable of making reliable predictions in combination with GPR, and all linear models fail (sometimes dramatically). In other words, a compact descriptor incorporating chemical intuition (the relation between bond angles / distances and J is known since the 1950s¹²), while not capable of linearizing the problem, outperforms several sophisticated state-of-the-art descriptors when it comes to extrapolation.

This finding is indeed restricted to J , as would be expected from the fact that the intuition-based angles + distances descriptor is tailored to correlate with this property. For polarizabilities and dipole moments, all descriptors perform equally poorly for extrapolation. As far as predictions for the test set drawn from the small dicopper molecules are concerned, J interestingly behaves very similar to the electric polarizability α , with prediction quality being high for both. This may be rationalized by both properties encoding how the electronic structure responds to perturbations, and the evaluation of both depending on unoccupied molecular orbitals. For the dipole moment μ , representing a purely ground-state property, all regression models under consideration do not perform well. This distinction between α and μ was also found for a data

set of 6095 constitutional isomers of $C_7H_{10}O_2$ frequently studied in the literature, suggesting that this is a general trend and not specific to our dicopper data set.

Overall, our findings highlight the crucial importance of the descriptor, which, alone or in combination with a kernel function, encodes chemical structure–property relationships. Our data also suggest that chemical intuition can play a central role in finding suitable descriptors for a given problem, even though systematic and automated feature selection^{91,150–152} — on its own or combined with such intuition — has proven promising for spin-dependent properties¹⁵⁰ and may be an exciting avenue to explore more deeply in chemistry and materials science.

Acknowledgement

We thank Ömer Sahin for valuable discussions. Financial support of this research by the German Research Foundation (DFG) via projects HE 5675/6-1 (MPB, NM, and CH) and 389479699/GRK2455 (JP) is acknowledged. The authors thank the High-Performance Computing Center at University of Hamburg for computational resources.

Supporting Information Available

Python scripts to reproduce our results (including readily available descriptors), Cartesian coordinates and properties, division of complexes into test and training data set, data for further kernels and properties (α, μ) for extrapolation to experimental structures, t-SNE data with different perplexity factors, and data for the set of 6095 constitutional isomers of $C_7H_{10}O_2$ are provided in the Supporting Information.

A Gaussian Process Regression

A.1 Conceptual Introduction

For the simplest case of one training input \mathbf{x} with functional value $f(\mathbf{x})$ and one test data point \mathbf{x}_* with functional value $f_* = f(\mathbf{x}_*)$, a bivariate Gaussian distribution is assumed,

$$\begin{pmatrix} f \\ f_* \end{pmatrix} = \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} K & K_* \\ K_* & K_{**} \end{pmatrix} \right) \quad (16)$$

with a mean of zero and a covariance matrix \mathbf{K} . \mathbf{K} has the entries $K = k(\mathbf{x}, \mathbf{x})$, $K_* = k(\mathbf{x}, \mathbf{x}_*) = k(\mathbf{x}_*, \mathbf{x})$, $K_{**} = k(\mathbf{x}_*, \mathbf{x}_*)$, where k is a kernel function which encodes the correlation between chemical properties depending on distances in molecular-structure space (see below). $k(\mathbf{x}, \mathbf{x}_*)$ reaches a maximum if \mathbf{x} is equal to \mathbf{x}_* and approaches zero if \mathbf{x}_* is very different from \mathbf{x} . Therefore, if two inputs, i.e., two molecular structures, are very different, \mathbf{K} will be a unit matrix, resulting in a perfect bell-shape of the distribution (see left-hand side of Figure 10). The idea behind GPR is that if we choose a certain training input \mathbf{x} , we obtain a conditional distribution $p(f_* | \mathbf{x}_*, \mathbf{x}, f)$ which is also Gaussian, like cutting through an actual bell (the red lines in Figure 10). The mathematical relations for the mean (in this case zero) and the variance of $p(f_* | \mathbf{x}_*, \mathbf{x}, f)$ are established, including their generalization to higher dimensions^{31,76}, so that both the most likely functional value for a given test data point \mathbf{x}_* and its variance are known. If the test data point is close to a training data point, $k(\mathbf{x}, \mathbf{x}_*)$ will become nonzero, resulting in a distorted bivariate Gaussian (see right-hand side of Figure 10). This implies that f_* and f are correlated, pulling the prediction for f_* closer to $f(\mathbf{x})$ and reducing the variance.

\mathbf{x}_* and \mathbf{x} far from each other

\mathbf{x}_* and \mathbf{x} close to each other

$$\mathbf{K} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\mathbf{K} = \begin{pmatrix} 1 & -0.7 \\ -0.7 & 1 \end{pmatrix}$$

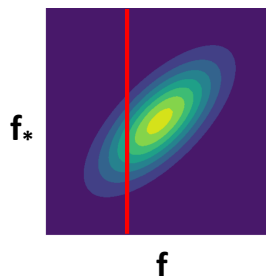
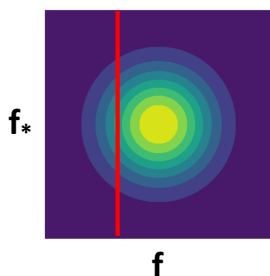
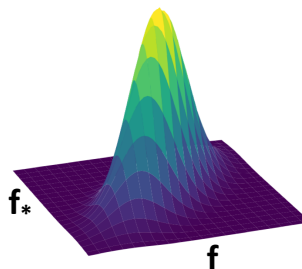
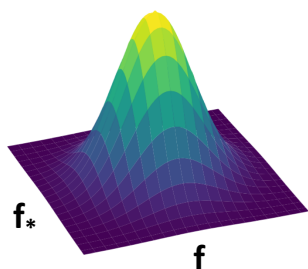


Figure 10: Illustration of how the covariance matrix \mathbf{K} (with elements determined by the kernel) determines the distribution of f_* if a the functional value f for a training input is known (red line in the bottom plots), depending on whether the test molecular structure encoded via a descriptor as \mathbf{x}_* is far from a molecular structure \mathbf{x} in the training set (left-hand side) or close to it (right-hand side). In other words, \mathbf{x} and \mathbf{x}_* determine what the bivariate Gaussian looks like, and $f = f(\mathbf{x})$ determines where to cut it, and all three together determine the most likely value and variance for f_* . Normalization has not been considered in these plots.

A.2 Further Information

The coefficients $\{\alpha_i\}$ of Eq. (2) are determined from the $(N \times N)$ -dimensional covariance matrix of the training set, \mathbf{K} , the noise variance, σ_n^2 , and the N -dimensional vector of training labels, \mathbf{y} ,

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} = (\mathbf{K} + \sigma_n^2 \mathbf{I}_N)^{-1} \mathbf{y} \quad (17)$$

where \mathbf{I}_N is the $(N \times N)$ -dimensional unity matrix. While the coefficients $\{\alpha_i\}$ are global, the coefficients $\{\beta_i^*\}$ of Eq. (3) depend on the queried molecule, represented by input vector \mathbf{x}_* ,

$$\boldsymbol{\beta}^* = \begin{pmatrix} \beta_1^* \\ \vdots \\ \beta_N \end{pmatrix} = (\mathbf{K} + \sigma_n^2 \mathbf{I}_N)^{-1} \mathbf{k}_* \quad (18)$$

where $\mathbf{k}_* = \mathbf{k}(\mathbf{X}, \mathbf{x}_*)$ is an N -dimensional covariance vector that measures similarity between the queried molecule and the N molecules of the training set,

$$\mathbf{k}_* = \begin{pmatrix} k(\mathbf{x}_1, \mathbf{x}_*) \\ \vdots \\ k(\mathbf{x}_N, \mathbf{x}_*) \end{pmatrix} \quad (19)$$

The GPR model studied here comprises three hyperparameters: the signal variance, σ_f^2 , and the length scale, l —see Eqs. (4), (5), and (6)—as well as the noise variance, σ_n^2 . The hyperparameters implicitly determine the hypothesis space, which can be interpreted as the space of possible functions. The actual regression step, cf. Eq. (2), which takes into account the training data, weights those functions according to how likely they are to explain the data. In GPR, hyperparameter optimization is performed in a Bayesian fashion, i.e., by maximizing the marginal likelihood $p(\mathcal{D} | \sigma_f^2, l, \sigma_n^2)$, which is the probability of observing the training data conditioned on the specific values of the hyperparameters,

$$\text{opt}(\sigma_f^2, l, \sigma_n^2) = \arg \max_{\sigma_f^2, l, \sigma_n^2} p(\mathcal{D} | \sigma_f^2, l, \sigma_n^2) \quad (20)$$

Assuming likelihood and prior to be Gaussian (as applies here), the marginal likelihood is Gaussian, too, which can be expressed in closed form,

$$\log p(\mathcal{D} | \sigma_f^2, l, \sigma_n^2) = -\frac{1}{2} \mathbf{y}^\top (\mathbf{K} + \sigma_n^2 \mathbf{I}_N)^{-1} \mathbf{y} - \frac{1}{2} \log |\mathbf{K} + \sigma_n^2 \mathbf{I}_N| - \frac{N}{2} \log 2\pi \quad (21)$$

B The Molecules under Study

The $[\mathbf{Cu}_2]$ input structures (Figure 2) were generated as described in Section 3.1. For all molecules, the copper atoms have an oxidation state of +II, resulting in (formally) one unpaired electron per copper atom. The two unpaired electrons can be coupled ferromagnetically (triplet) or antiferromagnetically (singlet). For most cases, the ground state is a singlet, resulting in a negative exchange coupling constant J (see Table 4 and Figure 11), with an average of $J = -1.00$ kJ/mol. The distributions and statistics of the remaining properties are provided in Table 4 and Figure 11.

Table 4: Statistical values of J , α , and μ for the $[\mathbf{Cu}_2]$ molecules as obtained from B3LYP/TZVP. Std refers to the standard deviation, and Min / Max to smallest / largest values in the data set.

	J [kJ/mol]	α [Bohr ³]	μ [D]
Average	-1.00	141.0	2.60
Std	1.62	32.93	1.64
Median	-0.42	142.0	2.53
Min	-6.50	59.77	0.00
Max	0.03	220.44	7.56

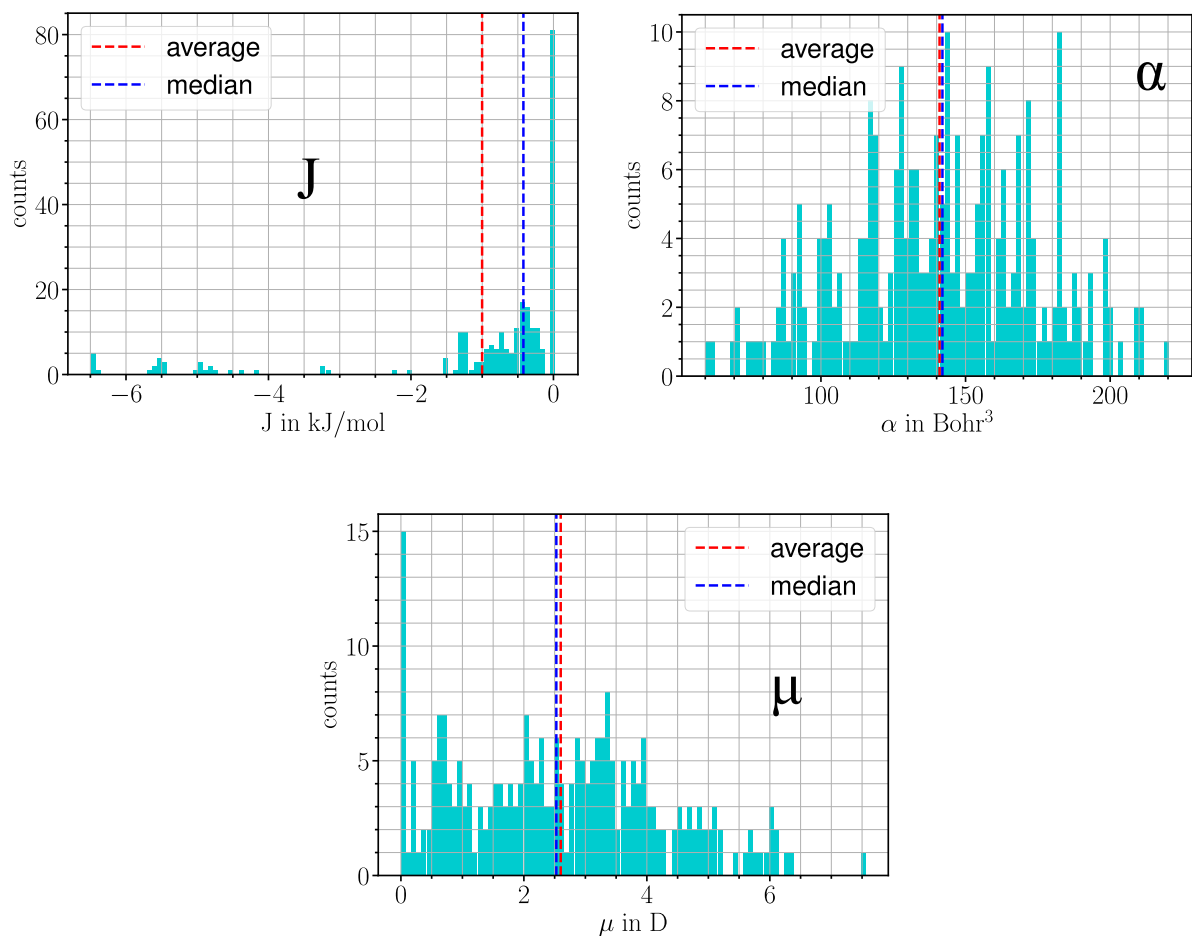


Figure 11: Distribution of J , μ and α for the $[\text{Cu}_2]$ molecules as obtained from B3LYP/TZVP in the high-spin state.

References

- (1) Berg, J. M.; Stryer, L.; Tymoczko, J. L.; Gatto, G. *Biochemistry*, 9th ed.; Freeman: New York, 2019.
- (2) Vogiatzis, K. D.; Polynski, M. V.; Kirkland, J. K.; Townsend, J.; Hashemi, A.; Liu, C.; Pidko, E. A. Computational Approach to Molecular Catalysis by 3d Transition Metals: Challenges and Opportunities. *Chem. Rev.* **2019**, *119*, 2453–2523.
- (3) Coronado, E. Molecular magnetism: from chemical design to spin control in molecules,

materials and devices. *Nature Rev. Mater.* **2020**, *5*, 87–104.

- (4) Gaita-Arino, A.; Luis, F.; Hill, S.; Coronado, E. Molecular spins for quantum computation. *Nat. Chem.* **2019**, *11*, 301–309.
- (5) Atzori, M.; Sessoli, R. The Second Quantum Revolution: Role and Challenges of Molecular Chemistry. *J. Am. Chem. Soc.* **2019**, *141*, 11339–11352.
- (6) Choi, K.-Y.; Wang, Z.; Nojiri, H.; van Tol, J.; Kumar, P.; Lemmens, P.; Bassil, B. S.; Kortz, U.; Dalal, N. S. Coherent Manipulation of Electron Spins in the Cu₃ Spin Triangle Complex Impregnated in Nanoporous Silicon. *Phys. Rev. Lett.* **2012**, *108*, 067206.
- (7) Affronte, M.; Troiani, F. In *Molecular Magnets: Physics and Applications*; Bartolomé, J., Luis, F., Fernández, J. F., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2014; pp 249–273.
- (8) Sanvito, S. Molecular spintronics. *Chem. Soc. Rev.* **2011**, 3336–3355.
- (9) von Kugelgen, S.; Freedman, D. E. A chemical path to quantum information. *Science* **2019**, *366*, 1070–1071.
- (10) Sierda, E.; Abadia, M.; Brede, J.; Elsebach, M.; Bugenhagen, B.; Prosenc, M. H.; Bazarnik, M.; Wiesendanger, R. On-Surface Oligomerization of Self-Terminating Molecular Chains for the Design of Spintronic Devices. *ACS Nano* **2017**, *11*, 9200–9206.
- (11) Thuijs, A. E.; Li, X.-G.; Wang, Y.-P.; Abboud, K. A.; Zhang, X.-G.; Cheng, H.-P.; Christou, G. Molecular analogue of the perovskite repeating unit and evidence for direct Mn^{III}-Ce^{IV}-Mn^{III} exchange coupling pathway. *Nature Commun.* **2017**, *8*, 500.
- (12) Anderson, P. W. Antiferromagnetism. Theory of Superexchange Interaction. *Phys. Rev.* **1950**, *72*, 350–356.
- (13) Anderson, P. W. New Approach to the Theory of Superexchange Interactions. *Phys. Rev.* **1959**, *115*, 2–13.

- (14) Kahn, O. *Molecular Magnetism*; Wiley VCH: New York, 1993.
- (15) Noodleman, L. Valence bond description of antiferromagnetic coupling in transition metal dimers. *J. Chem. Phys.* **1981**, *74*, 5737–5743.
- (16) Cramer, C. J.; Truhlar, D. G. Density functional theory for transition metals and transition metal chemistry. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10757–10816.
- (17) David, G.; Wennmohs, F.; Neese, F.; Ferré, N. Chemical Tuning of Magnetic Exchange Couplings Using Broken-Symmetry Density Functional Theory. *Inorg. Chem.* **2018**, *57*, 12769–12776.
- (18) Pantazis, D. A. Meeting the Challenge of Magnetic Coupling in a Triply-Bridged Chromium Dimer: Complementary Broken-Symmetry Density Functional Theory and Multireference Density Matrix Renormalization Group Perspectives. *J. Chem. Theory Comput.* **2019**, *15*, 938–948.
- (19) Jacob, C.; Reiher, M. Spin in Density-Functional Theory. *Int. J. Quantum Chem.* **2012**, *112*, 3661–3684.
- (20) Podewitz, M.; Herrmann, C.; Malassa, A.; Westerhausen, M.; Reiher, M. Spin–Spin interactions in polynuclear transition-metal complexes. *Chem. Phys. Lett.* **2008**, *451*, 301–308.
- (21) Neese, F. Definition of corresponding orbitals and the diradical character in broken symmetry DFT calculations on spin coupled systems. *J. Phys. Chem. Sol.* **2004**, *65*, 781–785.
- (22) Roemelt, M.; Pantazis, D. A. Multireference Approaches to Spin-State Energetics of Transition Metal Complexes Utilizing the Density Matrix Renormalization Group. *Adv. Theory Simul.* **2019**, *2*, 1800201.
- (23) Kawakami, T.; Miyagawa, K.; Sharma, S.; Saito, T.; Shoji, M.; Yamada, S.; Yamanaoka, S.; Okumura, M.; Nakajima, T.; Yamaguchi, K. UNO DMRG CAS CI calcula-

tions of binuclear manganese complex $\text{Mn(IV)}_2\text{O}_2(\text{NHCHCO}_2)_4$: Scope and applicability of Heisenberg model. *J. Comput. Chem.* **2018**, *40*, 333–341.

- (24) Krewald, V.; Pantazis, D. A. *Transition Metals in Coordination Environments*; Springer: Cham, 2019; pp 91–120.
- (25) Sharma, S.; Sivalingam, K.; Neese, F.; Chan, G. K.-L. Low-energy spectrum of iron-sulfur clusters directly from many-particle quantum mechanics. *Nature Chem.* **2014**, *6*, 927.
- (26) Manni, G. L.; Dobrutz, W.; Alavi, A. Compression of Spin-Adapted Multiconfigurational Wave Functions in Exchange-Coupled Polynuclear Spin Systems. *J. Chem. Theory Comput.* **2020**, *16*, 2202–2215.
- (27) Stein, C. J.; Reiher, M. Automated Selection of Active Orbital Spaces. *J. Chem. Theory Comput.* **2016**, *12*, 1760–1771.
- (28) Malmqvist, P. Å.; Pierloot, K.; Shahi, A. R. M.; ; Cramer, C. J.; Gagliardi, L. The restricted active space followed by second-order perturbation theory method: Theory and application to the study of CuO_2 and Cu_2O_2 . *J. Chem. Phys.* **2008**, *128*, 204109.
- (29) Cramer, C. J.; Kinal, A.; Włoch, M.; Piecuch, P.; Gagliardi, L. Theoretical Characterization of End-On and Side-On Peroxide Coordination in Ligated Cu_2O_2 Models. *J. Phys. Chem. A* **2006**, *110*, 11557–11568.
- (30) Sayfutyarova, E. R.; Sun, Q.; Chan, G. K.-L.; Knizia, G. Automated Construction of Molecular Active Spaces from Atomic Valence Orbitals. *J. Chem. Theory. Comput.* **2017**, *13*, 4063–4078.
- (31) Murphy, K. P. *Machine learning: A probabilistic perspective*; Adaptive computation and machine learning series; MIT Press: Cambridge, Mass., 2012.
- (32) Bishop, C. M. *Pattern Recognition and Machine Learning*; Springer: New York, 2006.

- (33) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.
- (34) 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.
- (35) Duan, C.; Janet, J. P.; Liu, F.; Nandy, A.; ; Kulik, H. J. Learning from Failure: Predicting Electronic Structure Calculation Outcomes with Machine Learning Models. *J. Chem. Theory Comput.* **2019**, *15*, 2331–2345.
- (36) Janet, J. P.; Liu, F.; Nandy, A.; Duan, C.; Yang, T.; Lin, S.; Kulik, H. J. Designing in the Face of Uncertainty: Exploiting Electronic Structure and Machine Learning Models for Discovery in Inorganic Chemistry. *Inorg. Chem.* **2019**, *58*, 10592–10606.
- (37) Aspuru-Guzik, A.; Lindh, R.; Reiher, M. The Matter Simulation (R)evolution. *ACS Cent. Sci.* **2018**, *4*, 144–152.
- (38) Dral, P. O. Quantum Chemistry in the Age of Machine Learning. *J. Phys. Chem. Lett.* **2020**, *11*, 2336–2347.
- (39) Draxl, C.; Scheffler, M. In *Handbook of Materials Modeling*; Andreoni, W., Yip, S., Eds.; Springer: Cham, 2020; pp 49–73.
- (40) Jablonka, K. M.; Ongari, D.; Moosavi, S. M.; Smit, B. Big-Data Science in Porous Materials: Materials Genomics and Machine Learning. **2020**, arXiv:2001.06728.
- (41) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547–555.

- (42) Fabrizio, A.; Meyer, B.; Fabregat, R.; Corminboeuf, C. Quantum Chemistry Meets Machine Learning. *Chimia* **2019**, *73*, 983–989.
- (43) Fabrizio, A.; Briling, K.; Grisafi, A.; Corminboeuf, C. Learning (from) the Electron Density: Transferability, Conformational and Chemical Diversity. *Chimia* **2020**, *74*, 232–236.
- (44) Hellström, M.; Behler, J. In *Handbook of Materials Modeling*; Andreoni, W., Yip, S., Eds.; Springer: Cham, 2020; pp 661–680.
- (45) von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Exploring chemical compound space with quantum-based machine learning. *Nature Rev. Chem.* **2020**,
- (46) Noé, F.; Tkatchenko, A.; Müller, K. R.; Clementi, C. Machine learning for molecular simulation. *Annu. Rev. Phys. Chem.* **2020**, *71*, 361–390.
- (47) Häse, F.; Galvan, I. F.; Aspuru-Guzik, A.; Lindh, R.; ; Vacher, M. How machine learning can assist the interpretation of ab initio molecular dynamics simulations and conceptual understanding of chemistry. *Chem. Sci.* **2019**, *10*, 2298–2307.
- (48) Musil, F.; Ceriotti, M. Machine Learning at the Atomic Scale. *Chimia* **2019**, *73*, 972–982.
- (49) Schleder, G. R.; Padilha, A. C. M.; Rocha, A. R.; Dalpian, G. M.; Fazzio, A. Ab Initio Simulations and Materials Chemistry in the Age of Big Data. *J. Chem. Inf. Model.* **2020**, *60*, 452–459.
- (50) Haghightlari, M.; Hachmann, J. Advances of machine learning in molecular modeling and simulation. *Curr. Opin. Chem. Eng.* **2019**, *23*, 51–57.
- (51) Chen, C.; Zuo, Y.; Ye, W.; Li, X.; Deng, Z.; Ong, S. P. A Critical Review of Machine Learning of Energy Materials. *Adv. Energy Mater.* **2020**, *10*, 1903242.

- (52) Barnard, A. S.; Motevalli, B.; Parker, A. J.; Fischer, J. M.; Feigl, C. A.; Opletal, G. Nanoinformatics, and the big challenges for the science of small things. *Nanoscale* **2019**, *11*, 19190–19201.
- (53) Albrecht, T.; Slabaugh, G.; Alonso, E.; Al-Arif, S. M. R. Deep learning for single-molecule science. *Nanotechnology* **2017**, *28*, 423001.
- (54) Janet, J. P.; Chan, L.; Kulik, H. J. Accelerating Chemical Discovery with Machine Learning: Simulated Evolution of Spin Crossover Complexes with an Artificial Neural Network. *J. Phys. Chem. Lett.* **2018**, *9*, 1064–1071.
- (55) 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.
- (56) Nandy, A.; Duan, C.; Janet, J. P.; Gugler, S.; Kulik, H. J. Strategies and Software for Machine Learning Accelerated Discovery in Transition Metal Chemistry. *Ind. Eng. Chem. Res* **2018**, *57*, 13973–13986.
- (57) Kulik, H. J. Making machine learning a useful tool in the accelerated discovery of transition metal complexes. *WIREs Comput Mol Sci.* **2020**, *10*.
- (58) Sanvito, S.; Zic, M.; Nelson, J.; Archer, T.; Oses, C.; Curtarolo, S. In *Handbook of Materials Modeling*; Andreoni, W., Yip, S., Eds.; Springer: Cham, 2020; pp 351–373.
- (59) Lunghi, A.; Sanvito, S. Surfing Multiple Conformation-Property Landscapes via Machine Learning: Designing Single-Ion Magnetic Anisotropy. *J. Phys. Chem. C* **2020**, *10*, 5802–5806.
- (60) Herrmann, C.; Elmisz, J. Electronic communication through molecular bridges. *Chem. Commun.* **2013**, *49*, 10456–10458.

- (61) Kirk, M. L.; Shultz, D. A.; Depperman, E. C.; Brannen, C. L. Donor-Acceptor Biradicals as Ground State Analogues of Photoinduced Charge Separated States. *J. Am. Chem. Soc.* **2007**, *129*, 1937–1943.
- (62) Kirk, M. L.; Shultz, D. A.; Stasiw, D. E.; Lewis, G. F.; Wang, G.; Brannen, C. L.; Sommer, R. D.; Boyle, P. D. Superexchange Contributions to Distance Dependence of Electron Transfer/Transport: Exchange and Electronic Coupling in Oligo(para-Phenylene)- and Oligo(2,5-Thiophene)-Bridged Donor–Bridge–Acceptor Biradical Complexes. *J. Am. Chem. Soc.* **2013**, *135*, 17144–17154.
- (63) Blondin, G.; Girerd, J. J. Interplay of electron exchange and electron transfer in metal polynuclear complexes in proteins or chemical models. *Chem. Rev.* **1990**, *90*, 1359–1376.
- (64) Tuzek, F.; Solomon, E. I. Charge-Transfer States and Antiferromagnetism of Bridged Cu Dimers: Application to Oxyhemocyanin. *J. Am. Chem. Soc.* **1994**, *116*, 6916–6924.
- (65) Shil, S.; Sarbadhikary, P.; Misra, A. Effect of length on the transport and magnetic properties of diradical substituted molecular wires. *RSC Adv.* **2016**, *6*, 99467–99474.
- (66) Sarbadhikary, P.; Shil, S.; Misra, A. Magnetic and transport properties of conjugated and cumulated molecules: the π -system enlightens part of the story. *Phys. Chem. Chem. Phys.* **2018**, *20*, 9364–9375.
- (67) Proppe, J.; Herrmann, C. Communication through molecular bridges: Different bridge orbital trends result in common property trends. *J. Comput. Chem.* **2015**, *36*, 201–209.
- (68) Herrmann, C. Electronic Communication as a Transferable Property of Molecular Bridges? *J. Phys. Chem. A* **2019**, *123*, 10205–10223.
- (69) Wang, C.-I.; Braza, M. K. E.; Claudio, G. C.; Nellas, R. B.; Hsu, C.-P. Machine Learning for Predicting Electron Transfer Coupling. *J. Phys. Chem. A* **2019**, *123*, 7792–7802.

- (70) Saeki, A.; Kranthiraja, K. A high throughput molecular screening for organic electronics via machine learning: present status and perspective. *Jap. J. Appl. Phys.* **2019**, *59*, SD0801.
- (71) Krämer, M.; Dohmen, P. M.; Xie, W.; Holub, D.; Christensen, A. S.; Elstner, M. Charge and Exciton Transfer Simulations Using Machine-Learned Hamiltonians. *J. Chem. Theory Comput.* **2020**,
- (72) Häse, F.; Kreisbeck, C.; Aspuru-Guzik, A. Machine learning for quantum dynamics: deep learning of excitation energy transfer properties. *Chem. Sci.* **2017**, *8*, 8419–8426.
- (73) Scalia, G.; Grambow, C. A.; Pernici, B.; Li, Y.-P.; Green, W. H. Evaluating Scalable Uncertainty Estimation Methods for Deep Learning-Based Molecular Property Prediction. *J. Chem. Inf. Model.* **2020**, *60*, 2697–2717.
- (74) Janet, J. P.; Duan, C.; Yang, T.; Nandy, A.; Kulik, H. J. A quantitative uncertainty metric controls error in neural network-driven chemical discovery. *Chem. Sci.* **2019**, *10*, 7913–7922.
- (75) Musil, F.; Willatt, M. J.; Langovoy, M. A.; Ceriotti, M. Fast and accurate uncertainty estimation in chemical machine learning. *J. Chem. Theor. Comput.* **2019**, *15*, 906–915.
- (76) Rasmussen, C. E.; Williams, C. K. I. *Gaussian processes for machine learning*, 3rd ed.; Adaptive computation and machine learning; MIT Press: Cambridge, Mass., 2008.
- (77) Panosetti, C.; Engelmann, A.; Nemeč, L.; Reuter, K.; Margraf, J. T. Learning to Use the Force: Fitting Repulsive Potentials in Density-Functional Tight-Binding with Gaussian Process Regression. *J. Chem. Theory Comput.* **2020**, *16*, 2181–2191.
- (78) Proppe, J.; Gugler, S.; Reiher, M. Gaussian Process-Based Refinement of Dispersion Corrections. *J. Chem. Theory Comput.* **2019**, *15*, 6046–6060.

- (79) Hashimoto, W.; Tsuji, Y.; Yoshizawa, K. Optimization of Work Function via Bayesian Machine Learning Combined with First-Principles Calculation. *J. Phys. Chem. C* **2020**,
- (80) Raimbault, N.; Grisafi, A.; Ceriotti, M.; Rossi, M. Using Gaussian process regression to simulate the vibrational Raman spectra of molecular crystals. *New J. Phys.* **2019**, *21*, 105001.
- (81) Denzel, A.; Haasdonk, B.; Kästner, J. Gaussian Process Regression for Minimum Energy Path Optimization and Transition State Search. *J. Phys. Chem. A* **2019**, *123*, 9600–9611.
- (82) Meyer, R.; Hauser, A. W. Geometry optimization using Gaussian process regression in internal coordinate systems. *J. Chem. Phys.* **2020**, *152*, 084112.
- (83) Denzel, A.; Kästner, J. Gaussian process regression for geometry optimization. *J. Chem. Phys.* **2018**, *148*, 094114.
- (84) Schmitz, G.; Christiansen, O. Gaussian process regression to accelerate geometry optimizations relying on numerical differentiation. *J. Chem. Phys.* **2018**, *148*, 241704.
- (85) Raggi, G.; Galván, I. F.; Ritterhoff, C. L.; Vacher, M.; Lindh, R. Restricted-Variance Molecular Geometry Optimization Based on Gradient-Enhanced Kriging. *J. Chem. Theory Comput.* **2020**, *16*, 3989–4001.
- (86) Dai, J.; Krems, R. V. Interpolation and Extrapolation of Global Potential Energy Surfaces for Polyatomic Systems by Gaussian Processes with Composite Kernels. *J. Chem. Theory Comput.* **2020**, *16*, 1386–1395.
- (87) Song, Q.; Zhang, Q.; Meng, Q. Revisiting the Gaussian process regression for fitting high-dimensional potential energy surface and its application to the $\text{OH} + \text{HO}_2 \rightarrow \text{O}_2 + \text{H}_2\text{O}$ reaction. *J. Chem. Phys.* **2020**, *152*, 134309.
- (88) Simm, G. N.; Reiher, M. Error-Controlled Exploration of Chemical Reaction Networks with Gaussian Processes. *J. Chem. Theory Comput.* **2018**, *14*, 5238–5248.

- (89) Ulissi, Z. W.; Medford, A. J.; Bligaard, T.; Nørskov, J. K. To address surface reaction network complexity using scaling relations machine learning and DFT calculations. *Nature Commun.* **2017**, *8*, 14621.
- (90) Sutton, C.; Boley, M.; Ghiringhelli, L. M.; Rupp, M.; Vreeken, J.; Scheffler, M. Identifying Domains of Applicability of Machine Learning Models for Materials Science. **2019**, ChemRxiv, DOI 10.26434/chemrxiv.9778670.v2.
- (91) Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M. Big Data of Materials Science: Critical Role of the Descriptor. *Phys. Rev. Lett.* **2015**, *114*, 105503.
- (92) Jung, H.; Stocker, S.; Kunkel, C.; Oberhofer, H.; Han, B.; Reuter, K.; Margraf, J. T. Size-Extensive Molecular Machine Learning with Global Descriptors. **2019**, ChemRxiv, DOI 10.26434/chemrxiv.10002020.v1.
- (93) Pozdnyakov, S. N.; Willatt, M. J.; Bartok, A. P.; Ortner, C.; Csanyi, G.; Ceriotti, M. On the Completeness of Atomic Structure Representations. **2020**, arXiv:2001.11696.
- (94) Collins, C. R.; Gordon, G. J.; von Lilienfeld, O. A.; Yaron, D. J. Constant size descriptors for accurate machine learning models of molecular properties. *J. Chem. Phys.* **2018**, *148*, 241718.
- (95) 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.
- (96) Jäger, M. O. J.; Morooka, E. V.; Canova, F. F.; Himanen, L.; Foster, A. S. Machine learning hydrogen adsorption on nanoclusters through structural descriptors. *NPJ Comput. Mater.* **2018**, *4*, 37.
- (97) Langer, M. F.; Goëßmann, A.; Rupp, M. Representations of Molecules and Materials for

Interpolation of Quantum-Mechanical Simulations via Machine Learning. 2020; arXiv 2003.12081.

- (98) Duvenaud, D.; Lloyd, J. R.; Grosse, R.; Tenenbaum, J. B.; Ghahramani, Z. Structure Discovery in Nonparametric Regression through Compositional Kernel Search. 2013; arXiv:1302.4922.
- (99) Chugh, T.; Rahat, A.; Palar, P. S. *Machine Learning, Optimization, and Data Science*; 2019; pp 579–591.
- (100) Hay, P. J.; Thibeault, J. C.; Hoffmann, R. Orbital Interactions in Metal Dimer Complexes. *J. Am. Chem. Soc.* **1975**, *97*, 4884–4899.
- (101) Gugler, S.; Janet, J. P.; Kulik, H. J. Enumeration of de novo inorganic complexes for chemical discovery and machine learning. *Mol. Syst. Des. Eng.* **2020**, *5*, 139–152.
- (102) Solomon, E. I.; Sundaram, U. M.; Machonkin, T. E. Multicopper Oxidases and Oxygenases. *Chem. Rev.* **1996**, *96*, 2563–2606.
- (103) Roth, A.; Becher, J.; Herrmann, C.; Görls, H.; Vaughan, G.; Reiher, M.; Klemm, D.; Plass, W. Large Trinuclear Copper(II) Complexes Derived from Schiff-Bas e Ligands Based on a 6-Amino-6-deoxy-glucopyranoside: Structural and Magnetic Characterization. *Inorg. Chem.* **2006**, *45*, 10066–10076.
- (104) Ülkü K k am-Demir.; Goldman, A.; Esrafilı, L.; Gharib, M.; Morsali, A.; Weingart, O.; Janiak, C. Coordinatively unsaturated metal sites (open metal sites) in metal–organic frameworks: design and applications. *Chem. Soc. Rev.* **2020**, Advance Article.
- (105) Khan, A. H.; Peikert, K.; Hoffmann, F.; Fr ba, M.; Bertmer, M. Nitric Oxide Adsorption in Cu₃btc₂-Type MOFs—Physisorption and Chemisorption as NONOates. *J. Phys. Chem. C* **2019**, *123*, 4299–4307.

- (106) Kaskel, S., Ed. *The Chemistry of Metal-Organic Frameworks: Synthesis, Characterization, and Applications*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, 2016.
- (107) Roemelt, M.; Guo, S.; Chan, G. K.-L. A projected approximation to strongly contracted N-electron valence perturbation theory for DMRG wavefunctions. *J. Chem. Phys.* **2016**, *144*, 204113.
- (108) Liechtenstein, A. I.; Katsnelson, M. I.; Antropov, V. P.; Gubanov, V. A. Local Spin Density Functional Approach to the Theory of Exchange Interactions in Ferromagnetic Metals and Alloys. *J. Magn. Magn. Mater.* **1987**, *67*, 65–74.
- (109) Han, M.; Ozaki, T.; Yu, J. Electronic structure, magnetic interactions, and the role of ligands in Mn_n ($n=4,12$) single-molecule magnets. *Phys. Rev. B* **2004**, *70*, 184421.
- (110) Steenbock, T.; Tasche, J.; Lichtenstein, A.; Herrmann, C. A Green’s-function approach to exchange spin coupling as a new tool for quantum chemistry. *J. Chem. Theory Comput.* **2015**, *11*, 5651–5664.
- (111) Duvenaud, D. K. Automatic model construction with Gaussian processes. Ph.D. thesis, University of Cambridge, Cambridge, UK, 2014.
- (112) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; WILEY-VCH Verlag GmbH: Weinheim, 2000.
- (113) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B* **2013**, *87*.
- (114) Pronobis, W.; Tkatchenko, A.; Müller, K.-R. Many-Body Descriptors for Predicting Molecular Properties with Machine Learning: Analysis of Pairwise and Three-Body Interactions in Molecules. *J. Chem. Theory Comput.* **2018**, *14*, 2991–3003.
- (115) Huo, H.; Rupp, M. Unified Representation of Molecules and Crystals for Machine Learning. 2018; arXiv:1704.06439.

- (116) Himanen, L.; Jäger, M. O. J.; Morooka, E. V.; Filippo, F.; Canova, F.; Ranawat, Y. S.; Z.Gao, D.; Rinke, P.; Foster, A. S. Dscribe: Library of descriptors for machine learning in materials science. *Comput. Phys. Commun.* **2020**, *247*, 106949.
- (117) Han, W.-G.; Lovell, T.; Liu, T.; Noodleman, L. Density Functional Study of a μ -1,1-Carboxylate Bridged Fe(III)-O-Fe(IV) Model Complex. 2. Comparison with Ribonucleotide Reductase Intermediate X. *Inorg. Chem.* **2004**, *43*, 613–621.
- (118) Peralta, J. E.; Barone, V. Magnetic exchange couplings from noncollinear spin density functional perturbation theory. *J. Chem. Phys.* **2008**, *129*, 194107.
- (119) Philips, J. J.; Peralta, J. E. Towards the blackbox computation of magnetic exchange coupling parameters in polynuclear transition-metal complexes: Theory, implementation, and application. *J. Chem. Phys.* **2013**, *138*, 174115.
- (120) Joshi, R. P.; Phillips, J. J.; Peralta, J. E. Magnetic Exchange Couplings in Heterodinuclear Complexes Based on Differential Local Spin Rotations. *J. Chem. Theory Comput.* **2016**, *12*, 1728–1734.
- (121) Pernot, P.; Huang, B.; Savin, A. Impact of non-normal error distributions on the benchmarking and ranking of Quantum Machine Learning models. 2020; arXiv:2004.02524.
- (122) Brunken, C.; Reiher, M. Self-Parametrizing System-Focused Atomistic Models. *J. Chem. Theory Comput.* **2020**, *16*, 1646–1665.
- (123) Grimme, S. A General Quantum Mechanically Derived Force Field (QMDF) for Molecules and Condensed Phase Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 4497–4514.
- (124) Grimme, S.; Bannwarth, C.; Caldeweyher, E.; Pisarek, J.; Hansen, A. A general intermolecular force field based on tight-binding quantum chemical calculations. *J. Chem. Phys.* **2017**, *147*, 161708.

- (125) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.
- (126) Bursch, M.; Neugebauer, H.; Grimme, S. Structure Optimisation of Large Transition-Metal Complexes with Extended Tight-Binding Methods. *Angew. Chem. Int. Ed.* **2019**, *58*, 11078–11087.
- (127) Taylor, M. G.; Yang, T.; Lin, S.; Nandy, A.; Janet, J. P.; Duan, C.; Kulik, H. J. Seeing Is Believing: Experimental Spin States from Machine Learning Model Structure Predictions. *J. Phys. Chem. A* **2020**, *124*, 3286–3299.
- (128) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (129) Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33*, 8822–8824.
- (130) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- (131) TURBOMOLE V7.1, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from <http://www.turbomole.com>.
- (132) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kömel, C. Electronic Structure Calculation on Workstation Computers: The Program System Turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165–169.
- (133) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange,. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

- (134) 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* **1988**, *37*, 785–789.
- (135) Schäfer, A.; Huber, C.; Ahlrichs, R. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J. Chem. Phys* **1994**, *100*, 5829–5835.
- (136) Frisch, M. J. et al. Gaussian 09 Revision A.1. Gaussian Inc. Wallingford CT 2009.
- (137) Deffner, M.; Groß, L.; Steenbock, T.; Voigt, B. A.; Zöllner, M. S.; Solomon, G. C.; Herrmann, C. 2008-2020; ARTAIOS — a postprocessing framework for electronic structure calculations, <https://www.chemie.uni-hamburg.de/institute/ac/arbeitsgruppen/herrmann/software/artaios.html>.
- (138) Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
- (139) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data* **2014**, *1*, 140022.
- (140) Estes, E. D.; Hatfield, W. E.; Hodgson, D. J. Structural characterization of di- μ -hydroxo-bis (N, N, N', N'-tetraethylethylenediamine) dicopper (II) perchlorate, [Cu (teen) OH] 2 (ClO₄) 2. *Inorg. Chem.* **1974**, *13*, 1654–1657.
- (141) Zheng, X.; Kawae, T.; Kashitani, Y.; Li, C.; Tateiwa, N.; Takeda, K.; Yamada, H.; Xu, C.; Ren, Y. Unconventional magnetic transitions in the mineral clinoatacamite Cu₂Cl(OH)₃. *Phys. Rev. B* **2005**, *71*, 052409.
- (142) Zheng, X.; Kubozono, H.; Nishiyama, K.; Higemoto, W.; Kawae, T.; Koda, A.; Xu, C. Coexistence of long-range order and spin fluctuation in geometrically frustrated clinoatacamite Cu₂Cl(OH)₃. *Phys. Rev. Lett.* **2005**, *95*, 057201.
- (143) Banks, M.; Kremer, R.; Hoch, C.; Simon, A.; Ouladdiaf, B.; Broto, J.-M.; Rakoto, H.;

- Lee, C.; Whangbo, M.-H. Magnetic ordering in the frustrated Heisenberg chain system cupric chloride CuCl₂. *Phys. Rev. B* **2009**, *80*, 024404.
- (144) Saraireh, S. A.; Altarawneh, M. Electronic structure of the CuCl₂ (100) surface: A DFT first-principle study. *J. Nanomater.* **2012**, *2012*.
- (145) Maaten, L. v. d.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res* **2008**, *9*, 2579–2605.
- (146) Li, Y.-P.; Han, K.; Grambow, C. A.; Green, W. H. Self-Evolving Machine: A Continuously Improving Model for Molecular Thermochemistry. *J. Phys. Chem. A* **2019**, *123*, 2142–2152.
- (147) Janet, J. P.; Kulik, H. J. Predicting electronic structure properties of transition metal complexes with neural networks. *Chem. Sci.* **2017**, *8*, 5137–5152.
- (148) Kudo, M.; Sklansky, J. Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition* **2000**, *33*, 25 – 41.
- (149) Hua, J.; Xiong, Z.; Lowey, J.; Suh, E.; Dougherty, E. R. Optimal number of features as a function of sample size for various classification rules. *Bioinformatics* **2005**, *21*, 1509–1515.
- (150) 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.
- (151) Ouyang, R.; Curtarolo, S.; Ahmetcik, E.; Scheffler, M.; Ghiringhelli, L. M. SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Phys. Rev. Mater.* **2018**, *2*.
- (152) Le, T. T.; Fu, W.; Moore, J. H. Scaling tree-based automated machine learning to biomedical big data with a feature set selector. *Bioinformatics* **2020**, *36*, 250–256.

Graphical TOC Entry

