THE JOURNAL OF PHYSICAL CHEMISTRY C

Article

pubs.acs.org/JPCC

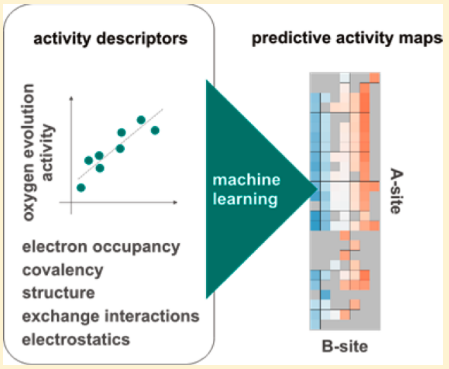
Descriptors of Oxygen-Evolution Activity for Oxides: A Statistical Evaluation

Wesley T. Hong,*[,†] Roy E. Welsch,[‡] and Yang Shao-Horn*[,†,§]

[†]Department of Materials Science & Engineering, [‡]Sloan School of Management, [§]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

Supporting Information

ABSTRACT: Catalysts for oxygen electrochemical processes are critical for the commercial viability of renewable energy storage and conversion devices such as fuel cells, artificial photosynthesis, and metal-air batteries. Transition metal oxides are an excellent system for developing scalable, non-noble-metal-based catalysts, especially for the oxygen evolution reaction (OER). Central to the rational design of novel catalysts is the development of quantitative structure−activity relationships, which correlate the desired catalytic behavior to structural and/or elemental descriptors of materials. The ultimate goal is to use these relationships to guide materials design. In this study, 101 intrinsic OER activities of 51 perovskites were compiled from five studies in literature and additional measurements made for this work. We explored the behavior and performance of 14 descriptors of the metal−oxygen bond strength using a number of statistical approaches, including factor analysis and linear regression models. We found that these descriptors can be classified into five descriptor families and identify electron occupancy and metal−oxygen covalency as the dominant influences on the OER activity. However, multiple descriptors still need to be considered in order to develop strong predictive relationships, largely outperforming the use of only one or two descriptors (as conventionally done in the field). We confirmed that the number of d electrons, charge-transfer energy (covalency), and optimality of e_g occupancy play the important roles, but found that structural factors such as M−O−M bond angle and tolerance factor are relevant as well. With these tools, we demonstrate how statistical learning can be used to draw novel physical insights and combined with data mining to rapidly screen OER electrocatalysts across a wide chemical space.



INTRODUCTION

A central theme in modern materials science is to identify materials with properties tailored to specific applications. However, materials selection is severely hindered where accurate physical models of functional properties are unavailable. Discovering catalyst chemistries with unprecedentedly high activities for the oxygen evolution reaction (OER) is a prototypical example of materials design that has struggled with an efficient path to progress for this reason.[1] To date, two main approaches have been used to determine material candidates from both experimental and computational catalysis work. On one end, quantitative structure−activity relationships (QSARs) have related the OER activity of catalysts to simple chemical frameworks[2,3] and/or adsorption energetics from density functional theory calculations[4−6] (i.e., descriptors). This approach provides physical insights but explores a select few compositions out of the large potential chemical space, limiting reliable testing of predictions.[7] Consequently, such studies have primarily succeeded at rationalizing observed activity trends rather than predicting them. On the other end, high-throughput combinatorial experiments have enabled a much wider investigation of material compositions.[8−12] Combining these experiments with objective benchmarking[13,14] is one approach to screen OER catalysts for a broad range of chemistries with self-consistent measurements.[15] However, detailed characterization and understanding of the physical origins responsible for OER activity are often lost at the expense of broadening the scope.

One promising way to address these limitations is to synthesize and learn from the wealth of rigorously collected data from different experimental studies. The statistical learning paradigm is a growing research direction in materials science[16−18] that embraces this approach. Specifically, supervised learning methods have gained traction as a large-scale approach for mapping constituent properties to system properties using historic data.[16] While the technique has been applied for compositional design in topics such as bandgaps in optical and photovoltaic materials,[19] lattice parameters in perovskites,[20,21] heterogeneous catalyst performance,[22,23] and zeolite synthesis,[24] its use has been notably absent in the field of oxygen electrocatalysis.

In the case of OER catalysts, QSARs have largely focused on descriptors that may control the surface metal (M) − oxygen (O) bonding, which dictates the OER activity.[1,4] Dating as far back as

Received: October 14, 2015
Revised: November 30, 2015
Published: December 16, 2015

DOI: 10.1021/acs.jpcc.5b10071
J. Phys. Chem. C 2016, 120, 78−86

1977, highly active oxides for the OER have been linked to a large number of parameters, including the transition-metal redox couple,[25] electrical conductivity,[25] transition-metal d-electron count,[26] transition-metal e_g occupancy,[2] metal−oxygen covalency[27] (a.k.a charge-transfer energy[27,28]), σ^*-band filling,[3,29] oxidation state,[5] and O p-band center.[6,30] The field thus stands to benefit greatly from understanding the relationships among descriptors and identifying the most influential one(s).

Here we demonstrate how statistical learning can be used to address ambiguities in the reported descriptors for OER activity, as well as complement modern materials informatics tools such as the Materials Project[31] to enable large-scale oxide screening. In this study, we focus on oxides in the perovskite family (ABO_3, where A is a rare-earth or alkaline-earth metal and B is a transition metal) to demonstrate the statistical method. Compiling 101 OER activity measurements on 51 unique oxide chemistries from previous works[2,6,26,29,32] and our own experimental measurements, we assess the predictive power of structure−activity descriptors proposed to date and find that they generally have poor predictive accuracies when used alone. Factor analysis demonstrates that there are at least five types of descriptors that can influence the OER activity, with transition metal electron occupancy and metal−oxygen covalency having the strongest effect. Moreover, QSARS with up to nine descriptors produce the most accurate and precise predictions, demonstrating the need to consider multiple descriptors. With these tools, we demonstrate how statistical learning can be used to draw novel physical insights and combined with data mining to rapidly screen OER catalysts across a wide chemical space.

EXPERIMENTAL METHODS

Data Preparation. In this study, we compiled 101 OER activity measurements on 51 unique perovskite chemistries from literature[2,6,26,29,32] and experimental measurements made for this work. The OER activity measurements made in this work used the same method published previously.[2,6] For consistent comparisons between studies, the OER activity was referenced to the reversible hydrogen electrode (RHE) scale, and all measurements examined were performed in the pH 13−14 range. We assume that the choice of electrolyte in these studies (1.0 M NaOH, and 0.1 and 1.0 M KOH) does not substantially modify the relative OER activities of catalysts, i.e., the mechanism does not change within the pH range or due to the choice of electrolyte cation. Although OER activity is typically measured by its log-current vs voltage response, there is no single metric for intrinsic OER activity (e.g., mA/cm^2_{oxide} under a given overpotential). The OER activities aggregated for this work use several different metrics, summarized in Table 1. We note that the purpose of this study is not to identify detailed mechanistic insights, which may depend heavily on the pH studied.[33] Rather, we focus on understanding the key material properties needed for high OER activity. We also draw attention to some of the selection biases that occur in the data. Notably, the data is aggregated from peer-reviewed publications, which may bias the data set toward higher activity materials. In addition, 39 of 51 perovskites used in this study also contain varying quantities of lanthanum. Nevertheless, the data set includes a range of alkaline-earth and rare-earth cation substitutions, allowing insights to be drawn regarding the influence of the A-site cation on the OER activity. The full data set, including oxide composition, electrolyte composition, OER activity metric, and OER activity, is provided as a CSV file in the Supporting Information.

Table 1. Summary of OER Activity Data Analyzed in This Study[a]

ref	OER activity metric	electrolyte	no. oxides measured
Matsumoto et al.[29]	η @ 40 mA/cm^2_{geo}	1 M KOH	16
Matsumoto et al.[29]	η @ 60 mA/cm^2_{geo}	1 M KOH	16
Bockris, Otagawa[26]	log i_{oxide} @ 1.53 V vs RHE	1 M NaOH	15
Jain et al.[32]	η @ 10 mA/cm^2_{geo}	1 M KOH	4
Jain et al.[32]	η @ 100 mA/cm^2_{geo}	1 M KOH	4
Suntivich et al.[2]	η @ 50 $\mu A/cm^2_{oxide}$	0.1 M KOH	12
Suntivich et al.[2]	log i_{oxide} @ 1.60 V vs RHE	0.1 M KOH	12
Grimaud et al.[6]	η @ 500 $\mu A/cm^2_{oxide}$	0.1 M KOH	6
Grimaud et al.[6]	log i_{oxide} @ 1.60 V vs RHE	0.1 M KOH	6
this work	log i_{oxide} @ 1.60 V vs RHE	0.1 M KOH	10

[a]Subscripts for current densities define whether oxide or geometric surface area was used for normalization.

Here, we used a simple transformation to compare the OER activities reported using different metrics. As each study measured $LaCoO_3$, all OER activities were taken as relative changes, with respect to $LaCoO_3$:

$$y'_{ij} = \frac{y_{ij} - y_{\mathrm{LaCoO3},j}}{|y_{\mathrm{LaCoO3},j}|}$$

where y_{ij} represents the OER activity of an oxide i in study j. After this correction, the data contained two different metrics instead of eight: relative overpotential and relative log-current-density. The difference in the units results in different spreads (Figure S1, Supporting Information). To compare relative overpotential to relative log-current-density, the data were stratified into two groups based on their units. Each group was normalized by its respective standard deviation (σ_k) to obtain the same spread for the metrics (Figure S2, Supporting Information):

$$(\text{relative OER activity})_i = \frac{y'_{ik}}{\sigma_k}$$

where k denotes whether oxide i has units of relative overpotential or relative log-current-density. All subsequent analysis was performed on this "relative OER activity" (referenced to $LaCoO_3$), which has units of standard deviations (s.d.).

Figure 1 illustrates the relative OER activities for the oxides in this study, ordered by decreasing median activity. The colors in Figure 1 categorize the transition metal atom(s) of each oxide, which is typically considered to hold the primary influence on the OER. The ranges of the boxplots in Figure 1 result from the spread in the OER activity across different studies, which are reasonably small for all oxides aside from $LaCrO_3$. It is evident from Figure 1 that oxides with Co, Co/Fe, or Ni have higher OER activities than those with V, Cr, Mn, or Fe. The standard deviation of the relative OER activity across the studies investigated here is ∼0.5 s.d. for the most reported material, $La_{0.6}Sr_{0.4}CoO_3$ (Table S2, Supporting Information). This error is equivalent to 30 mV overpotential at 50 $\mu A/cm^2_{oxide}$ and is similar to the experimental error within Suntivich et al.[2] Consequently, this error from comparing different OER activity

DOI: 10.1021/acs.jpcc.5b10071
J. Phys. Chem. C 2016, 120, 78−86

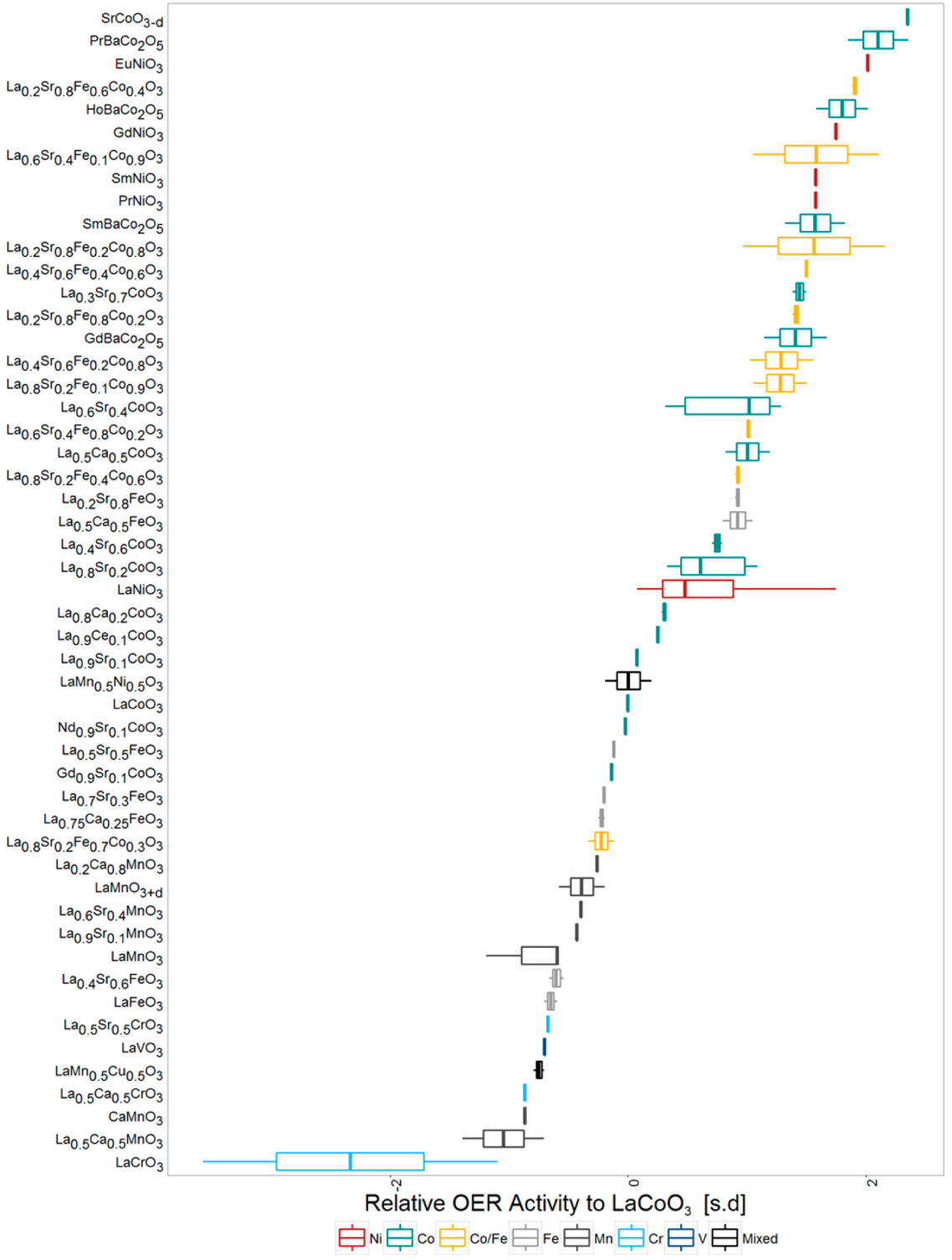


Figure 1. Boxplot of the relative OER activities with respect to $LaCoO_3$ (units of standard deviations) for 101 measurements on 51 unique perovskite chemistries from literature[2,6,26,29,32] and measurements in this study. The data are sorted by median relative OER activity. Colors denote the transition-metal site chemistry. Lines illustrate the range of reported relative OER activities, the ends of the box show the first and third quartiles, and the band inside the box is the median.

metrics appears to be comparable to experimental errors after applying the standardization process described above.

While the transition metal alone is qualitatively useful for organizing the OER activity of perovksites, it does not provide strong predictive accuracy. For example, the data show that cobalt-based oxides can range over 2 s.d. in the relative OER activity. Below we discuss that QSARs can be used to provide physical insights into OER activity with high predictive power.

Regression Methods. We considered 14 descriptors for developing QSARs, and details can be found in Table 2. Although catalysis of the OER is dictated by the catalyst surface chemistry, which may differ in stoichiometry or crystallinity from the bulk,

DOI: 10.1021/acs.jpcc.5b10071
J. Phys. Chem. C 2016, 120, 78−86

Table 2. List of Descriptor Variables and Their Physical Descriptions[a]

descriptors	details
d electrons[26]	The nominal number of transition-metal (B-site) d electrons based on its formal oxidation state. For mixed transition metals, the average number of d electrons was taken.
e_g electrons[2]	The nominal number of transition-metal (B-site) e_g electrons based on its formal oxidation state and spin state. For mixed transition metals, the more active transition metal was taken, per the e_g rules described in literature.[2,37]
optimality of e_g[2]	The absolute deviation of the nominal number of transition-metal (B-site) e_g electrons from the experimentally optimal value (1.2), as proposed in literature.[2]
oxidation state[5]	The formal oxidation state of the transition metal (B-site). For mixed transition metals, the average oxidation state was taken.
optimality of tolerance factor	The Goldschmidt tolerance factor (t) of the perovskite, calculated using the SPuDS software. Describes the ability of the crystal structure to accommodate the size of the ions (r_A, r_B, r_O for the A-site, B-site, and oxygen ions, respectively):[38] $t = \frac{r_A + r_O}{\sqrt{2}(r_B + r_O)}$ Optimality is defined as the absolute deviation of the tolerance factor from its ideal value, 1.0. Average site ionic radii were used for mixed metal oxides.
M−O−M bond angle (ave)	The average transition metal−oxygen-transition metal bond angle.
M−O bond length (ave)[39,40]	The average transition metal−oxygen bond length.
M−M distance (ave)	The average transition metal-transition metal distance.
Madelung potential, M	The on-site Madelung potential of the transition metal site (eV). Calculated using VESTA.[41]
Madelung potential, O	The on-site Madelung potential of the oxygen site (eV). Calculated using VESTA.[41]
ionization energy	The first ionization energy of the transition metal in its formal oxidation state.[36] Average ionization energy was used for noninteger oxidation states.
Hubbard U	Estimate of the Hubbard U interaction between transition-metal d electrons, using values defined above.[36]
charge-transfer energy[27]	Estimate of the charge-transfer energy between the transition metal and oxygen atoms, using values defined above.[36] The charge-transfer energy is defined by the difference in transition-metal and oxygen electronegativities in the oxide.
magnetic moment[42]	The magnetic moment of the transition metal, defined using the total spin.

[a]Values for the descriptors are provided as a.csv file in the Supporting Information. References to studies relating the descriptor to OER activity are noted where relevant.

Table 3. List of Regression Methods and Details of the Models Studied, Including Number of Descriptors and Optimization Criteria[a]

regression method	QSARs implemented
simple ordinary least-squares regression[43]	two models using a single descriptor each; these were selected based on previous descriptor studies (d electrons[4,5,26] and optimal e_g occupancy[2])
multiple ordinary least-squares regression[43]	a single model training all 14 descriptors; two models training descriptor pairs reported previously (d electrons and oxidation state,[5] and optimal e_g occupancy and charge-transfer energy[2,27])
forward selection, backward elimination[43]	forward selection and backward elimination algorithms for selecting descriptors by minimizing the Akaike information criterion[7]
penalized methods	ridge,[44] LASSO,[44] three elastic nets with mixing parameters α = 0.25, 0.50, 0.75 (designated EN1, EN2, and EN3, respectively),[44] and least angle regression (LAR)[45]
latent variable regression	factor analysis[43] optimized using the Kaiser criterion[46] (Figure S6, Supporting Information) to aggregate descriptors via dimension reduction, followed by least-squares regression

[a]Additional background on the regression methods are discussed in the Supporting Information.

bulk electronic structure information and physiochemical properties that are more easily defined have been shown to strongly influence surface reactivity trends of oxides.[1,5,6] This observed correlation has been rationalized by the connections between the surface chemistry of these oxides in a given electrolyte, bulk thermochemistry, and bulk electronic structure,[34] but remains the subject of further study. The 14 descriptors in this study included bulk physical properties related to the metal−oxygen bond strength of oxides, such as chemical formalisms (e.g., transition-metal oxidation state), simple models (e.g., the Goldschmidt tolerance factor,[35] charge-transfer energy,[36] or Hubbard U[36]), and experimental structural data (e.g., average metal−oxygen bond length). Although the charge-transfer energy and Hubbard U values used in this study capture general physical trends, they are not quantitatively accurate[28] because they require simple models to estimate them a priori.[28,36] We thus make notes when results are semiquantitative and should not be interpreted physically given these approximations.

To compare the prediction errors of different QSARs, the 101 measurements were randomly divided into training (81 points) and test (20 points) sets. The training set was used to fit ("train") the coefficients for the QSARs. 5-fold cross-validation (CV) was performed for each QSAR to determine optimal fitting parameters and select the most accurate model. The test data were then used to estimate how well the QSAR performs on data not used for fitting. All errors reported are mean absolute errors (MAE). Tables and figures of important statistics from the analyses are provided in the Supporting Information.

Five different linear regression methods were implemented to generate QSARs, as descriptor studies typically assume that linear free-energy relationships govern OER activity.[1] The regression methods studied in this work are summarized in Table 3. All descriptor values were standardized prior to training

DOI: 10.1021/acs.jpcc.5b10071
J. Phys. Chem. C 2016, 120, 78−86

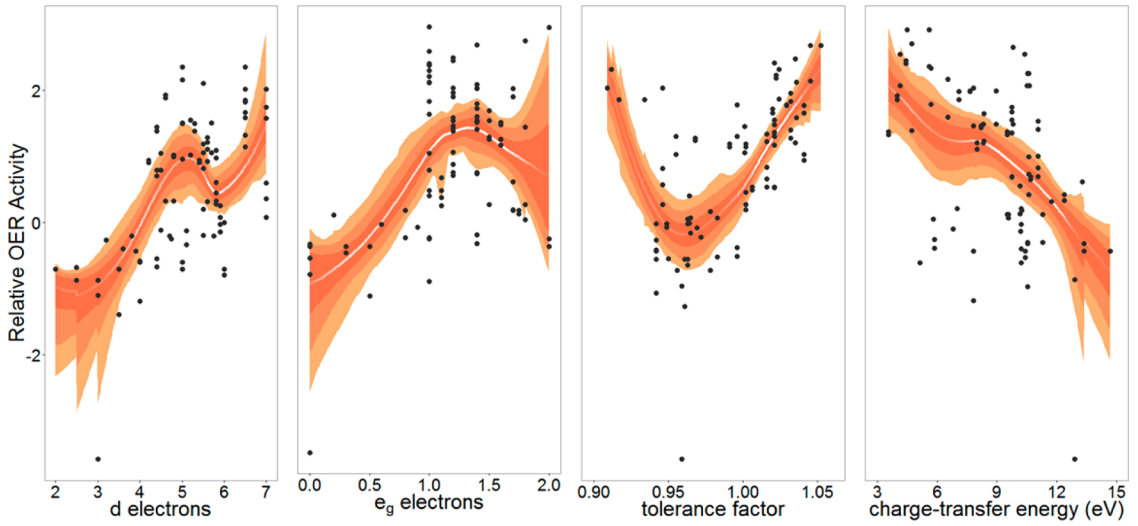


Figure 2. Confidence band plots of trends in relative OER activity for four descriptors: number of *d* electrons, number of e_g electrons, tolerance factor, and charge-transfer energy (Table 2). The full set of data (training and test) is plotted. A LOESS smoothing curve of the data is shown in white to visualize the trend, with 5000 nonparametrically bootstrapped smoothers used to determine the 68%, 95%, and 99.7% confidence bands (lighter shade indicates larger confidence interval).

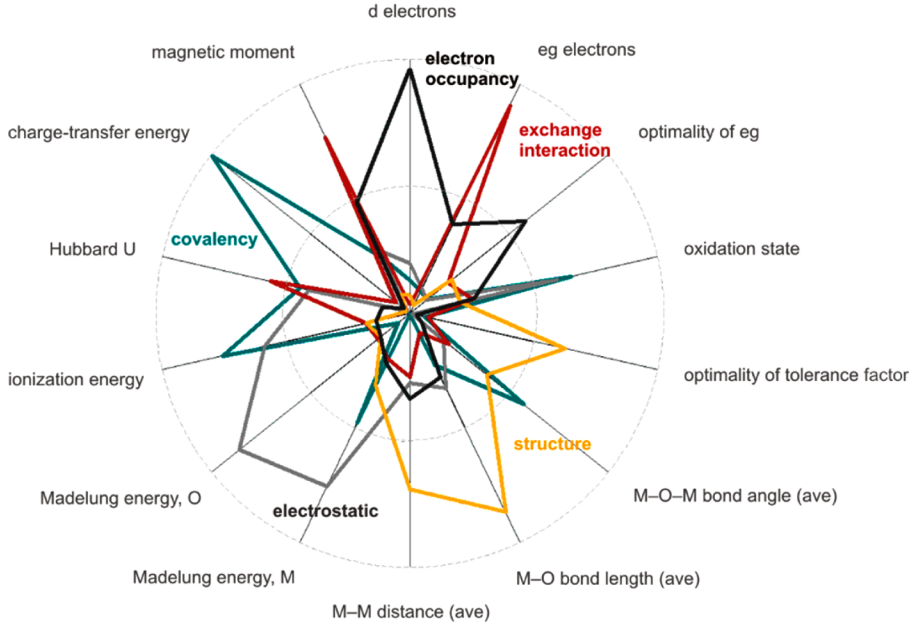


Figure 3. Loading magnitudes for the 14 descriptors obtained by factor analysis, using the Kaiser criterion to determine the optimal number of factors (5). Larger radial component indicates larger contribution of a descriptor to the factor. The factors can be considered as descriptor families, related to covalency (green), electrostatics (gray), structure (yellow), exchange interaction (red), and electron occupancy (dark gray).

the models to correct for differences in their scale. As a result, the coefficients from the regressions (β) correspond to the relative influence of descriptors on the relative OER activity. Test data relative OER activities and descriptor values were standardized using the means and standard deviations of the training data to simulate actual predictions. Additional details on descriptor standardization are discussed in the Supporting Information.

Predictive Mapping using the Materials Project Database. To map predictions of perovskite catalysts, we screened the relative OER activities of 101 first-row transition-metal perovskites using the QSAR with lowest CV prediction error. The perovskites selected were mined from the Materials Project database[31] using the *pymatgen* Python library,[47] taking the lowest hull energy structure to obtain descriptor values (specifically the M−O bond length, M−O−M bond angle, M−M distance, and Madelung site potentials). The data and source code for this screening are provided in the Supporting Information.

RESULTS AND DISCUSSION

Relationships between OER Descriptors. We first examine the OER activity trends for commonly used descriptors using the full data set in this study, which is nearly an order of magnitude larger in the number of data points explored by previous studies.[2,6,26,29,32] The relative OER activity is plotted as a function of number of d electrons, number of e_g electrons, tolerance factor, and charge-transfer energy in Figure 2. Despite large scatter in the data, general trends shown in Figure 2 are in good agreement with those reported previously: a linear trend

DOI: 10.1021/acs.jpcc.5b10071
J. Phys. Chem. C 2016, 120, 78−86

with number of d electrons[5,26] and an optimal range of values for the number of e_g electrons, consistent with the literature report of a volcano trend.[2] However, it is possible to obtain different trends when selecting only a subset of data points: for instance, near d^5 in the trend line for the number of d electrons. This ambiguity highlights the importance of exploring trends among a wide range of materials for general design strategies.

The relative OER activity shows comparable trends as a function of the tolerance factor and charge-transfer energy to the number of d electrons and e_g occupancy, which suggests the possibility of strong correlations among OER activity descriptors. From the correlation matrix of the 14 descriptors (Figure S7, Supporting Information), it is apparent that none of them is completely independent of the others. Of the 91 pairwise comparisons, 27 show fairly strong correlations ($|r| > 0.5$). Of note, the number of d electrons and optimality of e_g occupancy are well coupled ($r = -0.57$). These correlations highlight the importance of studying the behavior of descriptors together, where their respective influences on the relative OER activity of perovskites can be quantified.

We employed factor analysis to categorize the relationships among descriptors and identify those most effective at describing variations among perovskite chemistries.[48] The 14 descriptors are optimally reduced to five factors (Figure S6, Supporting Information), and their relative importance to each is given by the magnitude of their loadings[48] (Figure 3). Because each factor represents an underlying physical phenomenon that links the descriptors, they can be considered as descriptor families (Table 4). Descriptors with high loadings—including the number of d electrons, e_g occupancy, charge-transfer energy, Madelung potentials, and M−O bond length—are therefore the most useful because they govern the changes of underlying physical properties. The loading magnitudes also reveal unintuitive linkages among descriptors. In particular, although the M−O−M bond angle is a structurally defined descriptor, it is more strongly related to the covalency descriptors. The factor analysis thus provides a clearer interpretation of the physical consequence of tuning each descriptor, as well as the most important descriptors for each factor.

Table 4. Overview of Relationships among Descriptors Determined by Factor Analysis[a]

descriptor family	underlying physical phenomena	primary descriptors
covalency	hybridization of the bulk metal−oxygen bonds	charge-transfer energy
electrostatics	electrostatic effects on electrons and ions	Madelung potentials
structure	geometric effects associated with oxide crystal structure	M−O bond length
exchange interaction	electron exchange interactions	e_g electrons
electron occupancy	occupancy of the transition metal orbitals	d electrons

[a]The factors can be considered as descriptor families, related to covalency, electrostatics, structure, exchange interaction, and electron occupancy. Primary descriptors listed have loading magnitudes greater than 0.5.

Several descriptors, like the optimality of e_g and the optimality of the tolerance factor, do not have strong loadings in any factor, suggesting that they poorly describe the physical differences among materials when used alone. Notably, descriptors like the oxidation state and ionization energy have weak loadings in two factors, which indicate that they are strongly correlated with many other descriptors. The multiple correlations are also evident from the correlation matrix (Figure S7, Supporting Information), where oxidation state and ionization energy are coupled to the most descriptors (7). Strong correlations across factors therefore suggest that these descriptors are unsuitable for understanding the physical origins of OER activity, at least among the perovskites. Applying these methods across OER catalysts with a wider range of structures would allow for a broader study of the universality of these correlations across crystal families, such as the spinel, rutile, and/or pyrochlore oxides.

Comparing QSARs and Ranking OER Descriptors. We further evaluated various QSARs to gain deeper understanding of their predictive power and the relative importance of OER descriptors. Figure 4 shows the CV prediction errors for the regression models studied, estimated using the MAE. The CV error is an estimate of the error when applying the model to new data and is necessary for selecting the most accurate model. However, CV errors can overestimate the true prediction error; therefore, the prediction error was also estimated using the test data ("test error"), which provides a better estimate of the error magnitude.[7] Because the units of relative OER activity are standard deviations, the MAEs reported here are also in standard deviations.

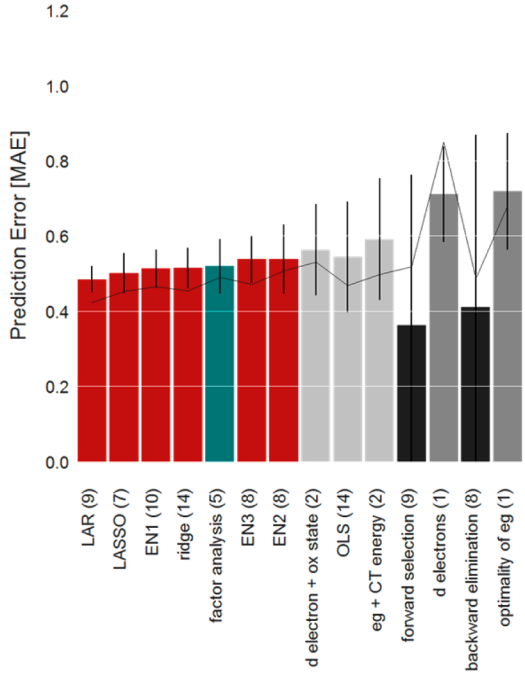


Figure 4. Mean absolute cross-validation (CV) errors (bars) and test errors (line) for the different QSARs. Relative OER activity prediction errors have units of standard deviations. Standard deviations in the MAE obtained from CV are shown by black error bars. Model classes are color coded: penalized regression (red), factor analysis (green), forward selection and backward elimination (light gray), simple OLS (gray), and multiple OLS (dark gray). The number of descriptors contributing to each model is noted parenthetically.

Single descriptors (d electrons and optimality of e_g) perform rather poorly, with mean CV errors around 0.7 s.d. and test errors around 0.7−0.9 s.d. The error magnitudes are considerably higher than the experimental error across different studies (~0.5 s.d.), indicating that the poor performance of these models is not

DOI: 10.1021/acs.jpcc.5b10071
J. Phys. Chem. C 2016, 120, 78−86

solely due to comparing measurements made by different groups. In fact, predictions from single descriptors are nearly as poor as predicting all oxides have the same OER activity as $LaCoO_3$ (MAE_{CV} = 0.83 s.d.; MAE_{test} = 1.15 s.d.). Poor predictions are avoidable by considering at least one additional descriptor. For instance, the combinations of using number of d electrons with oxidation state[5] ("d electron + ox state") or optimality of e_g occupancy with charge-transfer energy ("eg + CT energy")[2,27] both led to improvements in predictions ($MAE_{CV} \approx 0.6$ s.d.; $MAE_{test} \approx 0.5$ s.d.). However, the standard deviation in the MAE is still large (±0.2 s.d.), suggesting that predictions with these QSARs are good on average but not consistently reliable. Further addition of descriptors using a least-squares approach offers no substantial improvement and can actually weaken the model's reliability, as demonstrated by the OLS, forward-selection, and backward-elimination models.

The best predictions were obtained by using penalized methods and factor regression, both of which outperformed the least-squares approaches. Penalized and factor regression methods shrink the coefficients of some descriptors toward 0; these tools are thus particularly useful for neglecting descriptors that influence the relative OER activity via strong correlations with other variables such as the oxidation state. This results in an effective balance between the number of descriptor variables and their coefficients, accounting for the influence of more than two descriptors without weakening model precision. As evident from the test errors, the best models are capable of predicting the relative OER activity within 0.5 s.d., a substantial improvement over the single descriptors traditionally used and near the expected limit given the experimental error across studies.

In addition to generating accurate predictions, the magnitudes of the coefficients ($|\beta|$) from the QSARs describe the relative importance of OER descriptors. The factor analysis model provides a clear picture of the most important underlying physical phenomena (Table 5). Electron occupancy and

Table 5. Factor Regression Summary of Each Descriptor Family's Relative Importance to the Relative OER Activity ($|\beta|$) and *p* Values, which Indicate the Statistical Significances of $|\beta|$[a]

descriptor family	relative importance, β	*p* value
electron occupancy	0.65 ± 0.07	≪0.0001
covalency	−0.42 ± 0.07	≪0.0001
structure	0.12 ± 0.07	0.0866
exchange interaction	−0.11 ± 0.07	0.1188
electrostatic	−0.02 ± 0.07	0.7828

[a]See the Supporting Information for more detailed discussion.

covalency have the strongest influence on the relative OER activity, which validates the use of electronic descriptors for the OER activity,[2,4,5,26,27] rather than structural, magnetic, or electrostatic ones. It is unclear whether electron occupancy or covalency plays the stronger role in describing the relative OER activity given the standard deviation in their relative importances (see the Supporting Information for additional discussion). We note that this ambiguity is an instance in which the semi-quantitative nature of descriptors imposes limits on strict physical interpretation.

Even though only two of the factors are necessary to understand the dominant influence on the OER activity of perovskites, each factor is controlled by multiple descriptors. Figure 5 shows the $|\beta|$ values of the penalized regression models,

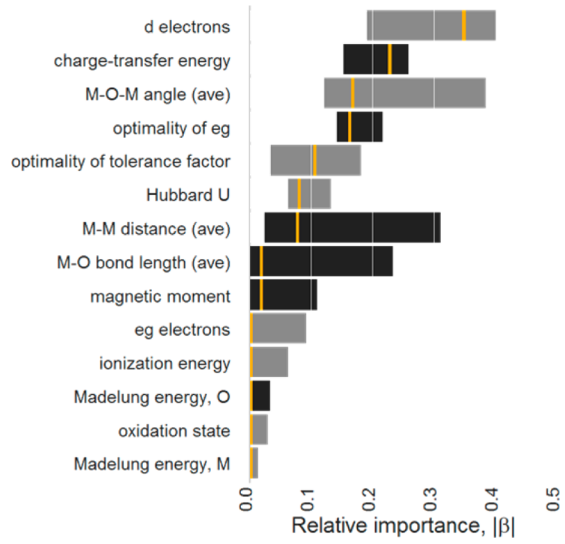


Figure 5. Relative importance of descriptor variables for the penalized regression models (elastic nets, LASSO, LAR). Descriptors are sorted by median (gold bar). Bar color indicates sign of the coefficient: negative (dark gray), positive (light gray). The relative importance of descriptors is more sensitive to the cost function used for regression (choice of model) than the optimization of the fit (Figure S8).

which illustrates the most important descriptors. The bars represent the range of $|\beta|$ values, color-coded dark gray for medians that are negatively correlated with the relative OER activity and light gray for those that are positively correlated. The top five descriptors include the three that have been discussed thoroughly in literature: d electrons, charge-transfer energy (covalency), and optimality of e_g occupancy. The lower importance of e_g occupancy optimality compared to the former two, along with its relatively weak contributions to factors in the factor analysis, suggests that it plays a more minor role in the OER activity than previously postulated.[2] Two additional secondary descriptors are also proposed here for the first time: optimality of tolerance factor and M−O−M bond angle. Despite their modest role in the factor analysis, the penalized models demonstrate that they actually play a fairly important role for generating good predictions.

In addition to identifying the most important descriptors, we also determined seven descriptors that are unnecessary for predicting the relative OER activity. These include properties such as the oxidation state, ionization energy, and Madelung potentials (all of which are more effectively captured through the charge-transfer energy), as well as the M−O bond length (captured by the tolerance factor) and exchange-interaction descriptors. The ability to rank and properly exclude potential descriptors is only made possible through large data sets and statistical learning methods that can deal with the strong correlations among descriptors.

Applications of Statistical Learning and Data Mining Tools. We conclude by discussing how this approach can complement the current materials design and selection process for oxide OER catalysts. A major advantage of obtaining larger experimental data sets is the ability to estimate reliable prediction errors for first-principle OER descriptors such as band centers.[30,49] We perform such an analysis for the recently proposed O 2p-band center descriptor[6,30] on a data set of 32 measurements on 12 oxide chemistries. The O 2p-band center values were calculated using the same method reported

DOI: 10.1021/acs.jpcc.5b10071
J. Phys. Chem. C 2016, 120, 78−86

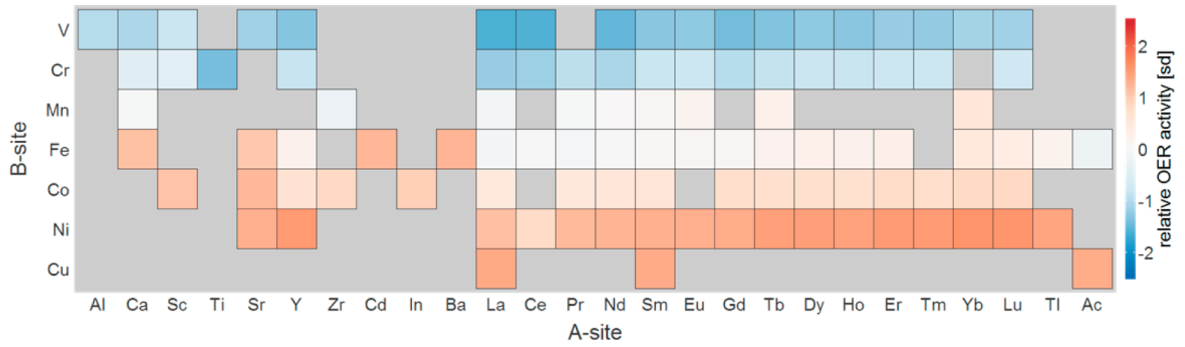


Figure 6. Heatmap of relative OER activity predictions for ABO_3 perovskites using the best-performing model (LAR) and data mined from the Materials Project database. Warmer colors indicate higher relative OER activity.

previously[6,30] and are provided in the CSV file in the Supporting Information. The mean absolute CV error using 5-fold CV was 0.634 ± 0.279 s.d., which is on par with using optimality of e_g and charge-transfer energy as descriptors (0.592 ± 0.162 s.d.) and notably worse than the best-performing model (0.484 ± 0.035 s.d.). Thus, the O 2p-band center is a single descriptor that effectively describes both electron occupancy and covalency, but aggregating simple descriptors can still generate a more powerful predictive tool for identifying perovskite OER catalysts.

Another application is rapid virtual screening of relative OER activities. Because the models are simple to calculate using a priori descriptors, large-scale screening can be performed in short times. Using ABO_3 perovskites from the Materials Project database, estimates of the relative OER activity from the best-performing model (LAR) are illustrated in a heatmap in Figure 6. A clear trend with transition metal appears that is similar to that observed in Figure 1, where higher activities occur for late-transition metal oxides. Interestingly, isovalent substitutions of the A-site cation can have a moderate influence on the relative OER activity, which may be related to the influence of structural descriptors such as the tolerance factor and average M−O−M bond angle. However, it is important to note again that these models are tools for identifying promising systems and exploring potentially interesting differences among materials; there remains a strong need to investigate trends using more physically meaningful approaches to fully understand the effects of perovskite chemistry on the OER activity.

CONCLUSION

We have aggregated 101 observations of 51 perovskites from literature and our own experimental data. We find that 14 descriptors of the metal−oxygen bond strength fall under five basic descriptor families, associated with the oxide's metal−oxygen covalency, electrostatics, structure, exchange interactions, and transition-metal electron occupancy. Two of these families, covalency and electron occupancy, hold the strongest influence on the OER activity, but a linear combination of nine descriptors is still necessary for obtaining the best predictive models, largely outperforming those that only include one or two descriptors. The number of d electrons and charge-transfer energy (covalency) were found to play the most important roles, while other factors such as optimality of e_g occupancy, M−O−M bond angle, and tolerance factor were found to be secondary descriptors. The tools explored in this study illustrate how statistical learning is a powerful approach for improving descriptor-based design strategies and can complement material databases to generate high-throughput predictions across a wide chemical space.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcc.5b10071.

Additional discussions on the OER activity standardization, regression models, analysis protocol, descriptor correlations, select regression model statistics (PDF)

Data file of the OER activities and descriptor values used (.csv), source code used for the analysis (R CRAN), and source code used for mining and extracting descriptor values for Materials Project predictions (Python) (ZIP)

AUTHOR INFORMATION

Corresponding Authors

*(W.T.H.) E-mail: whong@mit.edu. Phone: +1 (617) 324-3718.
*(Y.S.-H.) E-mail: shaohorn@mit.edu. Phone: +1 (617) 253-2259.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Kevin J. May, Alexis Grimaud, and Yueh-Lin Lee for collecting data used for the statistical analysis. This work was supported in part by the U.S. Department of Energy (SISGR DE-SC0002633) and the Skoltech-MIT Center for Electrochemical Energy Storage.

REFERENCES

(1) Hong, W. T.; Risch, M.; Stoerzinger, K. A.; Grimaud, A.; Suntivich, J.; Shao-Horn, Y. Toward the Rational Design of Non-Precious Transition Metal Oxides for Oxygen Electrocatalysis. *Energy Environ. Sci.* **2015**, *8*, 1404−1427.

(2) Suntivich, J.; May, K. J.; Gasteiger, H. A.; Goodenough, J. B.; Shao-Horn, Y. A Perovskite Oxide Optimized for Oxygen Evolution Catalysis from Molecular Orbital Principles. *Science* **2011**, *334*, 1383−1385.

(3) Matsumoto, Y.; Sato, E. Electrocatalytic Properties of Transition Metal Oxides for Oxygen Evolution Reaction. *Mater. Chem. Phys.* **1986**, *14*, 397−426.

(4) Man, I. C.; Su, H.-Y.; Calle-Vallejo, F.; Hansen, H. A.; Martínez, J. I.; Inoglu, N. G.; Kitchin, J.; Jaramillo, T. F.; Nørskov, J. K.; Rossmeisl, J. Universality in Oxygen Evolution Electrocatalysis on Oxide Surfaces. *ChemCatChem* **2011**, *3*, 1159−1165.

(5) Calle-Vallejo, F.; Inoglu, N. G.; Su, H.-Y.; Martínez, J. I.; Man, I. C.; Koper, M. T. M.; Kitchin, J. R.; Rossmeisl, J. Number of Outer Electrons

DOI: 10.1021/acs.jpcc.5b10071
J. Phys. Chem. C 2016, 120, 78−86

as Descriptor for Adsorption Processes on Transition Metals and Their Oxides. *Chem. Sci.* **2013**, *4*, 1245−1249.
(6) Grimaud, A.; May, K. J.; Carlton, C. E.; Lee, Y. L.; Risch, M.; Hong, W. T.; Zhou, J.; Shao-Horn, Y. Double Perovskites as a Family of Highly Active Catalysts for Oxygen Evolution in Alkaline Solution. *Nat. Commun.* **2013**, *4*, 2439.
(7) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer-Verlag: Berlin, 2009.
(8) Xiang, C.; Suram, S. K.; Haber, J. A.; Guevarra, D. W.; Soedarmadji, E.; Jin, J.; Gregoire, J. M. High-Throughput Bubble Screening Method for Combinatorial Discovery of Electrocatalysts for Water Splitting. *ACS Comb. Sci.* **2014**, *16*, 47−52.
(9) Neyerlin, K. C.; Bugosh, G.; Forgie, R.; Liu, Z.; Strasser, P. Combinatorial Study of High-Surface-Area Binary and Ternary Electrocatalysts for the Oxygen Evolution Reaction. *J. Electrochem. Soc.* **2009**, *156*, B363.
(10) Rajan, K. Combinatorial Materials Sciences: Experimental Strategies for Accelerated Knowledge Discovery. *Annu. Rev. Mater. Res.* **2008**, *38*, 299−322.
(11) Zhang, C.; Fagan, R. D.; Smith, R. D. L.; Moore, S. A.; Berlinguette, C. P.; Trudel, S. Mapping the Performance of Amorphous Ternary Metal Oxide Water Oxidation Catalysts Containing Aluminum. *J. Mater. Chem. A* **2015**, *3*, 756−761.
(12) Haber, J. A.; Xiang, C.; Guevarra, D.; Jung, S.; Jin, J.; Gregoire, J. M. High-Throughput Mapping of the Electrochemical Properties of (Ni-Fe-Co-Ce)O_x Oxygen-Evolution Catalysts. *ChemElectroChem* **2014**, *1*, 524−528.
(13) McCrory, C. C.; Jung, S.; Peters, J. C.; Jaramillo, T. F. Benchmarking Heterogeneous Electrocatalysts for the Oxygen Evolution Reaction. *J. Am. Chem. Soc.* **2013**, *135*, 16977−16987.
(14) McCrory, C. C.; Jung, S.; Ferrer, I. M.; Chatman, S. M.; Peters, J. C.; Jaramillo, T. F. Benchmarking Hydrogen Evolving Reaction and Oxygen Evolving Reaction Electrocatalysts for Solar Water Splitting Devices. *J. Am. Chem. Soc.* **2015**, *137*, 4347−4357.
(15) Haber, J. A.; Cai, Y.; Jung, S.; Xiang, C.; Mitrovic, S.; Jin, J.; Bell, A. T.; Gregoire, J. M. Discovering Ce-rich Oxygen Evolution Catalysts, From High Throughput Screening to Water Electrolysis. *Energy Environ. Sci.* **2014**, *7*, 682.
(16) Rajan, K. Materials Informatics. *Mater. Today* **2005**, *8*, 38−45.
(17) Suram, S. K.; Haber, J. A.; Jin, J.; Gregoire, J. M. Generating Information-Rich High-Throughput Experimental Materials Genomes using Functional Clustering via Multitree Genetic Programming and Information Theory. *ACS Comb. Sci.* **2015**, *17*, 224−33.
(18) Kusne, A. G.; Gao, T.; Mehta, A.; Ke, L.; Nguyen, M. C.; Ho, K. M.; Antropov, V.; Wang, C. Z.; Kramer, M. J.; Long, C.; et al. On-the-Fly Machine-Learning for High-Throughput Experiments: Search for Rare-Earth-Free Permanent Magnets. *Sci. Rep.* **2014**, *4*, 6367.
(19) Dey, P.; Bible, J.; Datta, S.; Broderick, S.; Jasinski, J.; Sunkara, M.; Menon, M.; Rajan, K. Informatics-Aided Bandgap Engineering for Solar Materials. *Comput. Mater. Sci.* **2014**, *83*, 185−195.
(20) Majid, A.; Khan, A.; Javed, G.; Mirza, A. M. Lattice Constant Prediction of Cubic and Monoclinic Perovskites Using Neural Networks and Support Vector Regression. *Comput. Mater. Sci.* **2010**, *50*, 363−372.
(21) Majid, A.; Khan, A.; Choi, T.-S. Predicting Lattice Constant of Complex Cubic Perovskites Using Computational Intelligence. *Comput. Mater. Sci.* **2011**, *50*, 1879−1888.
(22) Corma, A.; Serra, J.; Serna, P.; Moliner, M. Integrating High-Throughput Characterization into Combinatorial Heterogeneous Catalysis: Unsupervised Construction of Quantitative Structure/Property Relationship Models. *J. Catal.* **2005**, *232*, 335−341.
(23) Smotkin, E. S.; Díaz-Morales, R. R. New Electrocatalysts by Combinatorial Methods. *Annu. Rev. Mater. Res.* **2003**, *33*, 557−579.
(24) Corma, A.; Moliner, M.; Serra, J.; Serna, P.; Díaz-Cabanas, O. A.; Baumes, L. A New Mapping/Exploration Approach for HT Synthesis of Zeolites. *Chem. Mater.* **2006**, *18*, 3287−3296.
(25) Tseung, A. C. C.; Jasem, S. Oxygen Evolution on Semiconducting Oxides. *Electrochim. Acta* **1977**, *22*, 31−34.
(26) Bockris, J. O. M. The Electrocatalysis of Oxygen Evolution on Perovskites. *J. Electrochem. Soc.* **1984**, *131*, 290−302.
(27) Suntivich, J.; Hong, W. T.; Lee, Y.-L.; Rondinelli, J. M.; Yang, W.; Goodenough, J. B.; Dabrowski, B.; Freeland, J. W.; Shao-Horn, Y. Estimating Hybridization of Transition Metal and Oxygen States in Perovskites from O K-edge X-ray Absorption Spectroscopy. *J. Phys. Chem. C* **2014**, *118*, 1856−1863.
(28) Zaanen, J.; Sawatzky, G. A. Systematics in Band Gaps and Optical Spectra of 3D Transition Metal Compounds. *J. Solid State Chem.* **1990**, *88*, 8−27.
(29) Matsumoto, Y.; Yamada, S.; Nishida, T.; Sato, E. Oxygen Evolution on $La_{1-x}Sr_xFe_{1-y}Co_yO_3$ Series Oxides. *J. Electrochem. Soc.* **1980**, *127*, 2360−2364.
(30) Lee, Y.-L.; Kleis, J.; Rossmeisl, J.; Shao-Horn, Y.; Morgan, D. Prediction of Solid Oxide Fuel Cell Cathode Activity with First-Principles Descriptors. *Energy Environ. Sci.* **2011**, *4*, 3966−3970.
(31) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; et al. The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, *1*, 011002.
(32) Jain, A. N.; Tiwari, S. K.; Singh, R. N.; Chartier, P. Low-temperature Synthesis of Perovskite-type Oxides of Lanthanum and Cobalt and Their Electrocatalytic Properties for Oxygen Evolution in Alkaline Solutions. *J. Chem. Soc., Faraday Trans.* **1995**, *91*, 1871−1875.
(33) Giordano, L.; Han, B.; Risch, M.; Hong, W. T.; Rao, R. R.; Stoerzinger, K. A.; Shao-Horn, Y. pH Dependence of OER Activity of Oxides: Current and Future Perspectives. *Catal. Today* **2016**, *262*, 2−10.
(34) Calle-Vallejo, F.; Díaz-Morales, O. A.; Kolb, M. J.; Koper, M. T. M. Why Is Bulk Thermochemistry a Good Descriptor for the Electrocatalytic Activity of Transition Metal Oxides? *ACS Catal.* **2015**, *5*, 869−873.
(35) Goldschmidt, V. M. *Naturwissenschaften* **1926**, *14*, 477−485.
(36) Torrance, J. B.; Lacorre, P.; Asavaroengchai, C.; Metzger, R. M. Why are Some Oxides Metallic, While Most are Insulating? *Phys. C* **1991**, *182*, 351−364.
(37) Suntivich, J.; Gasteiger, H. A.; Yabuuchi, N.; Nakanishi, H.; Goodenough, J. B.; Shao-Horn, Y. Design Principles for Oxygen-Reduction Activity on Perovskite Oxide Catalysts for Fuel Cells and Metal-Air Batteries. *Nat. Chem.* **2011**, *3*, 546−550.
(38) Lufaso, M. W.; Woodward, P. M. Prediction of the Crystal Structures of Perovskites Using the Software Program *SPuDS*. *Acta Crystallogr., Sect. B: Struct. Sci.* **2001**, *B57*, 725−738.
(39) Zhou, W.; Sunarso, J. Enhancing Bi-functional Electrocatalytic Activity of Perovskite by Temperature Shock: A Case Study of $LaNiO_{3-\delta}$. *J. Phys. Chem. Lett.* **2013**, *4*, 2982−2988.
(40) Ezbiri, M.; Allen, K. M.; Galvez, M. E.; Michalsky, R.; Steinfeld, A. Design Principles of Perovskites for Thermochemical Oxygen Separation. *ChemSusChem* **2015**, *8*, 1966−71.
(41) Momma, K.; Izumi, F. *VESTA 3* for Three-Dimensional Visualization of Crystal, Volumetric and Morphology Data. *J. Appl. Crystallogr.* **2011**, *44*, 1272−1276.
(42) Iwakura, C.; Nishioka, M.; Tamura, H. Oxygen Evolution on Spinel-type Ferrite Film Electrodes. *Nippon Kagaku Kaishi* **1982**, 1136−1140.
(43) R Core Team *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2014.
(44) Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Soft.* **2010**, *33*, 1−22.
(45) Hastie, T.; Efron, B. *lars: Least Angle Regression, Lasso and Forward Stagewise* **2013**.
(46) Raiche, G. *An R Package for Parallel Analysis and Non Graphical Solutions to the Cattell Scree Test*, R package version 2.3.3; 2010.
(47) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A Robust, Open Source Python Library for Materials Analysis. *Comput. Mater. Sci.* **2013**, *68*, 314−319.
(48) Child, D. *The Essentials of Factor Analysis*, 3rd ed.; Bloomsbury Academic: London, 2006.
(49) Hammer, B.; Norskov, J. K. Electronic Factors Determining the Reactivity of Metal Surfaces. *Surf. Sci.* **1995**, *343*, 211−220.

DOI: 10.1021/acs.jpcc.5b10071
J. Phys. Chem. C 2016, 120, 78−86