CO Oxidation Catalyzed by Perovskites: The Role of Crystallographic Distortions Highlighted by Systematic Experiments and AI

Giulia Bellini,[a] Gregor Koch,[a] Frank Girgsdies,[a] Jinhu Dong,[a] Spencer J. Carey,[a] Olaf Timpe,[a] Gudrun Auffermann,[b] Matthias Scheffler,[c] Robert Schlögl,[a] Lucas Foppa,*[c] and Annette Trunschke*[a]

[a] Dr. Giulia Bellini, Dr. Gregor Koch, Dr. Frank Girgsdies, Dr. Jinhu Dong, Dr. Spencer J. Carey, Dr. Olaf Timpe, Prof. Dr. Robert Schlögl, Dr. Annette Trunschke*
Inorganic Chemistry Department
Fritz-Haber-Institut der Max-Planck-Gesellschaft
Faradayweg 4-6, 14195 Berlin, Germany
E-mail: trunschke@fhi-berlin.mpg.de

[b] Dr. Gudrun Auffermann
Max-Planck-Institut für Chemische Physik Fester Stoffe
Nöthnitzer Straße 40, 01187 Dresden, Germany

[c] Prof. Dr. Matthias Scheffler, Dr. Lucas Foppa*
The NOMAD Laboratory at the Fritz Haber Institute of the Max Planck Society
Faradayweg 4-6, 14195 Berlin, Germany
E-mail: foppa@fhi-berlin.mpg.de

Abstract: The identification of key materials' parameters that correlate with the catalytic performance can accelerate the development of improved heterogeneous catalysts and unveil the relevant underlying physical processes. However, the analysis of correlations is often hindered by inconsistent data. Besides, nontrivial, yet unknown relationships may be important, and the intricacy of the various processes may be significant. Here, we tackle these challenges for the CO oxidation reaction catalyzed by perovskites via a combination of rigorous experiments and artificial intelligence (AI). A series of 13 ABO_3 (*A* = La, Pr, Nd, Sm; *B* = Cr, Mn, Fe, Co) perovskites was synthesized, characterized, and tested in catalysis. To the resulting dataset, we applied the symbolic-regression SISSO approach. We identified an analytical expression as a descriptor for the activity that contains, as key materials' parameters, the normalized unit cell volume, the Pauling electronegativity of the elements *A* and *B*, and the ionization energy of the element *B*. Therefore, the activity is described by crystallographic distortions and by the chemical nature of *A* and *B* elements. The generalizability of the identified descriptor is confirmed by the good overall quality of the predictions for the activity of 3 additional ABO_3 and of 16 chemically more complex $AMn_{(1-x)}B'_xO_3$ (*A* = La, Pr, Nd; *B'* = Fe, Co Ni Cu, Zn) perovskites. These $AMn_{(1-x)}B'_xO_3$ materials contain substitutions of Mn at the *B* sites as well as chemical elements that were not part of the training set (Ni, Cu, Zn).

Introduction

Perovskites[1] are a promising materials' class[2] for energy-related applications and key to future catalysis technologies.[2b] Applications of perovskite oxides include environmental catalysis[3] and electrocatalysis.[2b] Due to the high flexibility of the perovskite structure in terms of composition, an immense number of compounds can be obtained by combining different elements in ABO_3 materials and by varying the stoichiometry, e.g., in $AB_{(1-x)}B'_xO_3$ or $A_{(1-x)}A'_xBO_3$ systems. In the perovskite structure, relatively large *A* cations (e.g., lanthanides) occupy sites that are coordinated with 12 oxygen anions (O^{2-}) and smaller *B* cations (e.g., transition metals) are coordinated with 6 oxygen anions to form BO_6 corner-sharing octahedra. The stability of the perovskite structure relative to other possible structures and the distortion of the crystal structure can be estimated from the ratio of the ionic radii of the constituent elements via tolerance factors.[4] The properties and catalytic performance of perovskites can be adjusted by the composition. However, elucidating the underlying relationships between materials' parameters and the performance, known as *descriptors*, and efficiently exploring the virtually infinite space of possible perovskites are challenging tasks. This is because heterogeneous catalysis is governed by a high intricacy of many superimposed and entangled processes such as the interplay between bulk and surface properties, the surface reactions, and the dynamic restructuring of the material under working conditions. This prevents the explicit, atomistic modelling of the full catalytic progression.[5] Indeed, the surface motifs that are present under reaction conditions and the reaction mechanisms that operate on those surfaces are typically unknown and continuously adapt to the progress of the reaction and the associated changes in chemical potential.

The CO oxidation to CO_2 by molecular oxygen is an example of an important reaction for environmental catalysis and for the purification of hydrogen for energy generation.[6] This reaction can be catalyzed by perovskite oxides.[3, 6-7] However, the mechanisms operating on perovskites are not well understood. The performance has been related to different underlying processes, such as the transfer of lattice oxygen to reacting species and the redox processes occurring on *B* sites.[8] Accordingly, the activity was correlated to the perovskite capacity to provide lattice oxygen and to the abundance and redox potential of surface *B* species. Additionally, the choice of the *A* element can result in crystallographic strain[9] and thus in deviations from the ideal cubic perovskite structure, which can in turn affect the surface properties and catalytic performance.[10] In

particular, small *A* cations induce octahedral tilting, creating more covalent bonding structures compared to those obtained with larger *A* cations.[11] It turns out that the reported relationships describing the performance of perovskites in CO oxidation are partly contradictory. In particular, the conclusions depend on the composition space of the investigated perovskites. Furthermore, to the best of our knowledge, the influence of *A* and *B* elements has not yet been studied simultaneously and systematically.

In order to unveil descriptors and to design new materials, the identification of key materials' parameters correlated with the catalytic performance is crucial. Artificial intelligence (AI) has been increasingly applied to identify nonlinear correlations and complex patterns in data in materials science and catalysis.[12] However, the inconsistency of experimental data in heterogeneous-catalysis research hinders the application of AI to reported data (see ref. [13] and references therein). Indeed, the measured materials properties, e.g., surface compositions, depend on the experimental workflow of sample synthesis, the pretreatment, and the measurement conditions. Besides, the measured catalytic performance is strongly influenced by the reaction conditions, e.g., start-up procedure (induction period under the reaction feed, also referred to as catalyst formation or activation), temperature, contact time, or feed composition, because the surface and solid-state chemistry of the material is coupled with the chemistry of the catalytic reaction. Finally, the presence of small amounts of by-phases formed as minor impurities during synthesis can also lead to inconsistencies when searching for structure-performance relationships. In addition to data inconsistency, the extremely small number of materials that can be accessed by detailed experiments prevents the application of widely used AI methods such as neural networks. These AI methods typically require orders of magnitude more data points compared to the number of materials that are investigated by systematic experiments.

To address these issues, we recently proposed a combination of rigorous experimental protocols, which produce consistent data,[14] with the data-efficient symbolic-regression AI approach SISSO (Sure-Independence Screening and Sparsifying Operator).[15] This strategy enables the identification of the key physicochemical parameters correlated with the performance, out of many offered candidate descriptive parameters, termed *primary features*. These primary features characterize the materials and correlate with many potentially relevant underlying processes. In analogy to genes in biology, the key parameters identified in a SISSO analysis were called "materials genes", since they describe the materials behavior similar to how genes in biology and medicine describe characteristics such as hair color, i.e., they capture complex patterns without providing the full, explicit description of all underlying processes. Indeed, the relationships between the identified key materials parameters and the performance might be indirect. SISSO identified models in the form of analytical expressions. These expressions are the pursued descriptors. SISSO can be applied even to the small number of materials that is accessed by rigorous experiments in catalysis and physical chemistry. We note, however, that SISSO and AI in general provide a statistical description, which is associated to a certain variance or uncertainty and a certain domain of applicability.[16]

In this paper, we utilize the materials-genes concept to model and design perovskite catalysts, using the CO oxidation as a model reaction. Crucially, we analyze the influence of a wide range of *A* and *B* elements on perovskite reactivity simultaneously by considering 13 ABO_3 perovskites composed by combinations of the *A* elements La, Pr, Nd, and Sm with the *B* elements Cr, Mn, Fe, and Co. Based on this training set, we identify an analytical expression correlated with the activity that contains, as key parameters, the normalized unit cell volume, the electronegativity of *A* and *B* elements, and the ionization energy of the *B* element. This descriptor highlights the so far overlooked role of crystallographic distortions on the reactivity of the perovskites, in addition to catalyst properties encoded in the chemical nature of *A* and *B* elements. The identified expression is used to predict the activity of an additional set of 3 ABO_3 and 16 $AMn_{(1-x)}B'_xO_3$ (*A* = La, Pr, Nd; *B'* = Fe, Co Ni Cu, Zn;) perovskites that was not used for training. The good overall quality of the predictions suggest that the expression is generalizable.

Results and Discussion

Synthesis, characterization, and catalysis

We synthesized 13 ABO_3 (*A* = La, Pr, Nd, Sm; *B* = Cr, Mn, Fe, Co) perovskites by the solution combustion method.[17] The perovskites were characterized by inductively coupled plasma optical emission spectroscopy (ICP OES) or equivalent X-ray fluorescence analysis (XRF), oxygen analysis, X-ray diffraction (XRD), X-ray photoelectron spectroscopy (XPS), and N_2 adsorption. Then, we tested the perovskites in CO oxidation. The preparation procedure, characterization techniques and catalytic testing are described in detail in the Electronic Supporting Information (ESI).

XRD reveals that all samples present a perovskite-like structure (Table 1). The materials are phase-pure, with the exception of $NdFeO_3$, for which we identified small amounts of a by-phase (2% Nd_2O_3). The orthorhombic crystal structure is observed for the majority of the synthesized materials. Only $LaMnO_3$ and $LaCoO_3$ perovskites exhibit the rhombohedral crystal symmetry. In order to quantify the distortions of the perovskites with respect to the ideal cubic structure, we determined the observed tolerance factor t_{obs} based on the Rietveld refinement analysis using the expression

$$t_{obs} = \frac{<AO>_\mu}{<BO>_\mu * \sqrt{2}} \quad . \qquad \text{(Eq. 1)}$$

Here, $< AO >_\mu$ and $< BO >_\mu$ are the mean values of the interatomic *A*-O and *B*-O distances, respectively. The observed tolerance factor for a cubic perovskite is equal to 1.00. For the synthesized materials, the values of t_{obs} are in the range [0.978, 0.993] (Table 1). For a fixed *B* element, the t_{obs} values decrease according to the *A* element following the trend La > Pr > Nd > Sm. This shows that the octahedral tilting and crystallographic strain increases (and t_{obs} decreases) as the size of the *A* species decreases.[11b] Additional parameters were obtained using the Rietveld refinement analysis. These parameters were defined so

that materials presenting different crystal symmetry (orthorhombic and rhombohedral) could be compared in a consistent manner. The normalized cell volume (V') is an example of such parameter. The quantity V' is the unit-cell volume divided by the number of ABO_3 formula units in the unit cell. It measures the packing of the atoms in the structure and it is determined by the size of *A* and *B* cations and by the distortions of the octahedra. In total, 13 parameters characterizing the perovskite structures and their distortions were obtained from XRD (see full list in Table S1). The formulas for defining these parameters are explained in detail in the ESI and the complex crystallographic characteristics of perovskite oxides are discussed in detail elsewhere.[11b]

ICP OES (or XRF), oxygen analysis, and XPS were used to measure the *A*, *B*, and oxygen contents in the bulk and on the surface of the perovskites. The bulk and surface contents of *B*, measured for the 13 ABO_3 perovskites in atomic percentage and denoted x_B^{bulk} and x_B^{surf}, respectively, are shown in Table 1. The x_B^{bulk} values are similar among the 13 materials and they are in the range [0.17, 0.22]. In contrast, the x_B^{surf} values are more spread in the range [0.03, 0.22]. Thus, the surface stoichiometry varies more than the bulk stoichiometry among the considered materials. In particular, materials such as $LaFeO_3$ and $NdFeO_3$ present significantly lower content of *B* in the surface (0.03) compared to the bulk (0.21 and 0.22, respectively). However, there is no clear trend linking the bulk and surface compositions with the *A* and *B* elements. In total, ICP OES (or XRF), oxygen analysis, and XPS provided 6 parameters characterizing the perovskite bulk and surface compositions. Finally, the N_2 physisorption provided the specific surface area of the perovskites (Table 1). The values of surface area are in the range [0.60, 13.40] $m^2 \cdot g^{-1}$.

The perovskites were tested in CO oxidation in an automated reactor[18] according to a standard operating procedure. We assessed the catalytic performance by analyzing the temperature required to reach 10% CO conversion, hereafter denoted T_{10} (Table 1). The value of T_{10} reflects the activity, and the lower the T_{10}, the higher the activity. The T_{10} values are in the range [116, 435] °C, showing that the considered training set of perovskites present remarkably diverse activity. The T_{10} generally depends on the *B* element and increases in the following order: Co < Mn < Fe < Cr. However, there are exceptions, e.g., $T_{10}^{\text{SmCrO}_3} < T_{10}^{\text{SmFeO}_3}$ or $T_{10}^{\text{NdMnO}_3} < T_{10}^{\text{PrCoO}_3}$. The least and most active materials of the 13 perovskites in the training set are $PrCrO_3$ and $LaCoO_3$, respectively, with T_{10} values of 435 and 116 °C, respectively.

Table 1. Measured reactivity and some properties of the 13 ABO_3 (*A* = La, Pr, Nd, Sm; *B* = Cr, Mn, Fe, Co) perovskites used to train the SISSO model.

Material	Space Group	T_{10} (°C)[c]	t_{obs}[d]	V' (AA)[e]	x_B^{bulk}[f]	x_B^{surf}[g]	s_{BET} (m^2/g)[h]	Material ID[i]
$LaCrO_3$	*Pbnm*[a]	425	0.989	58.66	0.20	0.15	12.85	27499
$PrCrO_3$	*Pbnm*[a]	435	0.984	57.61	0.20	0.14	12.10	29306
$NdCrO_3$	*Pbnm*[a]	403	0.983	57.21	0.19	0.14	11.10	29307
$SmCrO_3$	*Pbnm*[a]	371	0.981	56.47	0.19	0.16	11.40	29310
$LaMnO_3$	*R*-3*cH*[b]	180	0.993	58.43	0.17	0.17	13.40	28837
$PrMnO_3$	*Pbnm*[a]	152	0.987	57.91	0.20	0.21	5.68	30276
$NdMnO_3$	*Pbnm*[a]	137	0.985	57.75	0.17	0.05	5.68	30283
$LaFeO_3$	*Pbnm*[a]	268	0.987	60.71	0.21	0.03	3.12	24878
$PrFeO_3$	*Pbnm*[a]	259	0.982	59.6	0.19	0.04	1.82	25041
$NdFeO_3$	*Pbnm*[a]	297	0.981	59.11	0.22	0.03	4.49	24890
$SmFeO_3$	*Pbnm*[a]	374	0.978	58.24	0.21	0.22	0.60	34224
$LaCoO_3$	*R*-3*cH*[b]	116	0.992	56.04	0.19	0.10	18.5	29645
$PrCoO_3$	*Pbnm*[a]	147	0.988	54.39	0.18	0.19	7.50	30380

[a] *Pbnm*: orthorhombic symmetry. [b]*R*-3*cH*: rhombohedral symmetry. [c] Temperature for 10% CO conversion. [d] Observed tolerance factor measured by XRD. [e] Normalized unit-cell volume measured by XRD. [f] Atomic fractions of *B* in the bulk measured by ICP OES. [g] Atomic fractions of *B* on the surface measured by XPS. [h] Specific surface area measured by N_2 physisorption. [i]Samples ID according to the internal database[19] of the Fritz Haber Institute for unique identification of the sample batches.

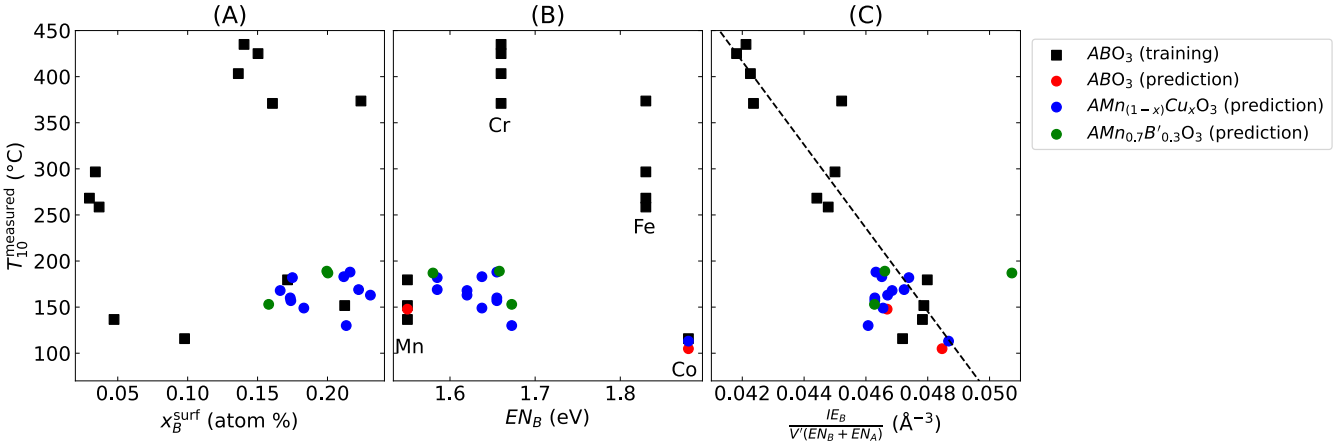

Figure 1. The activity of the 13 ABO_3 (*A* = La, Pr, Nd, Sm; *B* = Cr, Mn, Fe, Co) perovskites is not appropriately described by one single parameter, but it is well captured by the nonlinear analytical expression (descriptor) depending on multiple parameters identified by SISSO. The temperature for achieving 10% CO conversion (T_{10}) is plotted vs. (A) the atomic fraction of *B* on the surface measured by XPS (x_B^{surf}), (B) vs. the electronegativity of the *B* element (EN_B), and (C) vs. the descriptor identified by SISSO. This descriptor contains the key parameters reduced unit-cell volume (V'), electronegativity of *A* and *B* elements (EN_A and EN_B), and the ionization energy of *B* element (IE_B). The black squares correspond to the 13 ABO_3 perovskites of the training set. The red, blue, and green circles correspond to, respectively, 3 additional ABO_3 perovskites ($SmMnO_3$, $SmCoO_3$, and $NdCoO_3$), 12 $AMn_{(1-x)}Cu_xO_3$ (*A* = La, Pr) perovskites, and 4 $AMn_{0.7}B'_xO_3$ (*A* = Pr, Nd; *B'*= Fe, Co, Ni, Zn) perovskites. The SISSO descriptor correctly predicts the activity of most of the materials that were not used to train the model, displayed with circle markers. The dashed line in (C) corresponds to the SISSO model of Eq. 3.

In an attempt to identify simple, linear correlations describing the activity, we plotted T_{10} as a function of x_B^{surf} (Figure 1A, black squares), since the surface amount of *B* has been related to the activity of perovskites in CO oxidation.[8] However, we do not see a correlation between T_{10} and x_B^{surf}. No correlation is observed between T_{10} and the parameters x_A^{bulk}, x_B^{bulk}, x_O^{bulk}, $x_A^{surface}$, and $x_O^{surface}$ either (Figure S1). We have also verified whether T_{10} is correlated with the Pauling electronegativity EN of the *B* element, denoted EN_B (Figure 1B, black squares). We observe that T_{10} increases, reaches a maximum, and then decreases as a function of EN_B. However, there are significant differences in the T_{10} values according to the *A* elements when the *B* element is fixed. The labels in Figure 1B indicate the *B* elements associated with the points falling within the respective vertical lines. This shows that the choice of the *A* element has a significant influence on the activity, in addition to the *B* element. We identified no clear correlation between T_{10} and EN_A or the ionization energy (IE) of the *A* and *B* elements (Figure S1). In summary, nor the relative amount of *A*, *B* or oxygen in the bulk or on the surface neither the elemental properties of *A* or *B* elements alone are able to describe the reactivity of the perovskites towards CO oxidation across all the 13 ABO_3 materials.

AI analysis

In order to identify a descriptor that can appropriately capture the activity towards CO oxidation for the 13 ABO_3 (*A* = La, Pr, Nd, Sm; *B* = Cr, Mn, Fe, Co) materials, we applied the SISSO approach.[20] This method identifies models of a target of interest $y^{(SISSO)}$ as analytical expressions containing key parameters, out of many offered primary features. SISSO was used to derive, for instance, a tolerance factor indicating the perovskite stability[4b] that is more accurate than the Goldschmidt tolerance factor.[4a] In this paper, the target that we will model is T_{10}.

The SISSO approach starts with the collection of primary features. Then, a large number (e.g., up to millions) of expressions are generated by combining the primary features according to unary and binary mathematical operators such as addition, multiplication, division, and exponential. Finally, compressed sensing is applied to identify the few expressions, which combined by weighing coefficients, best correlate with the target for the dataset. The models identified by SISSO for an arbitrary target y have the general form

$$y^{(SISSO)} = c_0 + \sum_{i=1}^{N} c_i d_i, \text{ (Eq. 2)}$$

where c_i are fitted weighing coefficients and d_i are the selected analytical expressions, which correspond to components of a descriptor vector. Because the functional forms of SISSO models are flexible and can easily fit the rather small number of materials considered in this work, the appropriate model complexity needs to be determined. This is done to avoid identifying models that fit the training data but which are not generalizable. We evaluated the optimal complexity with respect to the predictability through a leave-one-material-out-cross-validation (CV) approach. Details of the SISSO analysis are provided in the ESI.[15a]

The 20 parameters obtained from all the characterization techniques were used as primary features in the SISSO analysis. Additionally, we included the elemental parameters electronegativity and ionization energy that reflect the chemical nature of the *A* and *B* elements of the perovskite, EN_A, EN_B, IE_A and IE_B. These physicochemical parameters are associated to free atoms. They were measured are are reported in the literature (see values in Table S3).[21] In total, 24 primary features were considered in our SISSO analysis. The full list of primary features is shown in Table S1 of the ESI. The values of these primary features for all the materials discussed in this paper are available in the dataset.xls supplementary file.

The best model identified by SISSO for T_{10} correctly captures the reactivity trend among the 13 ABO_3 perovskites (Figure 1C). This model has the following expression:

$$T_{10}^{(\mathrm{SISSO})} = 2316 - 45227\left(\frac{IE_B}{V\prime(EN_A + EN_B)}\right). \text{ (Eq. 3)}$$

The key parameters that appear in Eq. 3 are the normalized unit-cell volume (V') determined by XRD, the Pauling electronegativity of the *A* and *B* elements (EN_A and EN_B, respectively), and the ionization energy of the *B* element (IE_B). The relevance of V' can be associated to the role of crystallographic distortions. Indeed, octahedral tilting induced by different cation sizes[11a] can alter the packing and the electronic structure of the perovskites and thus the bonding patterns.[9] In particular, the balance between ionic and covalent bonding can change. These different bulk bonding patterns consequently translate into different surface reactivity. The relevance of IE_B reflects the redox activity of the B species, as the *B* element acts as a redox-active element in oxidation reactions catalyzed by perovskites. Finally, the relevance of EN_A and EN_B reflects the electronic properties of the catalyst, which modulates, for instance, the strength of adsorption of surface reactive species such as CO. The atomic fractions of *A*, *B,* and oxygen species in the bulk and on the surface were also offered as primary features, but they were not selected by SISSO as important parameters. This may imply that the density of surface redox-active sites is less important than their chemical nature for the considered materials set. Overall, the identified key parameters suggest that crystallographic distortions combined with the electronic properties of the catalyst encoded in the chemical nature of *A* and *B* elements are key for CO oxidation on the perovskites. The insights provided by our AI analysis reconcile the various, sometimes contradictory conclusions of previous studies[8c, 8g, 8k, 8l] and they stress that linear correlations are often not enough to describe heterogeneous catalysis. Moreover, Eq. 3 highlights that parameters characterizing atom packing and crystallographic distortions such as V' are required, in addition to physicochemical parameters associated to *A* and *B*, to correctly describe CO oxidation activity across a wide range of ABO_3 perovskite compositions. The crucial impact of the mentioned distortions has been overlooked by previous works, but it was identified here via a systematic consideration of *A* and *B* elements, rigorous experiments, and AI.

In order to assess the generalizability of the descriptor identified by SISSO, we verified the predictions of T_{10} given by Eq. 3 for 19 additional materials that were not used for training the model: 3 ABO_3 perovskites ($SmMnO_3$, $NdCoO_3$, and $SmCoO_3$), 12 $AMn_{(1-x)}Cu_xO_3$ (*A* = La, Pr) perovskites, and 4 $AMn_{0.7}B'_{0.3}O_3$ (*A* = Pr, Nd; *B'* = Fe, Co Ni, Zn) perovskites. The $AMn_{(1-x)}B'_xO_3$ perovskites present *B* and *B'* elements at the *B* sites, where Mn is substituted by *B'* at different proportions x in the range [0.0,0.4]. While Mn^{+3} or Mn^{+4} cations can be stabilized in perovskites, the stabilization of certain oxidation states of *B'* species such as Cu^{+3} is challenging. This motivated us to consider the substituted Mn-based perovskites $AMn_{(1-x)}Cu_xO_3$. We note that no substituted perovskite was considered in the training set. Besides, the elements Cu, Ni, and Zn were not present in any material of the training set. These additional materials were synthesized by solution combustion synthesis[11b, 17] and they were subjected to the same characterization methods and catalytic experiment as those used for the 13 ABO_3 training set materials. Experimental details are provided in ESI. The values of the primary features for these 19 materials are provided in the dataset.xls supplementary file. In order to evaluate the model of Eq. 3 for $AMn_{(1-x)}B'_xO_3$ materials, we took EN_B and IE_B as the composition average of the elemental properties of Mn and *B'* elements following

$$EN_B = (1-x)EN_{\mathrm{Mn}} + xEN_{\mathrm{B}}, \text{ (Eq. 4) and}$$

$$IE_B = (1-x)IE_{\mathrm{Mn}} + xIE_{\mathrm{B}\prime}. \text{ (Eq. 5)}$$

This is a reasonable approximation given that the temperature at which the materials were calcined is relatively low and thus the *B* and *B'* species are expected to be randomly distributed across the material. The T_{10} for the 19 additional materials are shown in Figure 1 as circles. The 3 ABO_3, 12 $AMn_{(1-x)}Cu_xO_3$, and 4 $AMn_{0.7}B'_{0.3}O_3$ perovskites are displayed in red, blue, and green, respectively. The activity of these materials is not well captured by x_B^{surf} or EN_B (Figure 1A and 1B). Indeed, the T_{10} values for the substituted $AMn_{(1-x)}B'_xO_3$ materials do not follow the T_{10} vs. EN_B trend that was apparent from the analysis of ABO_3 materials (Figure 1B, blue and green circles). However, the SISSO descriptor of Eq. 3 does capture the T_{10} values of the additional 19 perovskites with overall good accuracy (circles in Figure 1C). Interestingly, the high activity of $SmCoO_3$ and $NdCoO_3$ perovskites, which display T_{10} values lower than any material of the training set (105 and 113°C, respectively), is well captured by Eq. 3. The absolute prediction errors for $SmCoO_3$ and $NdCoO_3$ are 19 and 2 °C, respectively. These results show that the descriptor identified based on the 13 ABO_3 materials is valid beyond the materials space used for training and it can correctly describe highly active materials.

We analyzed the prediction errors (Table S2) to understand when the SISSO model might fail. The mean absolute prediction error evaluated on the 19 materials is 45 °C. However, the absolute prediction errors for $PrMn_{0.7}Zn_{0.3}O_3$ and $LaMn_{0.6}Cu_{0.4}O_3$ are 165 and 102 °C, respectively. These materials contain the elements Zn and Cu, which were not present in any material of the training set. Thus, the relatively high errors could be associated to the significantly different physical properties of these two elements compared to the elements constituting the materials of the training set. For instance, Zn displays the highest IP values among all elements considered in this work (Table S3). This results in a descriptor value that is higher than the descriptor values of the training set (Figure 1C). Additionally, the underlying processes governing the reactivity of these two materials could be different compared to processes dictating the activity in the training set due to the limited tendency of Cu and Zn to change the oxidation state and the resulting deviating redox properties. Indeed, Eq. 3 is expected to be valid for new materials as long as the reactivity of the examined compounds is governed by the same processes controlling the catalytic performance of the perovskites in the training set. In order to improve the domain of applicability of the SISSO model and cover larger portions of the materials space with high accuracy, the SISSO modelling can be combined with

systematic data acquisition in a sequential-learning setting. A SISSO-based sequential-learning approach will be discussed in an upcoming study.

Finally, we note that the parameters used as primary features in this work can be measured by relatively simple, widely accessible characterization techniques such as XRD or XPS. However, these parameters only reflect the thermodynamic properties of the materials in their precursor state (*ex situ*). Our target T_{10}, in turn, is obtained from a kinetic measurement. While our thermodynamic parameters are able to correctly describe T_{10} under fixed reaction conditions, e.g., feed composition, in this study, the inclusion of primary features that capture the interaction of the material with the reaction environment, e.g., by using *in situ* spectroscopy, might be crucial for the description of the reactivity in other systems or for systems that operate under significantly different environments, e.g., chemical potentials of the reaction mixture.[15a, 15b, 22]

Conclusion

We synthesized, characterized, and tested in CO oxidation 13 ABO_3 perovskites obtained by a systematic variation of *A* and *B* elements (*A* = La, Pr, Nd, Sm; *B* = Cr, Mn, Fe, Co) according to rigorous experimental procedures. The activity of these perovskites towards the catalytic CO oxidation cannot be easily correlated with the bulk or surface composition or with elemental properties of *A* or *B*, but it is well captured by AI. By applying a SISSO analysis, we identified a descriptor for the temperature necessary to achieve 10% conversion (T_{10}) as an analytical expression. This descriptor contains the key parameters normalized unit-cell volume determined by XRD, the Pauling electronegativity of the *A* and *B* elements, and the ionization energy of the *B* element. The descriptor is of more general validity than the quantities previously proposed as descriptors for activity. It can predict T_{10} of 3 additional ABO_3 materials and 16 substituted perovskites with the general formula $AMn_{(1-x)}(Cu, Fe, Co, Ni, Zn)_xO_3$ (*A* = La, Pr, Nd) with overall good accuracy. Crucially, this descriptor highlights the so far overlooked role of crystallographic strain, as the normalized unit cell volume is determined by the size of *A* and *B* cations as well as by the distortions of the octahedra within the perovskite structure. This work clearly shows that the relationships between the physical properties of perovskites and catalysis are nonlinear and they involve more than one parameter. The approach described in this paper accelerates the development of improved perovskite materials for catalysis by providing insights into the key materials' properties controlling catalysis and thus proposing approaches to the inorganic chemist or materials scientist that are not readily apparent.

Supporting Information

The Electronic Supporting Information (ESI) contains details on materials synthesis, characterization, and testing in the catalytic CO oxidation. Additionally, details on the SISSO analysis, as well as supplementary Tables S1, S2, and S3 and supplementary Figures S1 and S2 are provided in ESI. Additional references are cited within the ESI. [11b, 15a, 17-18, 20c, 21a, 23] The values of T_{10} and the values of the primary features for all materials discussed in this work are provided in the dataset.xls supplementary file.

Acknowledgements

L. F. and M. S. acknowledge the funding from the NOMAD Center of Excellence (European Union's Horizon 2020 research and innovation program, Grant Agreement No. 951786) and the ERC Advanced Grant TEC1p (European Research Council, Grant Agreement No 740233). We acknowledge Maike Hashagen for the N_2 physisorption experiments, Bernd Steinhauer, Jingyi Yang, and Rania Hanna for support in CO oxidation measurements.

Keywords: perovskite • CO oxidation • heterogeneous catalysis • artificial intelligence • symbolic regression

[1] a) V. M. Goldschmidt, *Naturwissenschaften* **1926**, *14*, 477-485; b) F. Fujishiro, C. Sasaoka, T. Ina, T. Sakuragi, M. Oishi, *J. Phys. Chem. C.* **2021**, *125*, 13283-13290; c) S. A. Oskoui, A. Niaei, H.-H. Tseng, D. Salari, B. Izadkhah, S. A. Hosseini, *ACS Comb. Sci.* **2013**, *15*, 609-621.

[2] a) J. Zhu, H. Li, L. Zhong, P. Xiao, X. Xu, X. Yang, Z. Zhao, J. Li, *ACS Catal.* **2014**, *4*, 2917-2940; b) J. Hwang, R. R. Rao, L. Giordano, Y. Katayama, Y. Yu, Y. Shao-Horn, *Science* **2017**, *358*, 751-756; c) M. A. Peña, J. L. G. Fierro, *Chem. Rev.* **2001**, *101*, 1981-2018.

[3] S. Royer, D. Duprez, *ChemCatChem* **2011**, *3*, 24-65.

[4] a) V. M. Goldschmidt, *Naturwissenschaften* **1926**, *14*, 477-485; b) C. J. Bartel, C. Sutton, B. R. Goldsmith, R. Ouyang, C. B. Musgrave, L. M. Ghiringhelli, M. Scheffler, *Sci. Adv.* **2019**, *5*, eaav0693.

[5] A. Bruix, J. T. Margraf, M. Andersen, K. Reuter, *Nat. Catal.* **2019**, *2*, 659-670.

[6] H.-J. Freund, G. Meijer, M. Scheffler, R. Schlögl, M. Wolf, *Angew. Chem. Int. Ed.* **2011**, *50*, 10064-10094.

[7] H. Tanaka, M. Misono, *Curr. Opin. Solid St.* **2001**, *5*, 381-387.

[8] a) B. C. Tofield, W. R. Scott, *J. Solid State Chem.* **1974**, *10*, 183-194; b) N. Mizuno, Y. Fujiwara, M. Misono, *Chem. Comm.* **1989**, 316-318; c) H. Yasuda, Y. Fujiwara, N. Mizuno, M. Misono, *Faraday Trans.* **1994**, *90*, 1183-1189; d) K. S. Chan, J. Ma, S. Jaenicke, G. K. Chuah, J. Y. Lee, *Appl. Cat. A.* **1994**, *107*, 201-227; e) J. Töpfer, J. B. Goodenough, *J. Solid State Chem.* **1997**, *130*, 117-128; f) K. Tabata, Y. Hirano, E. Suzuki, *Appl. Cat. A.* **1998**, *170*, 245-254; g) N. Taihei, I. Tatsumi, M. Makoto, *Bull. Chem. Soc. Jpn.* **1988**, *61*, 621-626; h) Y. Zhang-Steenwinkel, J. Beckers, A. Bliek, *Appl. Cat. A.* **2002**, *235*, 79-92; i) S. Miyoshi, J.-O. Hong, K. Yashiro, A. Kaimai, Y. Nigara, K. Kawamura, T. Kawada, J. Mizusaki, *Solid State Ion.* **2003**, *161*, 209-217; j)T. Kuznetsova, V. Sadykov, L. Batuev, L. Kurina, *React. Kinet. Catal. L.* **2005**, *86*, 257-265; k) M. Abdolrahmani, M. Parvari, M. Habibpoor, *Chin. J. Catal.* **2010**, *31*, 394-403; l)A. Tarjomannejad, A. Niaei, A. Farzi, D. Salari, P. R. Zonouz, *Catal. Lett.* **2016**, *146*, 1544-1551; m) P. R. Zonouz, A. Niaei, A. Tarjomannejad, *Int. J. Environ. Sci.* **2016**, *13*, 1665-1674; n) R. Mane, H. Kim, K. Han, H. Kim, S. S. Lee, H.-S. Roh, C. Lee, Y. Jeon, *Catal. Today* **2024**, *425*, 114347.

[9] a) E. Niwa, T. Sato, Y. Watanabe, Y. Toyota, Y. Hatakeyama, K. Judai, K. Shozugawa, M. Matsuo, T. Hashimoto, *J. Ceram. Soc. Jpn.* **2015**, *123*, 501-506; b) B. Levasseur, S. Kaliaguine, *J. Solid State Chem.* **2008**, *181*, 2953-2963.

[10] a) O. Parkash, P. Ganguly, G. R. Rao, C. N. R. Rao, D. S. Rajoria, V. G. Bhide, *Mater. Res. Bull.* **1974**, *9*, 1173-1176; b) M. Futai, C. Yonghua, Louhui, *React. Kinet. Catal. L.* **1986**, *31*, 47-54; cS. George, B. Viswanathan, *J. Colloid Interface Sci.* **1983**, *95*, 322-326.

[11] a) S. A. Hallweger, C. Kaußler, G. Kieslich, *Phys. Chem. Chem. Phys.* **2022**, *24*, 9196-9202; b) G. Koch, G. Bellini, F. Girgsdies, S. J. Carey, O. Timpe, T. Götsch, J. Kröhnert, G. Auffermann, M. Hashagen, R. Schlögl, A. Trunschke, *Chem. Mater.* **2024**, *36*, 5388-5404.

[12] a) Q. Tao, P. Xu, M. Li, W. Lu, *npj Comput. Mater.* **2021**, *7*, 23; b) B. Weng, Z. Song, R. Zhu, Q. Yan, Q. Sun, C. G. Grice, Y. Yan, W.-J. Yin, *Nat. Comm.* **2020**, *11*, 3513; c) T. Toyao, Z. Maeno, S. Takakusagi, T. Kamachi, I. Takigawa, K.-i. Shimizu, *ACS Catal.* **2020**, *10*, 2260-2297; d) S. Ma, Z.-P. Liu, *ACS Catal.* **2020**, *10*, 13213-13226; e) K. Takahashi, L. Takahashi, I. Miyazato, J. Fujima, Y. Tanaka, T. Uno, H. Satoh, K. Ohno, M. Nishida, K. Hirai, J. Ohyama, T. N. Nguyen, S. Nishimura, T. Taniike, *ChemCatChem* **2019**, *11*, 1146-1152; f) P. Schlexer Lamoureux, K. T. Winther, J. A. Garrido Torres, V. Streibel, M. Zhao, M. Bajdich, F. Abild-Pedersen, T. Bligaard, *ChemCatChem* **2019**, *11*, 3581-3601; g) R. Palkovits, S. Palkovits, *ACS Catal.* **2019**, *9*, 8383-8387; h) J. R. Kitchin, *Nat. Catal.* **2018**, *1*, 230-232; i) J. K. Nørskov, T. Bligaard, *Angew. Chem. Int. Ed.* **2013**, *52*, 776-777; j) L. Foppa, C. Sutton, L. M. Ghiringhelli, S. De, P. Löser, S. A. Schunk, A. Schäfer, M. Scheffler, *ACS Catal.* **2022**, *12*, 2223-2232; k) L. Foppa, L. M. Ghiringhelli, *Top. Catal.* **2022**, *65*, 196-206; l) J. Wang, H. Xie, Y. Wang, R. Ouyang, *J. Am. Chem. Soc.* **2023**, *145*, 11457-11465.

[13] Y. Liu, T. Zhao, W. Ju, S. Shi, *J. Materiomics* **2017**, *3*, 159-177.

[14] A. Trunschke, G. Bellini, M. Boniface, S. J. Carey, J. Dong, E. Erdem, L. Foppa, W. Frandsen, M. Geske, L. M. Ghiringhelli, F. Girgsdies, R. Hanna, M. Hashagen, M. Hävecker, G. Huff, A. Knop-Gericke, G. Koch, P. Kraus, J. Kröhnert, P. Kube, S. Lohr, T. Lunkenbein, L. Masliuk, R. Naumann d'Alnoncourt, T. Omojola, C. Pratsch, S. Richter, C. Rohner, F. Rosowski, F. Rüther, M. Scheffler, R. Schlögl, A. Tarasov, D. Teschner, O. Timpe, P. Trunschke, Y. Wang, S. Wrabetz, *Top. Catal.* **2020**, *63*, 1683-1699.

[15] a) L. Foppa, L. M. Ghiringhelli, F. Girgsdies, M. Hashagen, P. Kube, M. Hävecker, S. J. Carey, A. Tarasov, P. Kraus, F. Rosowski, R. Schlögl, A. Trunschke, M. Scheffler, *MRS Bull.* **2021**, *46*, 1016-1026; b) L. Foppa, F. Rüther, M. Geske, G. Koch, F. Girgsdies, P. Kube, S. J. Carey, M. Hävecker, O. Timpe, A. V. Tarasov, M. Scheffler, F. Rosowski, R. Schlögl, A. Trunschke, *J. Am. Chem. Soc.* **2023**, *145*, 3427-3442; c) R. Miyazaki, K. S. Belthle, H. Tüysüz, L. Foppa, M. Scheffler, *J. Am. Chem. Soc.* **2024**, *146*, 5433-5444.

[16] C. Sutton, M. Boley, L. M. Ghiringhelli, M. Rupp, J. Vreeken, M. Scheffler, *Nat. Comm.* **2020**, *11*, 4428.

[17] A. Varma, A. S. Mukasyan, A. S. Rogachev, K. V. Manukyan, *Chem. Rev.* **2016**, *116*, 14493-14586.

[18] A. Moshantaf, P. Oppermann, H. Junkes, https://gitlab.fhi.mpg.de/fhi-ac/ertl **2023**.

[19] C. P. Marshall, J. Schumann, A. Trunschke, *Angew. Chem. Int. Ed.* **2023**, *62*, e202302971.

[20] a) R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler, L. M. Ghiringhelli, *Phys. Rev. Mat.* **2018**, *2*, 083802; b) L. Foppa, T. A. R. Purcell, S. V. Levchenko, M. Scheffler, L. M. Ghringhelli, *Phys. Rev. Lett.* **2022**, *129*, 055301; c) T. A. R. Purcell, M. Scheffler, C. Carbogno, L. M. Ghiringhelli, *J. Open Source Soft.* **2022**, *7 (71)*, 3960.

[21] a) S. G. Lias, *"Ionization Energy Evaluation" in NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, National Institute of Standards and Technology, Gaithersburg MD, 20899; b) WebElements, httpss://www.webelements.com.

[22] G. Koch, M. Hävecker, P. Kube, A. Tarasov, R. Schlögl, A. Trunschke, *Front. Chem.* **2021**, *9*.

[23] a) G. Koch, M. Hävecker, D. Teschner, S. J. Carey, Y. Wang, P. Kube, W. Hetaba, T. Lunkenbein, G. Auffermann, O. Timpe, F. Rosowski, R. Schlögl, A. Trunschke, *ACS Catal.* **2020**, *10*, 7007-7020; b) M. P. Pechini, US Patent 3330697, 3331967; c) C.-C. Hwang, J.-S. Tsai, T.-H. Huang, C.-H. Peng, S.-Y. Chen, *J. Solid State Chem.* **2005**, *178*, 382-389; d) P. W. Cooper, S. R. Kurowski, *Introduction to the Technology of Explosives*, John Wiley & Sons, **1997**; e) J. J. Yeh, I. Lindau, *Atomic Data and Nuclear Data Tables* **1985**, *32*, 1-155; f) S. Tanuma, C. J. Powell, D. R. Penn, *Surface and interface analysis* **1993**, *20*, 77-89; g) M. C. Biesinger, B. P. Payne, A. P. Grosvenor, L. W. M. Lau, A. R. Gerson, R. S. C. Smart, *Appl. Surf. Sci.* **2011**, *257*, 2717-2730; h) M. F. Sunding, K. Hadidi, S. Diplas, O. M. Løvvik, T. E. Norby, A. E. Gunnæs, *J. Elec. Spec. Rel. Phen.* **2011**, *184*, 399-409; i) M. C. Biesinger, *Surface and Interface Analysis* **2017**, *49*, 1325-1334; j) H. Ogasawara, A. Kotani, R. Potze, G. A. Sawatzky, B. T. Thole, *Phys. Rev. B* **1991**, *44*, 5465; k*, Vol. 2023*.