Article

# Materials Genes of CO₂ Hydrogenation on Supported Cobalt Catalysts: An Artificial Intelligence Approach Integrating Theoretical and Experimental Data

Ray Miyazaki,* Kendra S Belthle, Harun Tüysüz, Lucas Foppa,* and Matthias Scheffler

Read Online

ACCESS | 📊 Metrics & More | 📰 Article Recommendations | 🆂🅸 Supporting Information

**ABSTRACT:** Designing materials for catalysis is challenging because the performance is governed by an intricate interplay of various multiscale phenomena, such as the chemical reactions on surfaces and the materials' restructuring during the catalytic process. In the case of supported catalysts, the role of the support material can be also crucial. Here, we address this intricacy challenge by a symbolic-regression artificial intelligence (AI) approach. We identify the key physicochemical parameters correlated with the measured performance, out of many offered candidate parameters characterizing the materials, reaction environment, and possibly relevant underlying phenomena. Importantly, these parameters are obtained by both experiments and ab initio simulations. The identified key parameters might be



called "materials genes", in analogy to genes in biology: they correlate with the property or function of interest, but the explicit physical relationship is not (necessarily) known. To demonstrate the approach, we investigate the CO₂ hydrogenation catalyzed by cobalt nanoparticles supported on silica. Crucially, the silica support is modified with the additive metals magnesium, calcium, titanium, aluminum, or zirconium, which results in six materials with significantly different performances. These systems mimic hydrothermal vents, which might have produced the first organic molecules on Earth. The key parameters correlated with the CH₃OH selectivity reflect the reducibility of cobalt species, the adsorption strength of reaction intermediates, and the chemical nature of the additive metal. By using an AI model trained on basic elemental properties of the additive metals (e.g., ionization potential) as physicochemical parameters, new additives are suggested. The predicted CH₃OH selectivity of cobalt catalysts supported on silica modified with vanadium and zinc is confirmed by new experiments.
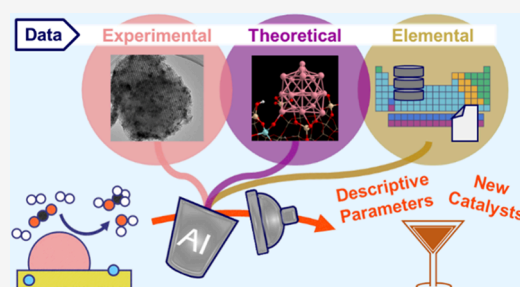
## 1. INTRODUCTION

Heterogeneous catalysis is one of the essential technologies in modern societies, since it has been utilized for decomposing toxic species, generating valuable chemicals, and for many more industrial applications.[1−6] Thus, there is a great demand for discovering and designing new catalytic materials that show higher performance than the available ones. Furthermore, catalysis is not only relevant to industrial applications but also links to fundamental science. For instance, organic molecules are generated by CO₂ hydrogenation at hydrothermal vents, which are fissures on the seafloor that discharge heated water.[10,11] This reaction is catalyzed by metal-containing materials. Because hydrothermal vents existed at the early Earth, they are considered as one of the candidate systems that produced the first organic molecules, eventually enabling the emergence of life.

Identifying key physicochemical parameters that describe the catalytic performance is a key step to design new catalytic materials and to understand the underlying phenomena. However, heterogeneous catalysis is governed by a complex and intricate interplay of several multiscale processes, such as transport of reactants, products, and heat in reactors, dynamical phase and structural transitions of catalytic materials during reactions, and chemical reactions on catalyst surfaces.[1−3,12] The time scales of those processes are also different. In particular, the catalytic material is often deposited on a support, for instance, metal nanoparticles are commonly supported on oxides. In these systems, the support material can also play a critical role in catalysis.[4,7−9] Thus, we need to consider a vast variety of physicochemical parameters that could describe the catalytic performance, and detailed atomic scale information is hard to be obtained from experimental studies, particularly under catalyst operating conditions. Additionally, there might be a higher complexity in real catalysis than what can be described by conventional theoretical modeling based on electronic structure calculations and statistical mechanics. In real catalysis, there might be not just one underlying process, but there is often a
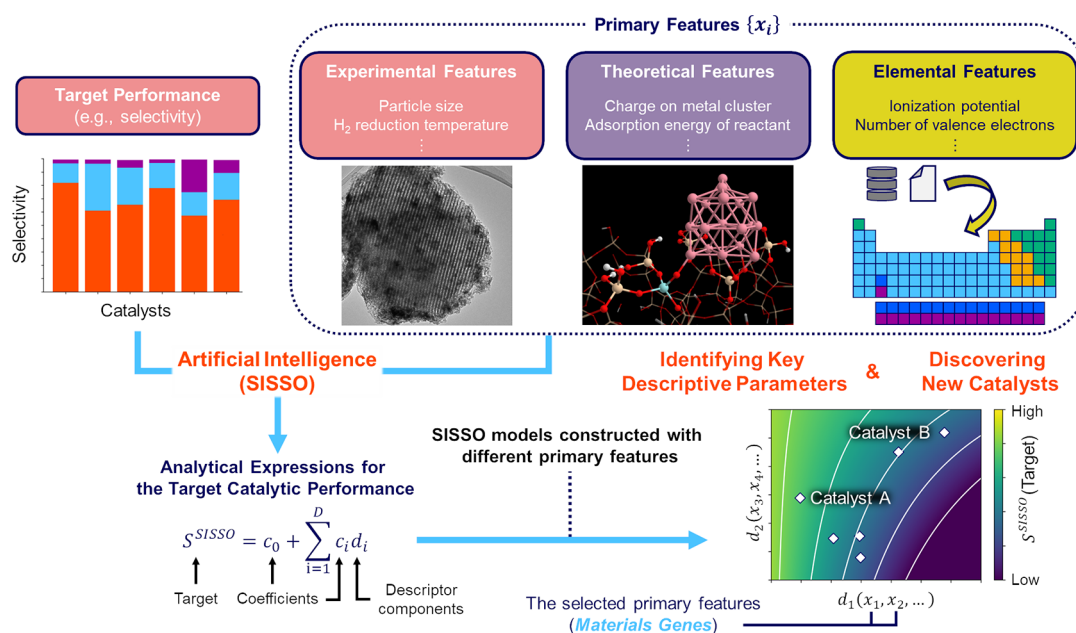
**Figure 1.** Schematic outline of the present study. By integrating information from different types of materials parameters (primary features) through the SISSO AI approach, we identify analytical expressions and the key descriptive parameters (materials genes) that are correlated to the measured (experimental) target catalytic performance. The SISSO models based on primary features with low acquisition costs also accelerate the discovery of new high-performance catalysts.

high intricacy of many underlying processes. Artificial intelligence (AI) may capture the catalytic progression better than previous theoretical/computational methods because it targets correlations and does not assume a single underlying physical model.

AI has been utilized in heterogeneous catalysis for discovering new catalytic materials and/or their design rules.[13−19] AI can access correlations describing the measured target catalytic performance without explicitly modeling all the underlying phenomena. In particular, the Sure-Independence Screening and Sparsifying Operator (SISSO)[20,21] has been adopted on data-centric approaches for heterogeneous catalysis.[22−25] Analytical expressions describing the target catalytic performance are identified by SISSO. The expressions contain few key parameters, out of many offered parameters that characterize the materials and might be correlated with the underlying processes triggering, favoring, or hindering the performance. Those key descriptive parameters composing the SISSO expressions have been called materials genes,[22] in analogy to genes in biology. Namely, the catalytic function of the material can be described by the combinations of the materials parameters analogously to how eye's color and health characteristics are determined by the combination of genes. Crucially, SISSO can identify potentially nonlinear, intricate correlations between (high-quality) small data sets (e.g., hundreds of target values), and an immense amount of candidate analytic functions (e.g., millions) is considered in the analysis. This makes SISSO suitable for applications in heterogeneous catalysis, where obtaining a large amount of consistent experimental data (i.e., data generated according to consistent and reproducible procedures) is usually time- and resource-consuming. On the other hand, either experimental or theoretical parameters were used in the previous SISSO studies on heterogeneous catalysis.[22−25] In the present study, we will combine these different types of data.

Here, we exploit experimental, theoretical, and elemental parameters (termed primary features) to efficiently model the catalytic performance and guide materials design (Figure 1). The experimental parameters consist of catalyst properties measured experimentally. The theoretical parameters consist of atomic-scale information obtained by density functional theory (DFT-RPBE) calculations with atomistic models, such as adsorption energies of intermediates and charge state of the catalysts. The elemental parameters are properties of the atoms composing the catalysts, such as ionization potential and atomic radius. These different types of parameters might capture different underlying processes, and they have different acquisition costs. In particular, the acquisition of elemental features does not require an experiment or a high-cost calculation with a large-scale atomistic model. Thus, the elemental parameters could provide AI models for predicting the performance of materials that were not yet synthesized or modeled with an atomistic simulation.

We focus on $CO_2$ hydrogenation catalyzed by cobalt nanoparticles supported on modified amorphous silica supports (Co/SBA-15 and Co/M-SBA-15 catalysts in Figure 2).[26] Those catalytic systems mimic the environment of hydrothermal vents, which are mainly formed by silica-rich mixtures of serpentinized peridotite and mafic materials. These materials could contain several metals, such as Al, Ca, and Mg. On the other hand, the reaction carried out in the gas phase, whereas the environment of the hydrothermal vent presents an aqueous environment. The adopted experimental setup corresponds to conditions of reduced water activity, which prevent hydrolysis,[27] and water is produced in our system via the reverse watergas shift reaction. Several organic molecules are formed by the $CO_2$ hydrogenation on the Co/M-SBA-15 catalysts (Figure S16 in the Supporting Information). We particularly attempt to elucidate the role of the modified amorphous silica support (M-SBA-15), in which different additive metals (e.g., M = Ti or Al) are introduced. In $CO_2$ hydrogenation catalysis, various roles of the support materials have been reported. For instance, the support can provide oxygen vacancies and isolated metal sites on its surfaces,
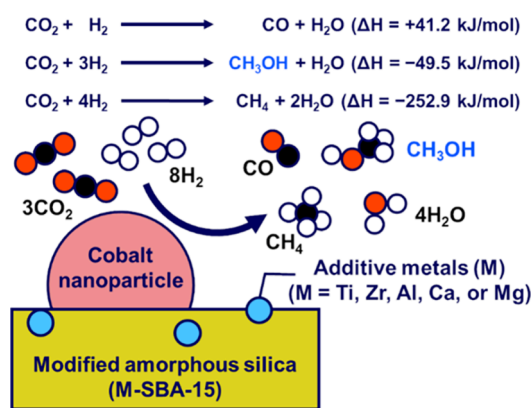
**Figure 2.** $CO_2$ hydrogenation by the Co/M-SBA-15 catalysts. In addition to the gas-phase molecules shown in the figure, liquid-phase organic molecules (e.g., formate) are formed by the reaction.[26] Formation enthalpies of the different reaction pathways ($\Delta H$ @ 298 K[35]) reflect the selectivity challenge in the $CO_2$ hydrogenation to methanol. The catalytic performance of the Co/SBA-15 and Co/M-SBA-15 catalysts is shown in Figure S16.

and it can modify the electronic structure and shape of the supported metal particles, or a mixing (alloying) between metal oxide supports and the supported metals might take place.[4,28,29] Indeed, in the studied systems, the selectivity depends on the type of incorporated additive metals (see Figure S16).[26] However, detailed, realistic microscopic modeling of catalytic processes on supported catalysts is rather challenging. Thus, design criteria for efficient support materials toward the desired catalytic performance have not been established yet. In the present study, we elucidate the most relevant properties of the modified silica supports and underlying catalytic mechanisms that lead to the formation of $CH_3OH$, $CH_4$, and CO by identifying the key descriptive parameters that correlate with the selectivity toward thses molecules. Furthermore, new additive metals that can improve the $CH_3OH$ selectivity are explored by using the SISSO model built only on elemental features (Figure 1). Those results can contribute to understand the key environments and the underlying mechanisms for the metabolic intermediate generation. Additionally, the obtained materials genes and candidate additive metals can be utilized to design catalysts for $CH_3OH$ production via $CO_2$ hydrogenation, which is a key catalytic reaction toward a sustainable society.[3,30−34]

## 2. DETAILS OF AI APPROACH

**2.1. SISSO Approach.** To identify the materials genes describing the $CO_2$ hydrogenation catalysis, we adopt the SISSO AI analysis[20,21] implemented in the SISSO++ code.[36] Physicochemical parameters potentially correlated to the target catalytic performance (primary features) are used as input. In this work, experimental characterization or first-principles calculations are used to obtain these primary features characterizing the catalyst materials and possible underlying processes. Then, mathematical operators, such as addition, division, and multiplication, are applied to the primary features for generating an immense number (up to millions) of analytic functions (descriptor candidates). SISSO then selects a few descriptor candidates and weighting coefficients by using the sure-independence screening (SIS) and the $l_0$ regularization. As a result, an analytical expression that shows the best correlation with the target performance ($S$) is obtained as a linear combination of the selected descriptors

$$S^{SISSO} = c_0 + \sum_{i=1}^{D} c_i d_i \tag{1}$$

Here, $c_0$ and $c_i$ are the weighting coefficients, and $d_i$ is the descriptor components. The primary features that appear in the expression of the descriptor components are the identified materials genes.

In this study, the multitask SISSO (MT-SISSO) transfer-learning approach[20] is adopted. This approach generates a different SISSO model for each of the predetermined tasks, but the descriptor components of these different models are the same. Just the weighting coefficients are fitted to the training data points of each task (i.e., the weighting coefficients are functions of the tasks).

**2.2. Primary Features and Target Performance.** We consider experimental, theoretical, and elemental primary features to model the catalytic performance in $CO_2$ hydrogenation. Experimental catalyst characterization primary features, simply denoted as the experimental primary features, are obtained from experimental characterization data of the Co/SBA-15 and Co/M-SBA-15 catalysts reported in our previous study.[26] These features are materials properties, such as the surface area per catalyst weight (specific surface area) measured by $N_2$ physisorption and the amount of desorbed $CO_2$ measured by TPD (temperature-programmed desorption). The total number of experimental primary features is 15, and they are listed in Table 1. Note that we also adopt measured $CO_2$ conversion ($X_{CO_2}^{RGSV}$) as an experimental primary feature to

**Table 1. Experimental Primary Features**[a]

| feature symbol | description | technique | unit |
|---|---|---|---|
| $X_{CO_2}^{RGSV}$ | $CO_2$ conversion | GC | % |
| $T_{H_2}$ | temperature of the first $H_2$ reduction signal | $H_2$-TPR | °C |
| $W_{Co}$ | cobalt loading | | wt % |
| $R_{M/Si}^{B}$ | bulk M/Si ratio | SEM-EDX | - |
| $R_{M/Si}^{S}$ | surface M/Si ratio | | - |
| $S_{surf}$ | specific surface area | | $m^2\ g^{-1}$ |
| $V_{pore}$ | total pore volume | $N_2$ physisorption | $cm^3\ g^{-1}$ |
| $r_{pore}$ | mean pore diameter | | Å |
| $TPD_{CO_2}^{mass}$ | amount of desorbed $CO_2$ per mass | | $\mu mol\ CO_2\ g^{-1}$ |
| $TPD_{CO_2}^{surf}$ | amount of desorbed $CO_2$ per surface area | $CO_2$-TPD | $\mu mol\ CO_2\ m^{-1}$ |
| $TPR_{H_2}^{mass}$ | amount of consumed $H_2$ per mass | | $\mu mol\ H_2\ g^{-1}$ |
| $TPR_{H_2}^{surf}$ | amount of consumed $H_2$ per surface area | $H_2$-TPR | $\mu mol\ H_2\ m^{-1}$ |
| $\nu_{C_5H_5N}$ | difference of wavelength of pyridine FT-IR[b] | pyridine FT-IR | $cm^{-1}$ |
| $d_{Co}$ | mean cobalt particle diameter | TEM | Å |

[a]Experimental primary features are based on the characterization schemes of ref 26. GC: gas chromatograph, TPR: temperature-programmed reduction, TPD: temperature-programmed desorption, SEM-EDX: scanning electron microscopy with energy dispersive X-ray analysis, FT-IR: Fourier transform infrared spectroscopy, TEM: transmission electron microscopy. [b]Difference of absorption wavelength of the via vibrational mode between pyridine in the gas phase and pyridine adsorbed on the catalysts.
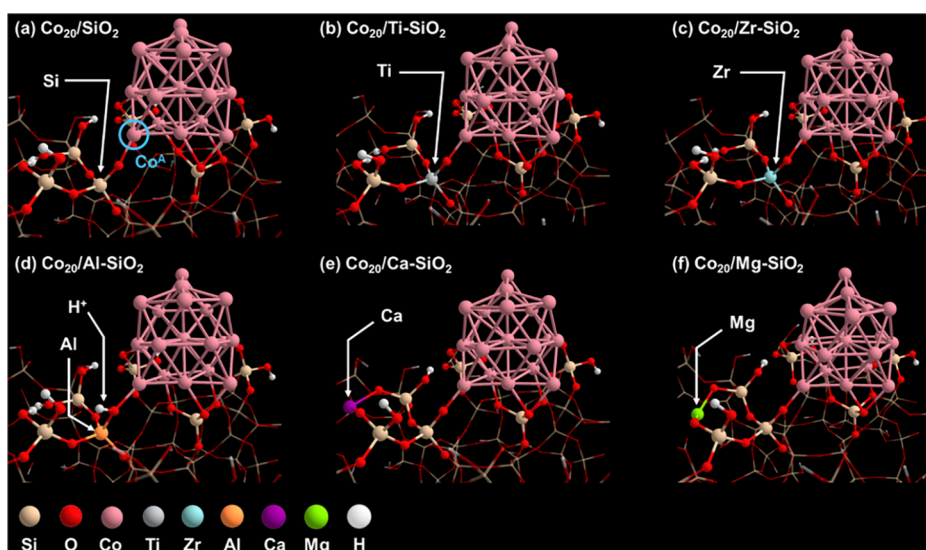
**Figure 3.** $Co_{20}/M\text{-}SiO_2$ models. The $Co_{20}$ cluster and $SiO_4$ that directly bond with $Co_{20}$ and/or M are highlighted as the ball and stick model. The other parts are shown as the wireframe model. The $Co_{20}/SiO_2$, $Co/Ti\text{-}SiO_2$, and $Co_{20}/Zr\text{-}SiO_2$ are reported previously in ref 26.

incorporate dependence of the $CH_3OH$ selectivity on the $CO_2$ conversion. More details are discussed later in this section.

Theoretical catalyst characterization primary features, simply denoted as the theoretical primary features, are obtained by electronic-structure calculations using atomistic models. We use $Co_{20}/M\text{-}SiO_2$ models,[26] where a $Co_{20}$ cluster[37] is supported on an amorphous silica surface slab,[38] as theoretical models of the $Co/SBA\text{-}15$ and $Co/M\text{-}SBA\text{-}15$ catalysts (Figure 3). The additive metal in the support (i.e., Ti, Zr, Al, Ca, or Mg) is incorporated in the silica surface. The coordination numbers of these additive metals in the models are chosen based on their stable formal oxidation number (+4: Ti and Zr, + 3: Al, and +2: Ca and Mg). As reported in our previous study,[26] a silicon atom in a $SiO_4$ unit that directly interacts with the cobalt cluster is replaced by Ti or Zr ($Co_{20}/Ti\text{-}SiO_2$ and $Co_{20}/Zr\text{-}SiO_2$ models in Figure 3b,c, respectively). In the $Co_{20}/Al\text{-}SiO_2$ model (Figure 3d), the same silicon atom is replaced by Al, and a proton is incorporated in the $AlO_4$ unit to keep charge of the system neutral with $Al^{3+}$. The incorporated site of the proton shown in Figure 3d (i.e., $-Al-O-Co-$) is the most stable one compared with the other three sites in the $AlO_4$ unit (i.e., $-Al-O-Si-$). Ca or Mg is incorporated in a bridge site between two $SiO_3-OH$ units by removing H atoms from OH groups in those units ($Co_{20}/Ca\text{-}SiO_2$ and $Co_{20}/Mg\text{-}SiO_2$ models in Figure 3e,f, respectively). We considered incorporating Ca or Mg into the bridge sites around the $SiO_4$ unit where silicon is substituted by the additive metal in the other $Co_{20}/M\text{-}SiO_2$ models, and the most stable one is adopted. Although Ca and Mg are not directly interacting with cobalt in the atomistic model, the calculated values of the theoretical primary features, such as adsorption energies (Figures S3−S8 in the Supporting Information), are different from those obtained with the $Co_{20}/SiO_2$ model, where no additive metal is included. By incorporating Ca or Mg, the local structure around the cobalt cluster is modified, and this can in turn change the materials properties, including the stability of adsorbed species, for example, the most stable adsorption sites. Spin-polarized DFT calculations with the RPBE exchange−correlation functional[39] were performed by using the FHI-aims code.[40] More computational details are given in section S2.1 in the Supporting Information.

The theoretical primary features are materials properties as well as quantities reflecting the interaction of the catalysts with reaction intermediates of $CO_2$ hydrogenation. For instance, the Hirshfeld charge of the additive metals in the silica support, the adsorption energies of the key species (e.g., $CO_2$ and oxygen atom), the formation energies of the proposed intermediates for the $CH_3OH$ formation (e.g., HCOO and $CH_3OH$), and the electronic and geometric structures of adsorbed $CO_2$ (e.g., O−C−O angle) are included as the theoretical primary features. Ten theoretical primary features are used in total. They are listed in Table 2. We note that reconstructions and dynamical processes on the catalysts during the reactions are not considered in the $Co_{20}/M\text{-}SiO_2$ models. However, we hope that these processes, which are missing in our static DFT-RPBE calculations, are captured by the SISSO analysis. The goal of this work is not a fully theoretical explanation of the experimental

**Table 2. Theoretical Primary Features[a]**

| feature symbol | description | technique | unit |
|---|---|---|---|
| $E_{ads}^{CO_2}$ | adsorption energy of $CO_2$ at the interfacial site | | eV |
| $\Delta_{C-O}$ | sum of C−O bond elongation of adsorbed $CO_2$[b] | | Å |
| $\Delta_{\angle O-C-O}$ | degree of O−C−O angle bending of adsorbed $CO_2$ | | degree |
| $q_{CO_2}$ | the Hirshfeld charge of adsorbed $CO_2$ | | e |
| $q_M$ | the Hirshfeld charge of the additive metal | | e |
| $E_{HCOO}$ | formation energy of HCOO at the interfacial site | DFT-RPBE | eV |
| $E_{COOH}$ | formation energy of COOH at the interfacial site | | eV |
| $E_{CO+O}$ | formation energy of CO + O at the interfacial site | | eV |
| $E_{CH_3O}$ | formation energy of $CH_3O$ at the interfacial site | | eV |
| $E_{ads}^{O}$ | adsorption energy of O atom at the interfacial site | | eV |

[a]Primary features obtained by $Co_{20}/M\text{-}SiO_2$ models. More details are given in section S2.1. [b]Summation of elongations of two C−O bonds of the adsorbed $CO_2$ compared to the bond lengths in the gas phase.

findings but rather the identification of theoretical and experimental parameters that correlate with the measured performance. Indeed, our approach also includes experimental parameters that could correlate with reconstructions. For instance, the temperature of the first $H_2$ reduction signal in TPR (temperature-programmed reduction) can capture the transitions between different oxide phases, which could affect the selectivity of the $CO_2$ hydrogenation.[30,41−43]

In addition to the experimental and theoretical primary features, we also employ elemental primary features that are atomic properties of the additive metals, such as ionization potential, number of valence electrons, or properties of atomic dimers reflecting the interaction of metal atoms with C, O, and H atoms. The eight elemental primary features are listed in Table 3.

**Table 3. Elemental Primary Features**

| feature symbol | description | technique | unit |
|---|---|---|---|
| $M_{cov}$ | covalent atomic radius | exp. data[44] | Å |
| $MC_{rad}$ | radius of $M^{1+}$ | DFT-PBE0[45] | Å |
| $E_{M-C}$ | formation energy of a M−C dimer | | eV |
| $E_{M-O}$ | formation energy of a M−O dimer | DFT-PBE0[a] | eV |
| $E_{M-H}$ | formation energy of a M−H dimer | | eV |
| IP | ionization potential | exp. data[46] | eV |
| PEN | Pauling electron negativity | exp. data[47] | - |
| $N_{VE}$ | number of valence electrons | - | - |

[a]More details are shown in section S2.2.

Our target quantity is the experimental $CH_3OH$ selectivity, $S_{CH_3OH}^{exp}$, for six catalysts (Co/SBA-15 and Co/M-SBA-15 catalysts with M = Ti, Zr, Al, Ca, or Mg) under four different reactant gas space velocities (RGSV).[26] These are velocities normalized with respect to the weight of the catalyst in the reactor. Within the MT-SISSO approach, each RGSV is treated as a different task, and six data points of $CH_3OH$ selectivity are included in each task. Thus, altogether we have 6 × 4 = 24 data points. Obviously, this is a very small number, and for most machine-learning approaches, this data situation would prevent a proper analysis. However, SISSO can identify potentially nonlinear, intricate correlations even in such challenging situation by offering an immensity of descriptor candidates. To avoid overfitting such small data set, the optimal complexity of the SISSO model with respect to its predictability is determined by cross-validation, as described in the next section.

Figure 4 shows $S_{CH_3OH}^{exp}$ and the experimental $CO_2$ conversion ($X_{CO_2}^{RGSV}$) values for the different materials and RGSVs. Because the selectivity depends on the conversion level, the methanol selectivity should be analyzed by considering the different $CO_2$ conversions achieved by the different catalysts. Thus, in the SISSO analysis, $CO_2$ conversion is offered as an experimental primary feature to incorporate the selectivity dependency with respect to conversion. At RGSV = 4000 ($cm^3 h^{−1} g_{cat}^{−1}$), where the selectivity and the $CO_2$ conversion are in a reasonable balance,[26] the most selective material is Co/Ti-SBA-15 with a selectivity value of 35.2% at the conversion of 1.0% (Figure 4). Co/Zr-SBA-15 shows the second highest $CH_3OH$ selectivity (27.8%), and the remaining materials have $S_{CH_3OH}^{exp}$= 14.6−20.1%. The conversions are in the range of 1.0−5.8%, and Co/Ti-SBA-15 and Co/SBA-15 show the lowest and highest conversion, respectively.
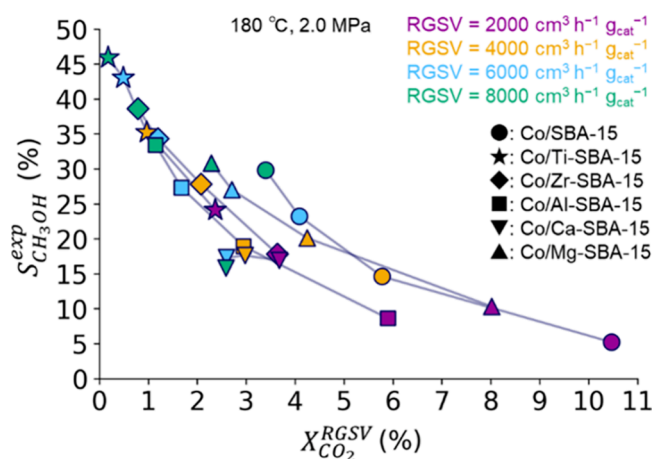


**Figure 4.** Experimental $CH_3OH$ selectivity ($S_{CH_3OH}^{exp}$) and $CO_2$ conversion ($X_{CO_2}^{RGSV}$) of the Co/SBA-15 and Co/M-SBA-15 catalysts at different RGSV.[26] The four colors label the different RGSV, and the six different shapes label the different catalysts. The lines connect RGSV values of the same catalysts. Reaction temperature and pressure are 180 °C and 2.0 MPa, respectively.

**2.3. Choice of SISSO Model Complexity.** The complexity of SISSO models is controlled by the number of times the mathematical operators are applied (rung: Q) to generate the descriptor candidates and by the number of descriptor components in the SISSO model (dimension: D). Thus, Q and D are hyperparameters and need to be chosen carefully. By increasing the Q and/or D, the complexity of the SISSO models also increases, and the training errors can be improved. However, the prediction errors do not necessarily decrease with increasing the complexity of the models.[48] Thus, the prediction errors commonly present a minimum at a certain Q and D.

In this study, the optimal Q and D with respect to predictability are determined by leave-one-material-out cross-validation (LOMO−CV).[22,23] This allows us to capture the trade-off between under- and overfitting of the models. In the LOMO−CV approach, data points related to one specific catalyst (e.g., Co/Ti-SBA-15) are removed from the training data. Then, the $CH_3OH$ selectivity of the left-out material is calculated (predicted) by the SISSO model trained on the remaining data, and root-mean-squared error (RMSE) of the prediction on the left-out material is obtained (Figure S1). This procedure is performed for all catalysts, and the average of the obtained RMSEs (CV-RMSE) is evaluated. The complexity (i.e., combination of Q and D) that shows the minimum CV-RMSE is selected as the optimal one.

We obtained models for the $CH_3OH$ selectivity considering all combinations with Q = 1,2 and D = 1,2,3. Additionally, we compared five different primary feature sets, containing (i) only the theoretical features (denoted "Theo"), (ii) only the experimental features (denoted "Exp"), (iii) the theoretical and experimental features (denoted "Theo + Exp"), (iv) the theoretical, experimental, and elemental features (denoted "Theo + Exp + Elem"), and (v) only the elemental features (denoted "Elem").

The number of candidate descriptors considered in our SISSO analysis is on the order of hundreds to millions (Table 4). For the models obtained with "Theo" and "Exp" primary feature sets, the optimal complexity identified by LOMO−CV is Q = 1

**Table 4. Number of Considered Primary Features and Generated Candidate Descriptors in the SISSO Analysis and the Optimal Complexity of the Models Identified by the LOMO−CV**

| primary feature set | number of primary features | number of candidate descriptors for $Q = 1$ | number of candidate descriptors for $Q = 2$ | Optimal $(Q, D)$ |
|---|---|---|---|---|
| Theo | 10 | 235 | 50,485 | (1, 1) |
| Exp | 15 | 304 | 63,110 | (1, 1) |
| Theo + Exp | 25 | 923 | 576,677 | (2, 1) |
| Theo + Exp + Elem | 33 | 1724 | 2,162,827 | (2, 1) |
| Elem | 8 | 154 | 24,652 | (1, 2) |

$D = 1$. Thus, in these cases, the optimal complexity is the lowest one considered. In contrast, the optimal $Q$ or $D$ is 2 for the remaining models (Table 4). These results indicate that only for the case of the last three primary feature sets a more complex model is also more predictive. Further details on the LOMO−CV are given in section S1 in the Supporting Information.

## 3. RESULTS AND DISCUSSION

### 3.1. Prediction Accuracy of SISSO Models with Different Primary Feature Sets.
Let us start with the analysis of prediction errors of SISSO models obtained for different primary feature sets at the optimal complexity. Figure 5 shows
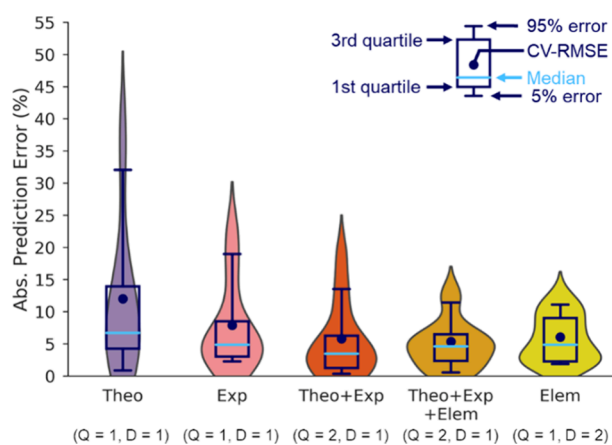


**Figure 5.** Box and violin plots of absolute prediction errors of LOMO−CV for SISSO models with different primary feature sets. The distribution of the prediction errors for the $CH_3OH$ selectivity shows how the combination of the different types of primary features improves the performance of the SISSO models. The upper and bottom whiskers of the box plots correspond to the 95th and 5th percentiles, respectively. The optimal complexity of each SISSO model (rung, dimension) is shown in parentheses.

the box and violin plots of the absolute prediction errors, which correspond to the absolute CV errors on left-out materials (i.e., $|\Delta M_i|$ in Figure S1). The width of the violin plots represents the density of data points with a certain error value. For the "Theo" and "Exp" models, which are built only with the theoretical or experimental primary features, respectively, the error distributions are broad (Figure 5). For example, their 95th percentiles of the prediction error distributions are 32.07 and 18.96%, respectively. Two catalysts have a particularly high prediction error for the "Theo" and "Exp" models. These are Co/Zr-SBA-15 and Co/SBA-15 (see Table S2). Thus, our offered primary features miss some of the relevant processes. When we offer

experimental and theoretical primary features ("Theo + Exp" models), the error distribution becomes narrower, and the 95% error of this model decreases to 13.53%. In particular, Co/Zr-SBA-15 is now described well (Table S2). The quartile values (length of the box) and CV-RMSE are also improved (decreased). The results indicate that both experimental and theoretical primary features and their combinations are important to model and describe the experimental $CH_3OH$ selectivity of the $CO_2$ hydrogenation. By further adding the elemental features along with experimental and theoretical features, a SISSO model with even lower prediction errors is obtained ("Theo + Exp + Elem" models in Figure 5). The 95% error of this model (11.43%) is now almost comparable with the standard deviation of the target calculated across the entire training set (10.48%). Note that the training error improves by increasing the number of the primary features because fitting gets better when more functions are offered. However, the prediction error may not improve because too many features increase the risk for causing overfitting.[48] Thus, the improvement of the prediction error in the "Theo + Exp" and "Theo + Exp + Elem" with respect to "Theo" or "Exp" models is not only due to the increment of the number of primary features but also due to the fact that more processes governing the $CH_3OH$ selectivity are synergistically captured by the combination among features.

### 3.2. Investigation of the Materials Genes.
Equation 2 is the SISSO model obtained with the theoretical, experimental, and elemental features (i.e., the "Theo + Exp + Elem" model) using the entire data set for training.

$$S_{CH_3OH}^{SISSO-Full} = c_0^{RGSV} + c_1^{RGSV} \left\{ \frac{M_{cov} \times X_{CO_2}^{RGSV}}{E_{CH_3O} \times T_{H_2}} \right\} \tag{2}$$

This model is based on a one-dimension descriptor. $M_{cov}$ is the covalent atomic radius[44] of the additive metals, and $X_{CO_2}^{RGSV}$ and $T_{H_2}$ are the experimental features, namely, the $CO_2$ conversion and the temperature of the first $H_2$ reduction signal. $E_{CH_3O}$ is the theoretical formation energy of $CH_3O$ intermediate at the interfacial site with respect to $CO_2$ and $H_2$ in the gas phase. The training RMSE of this model is 1.24%, and Figure 6 shows that the experimental $CH_3OH$ selectivity of each catalyst is well described.

The primary features that are selected by the SISSO model provide insights on the underlying processes. $T_{H_2}$ can be assigned to reduction temperature for $Co_3O_4$ to CoO. Thus, $T_{H_2}$ is related to the materials' reducibility and its phase transitions. Indeed, those reconstructions can be crucial in cobalt-based systems for the $CH_3OH$ formation from $CO_2$.[41] $X_{CO_2}^{RGSV}$ reflects the dependence of the selectivity on the conversion. $E_{CH_3O}$ corresponds to the stability of $CH_3O$, which is one of the key intermediates for the $CH_3OH$ formation as reported in previous studies about Cu catalysts supported on modified silica with highly dispersed hetero metal sites.[49−52] Because $CH_3O$ is recognized as one of the final intermediates for the $CH_3OH$ formation, the stability of $CH_3O$ affects the selectivity. For instance, lower stability makes the reaction pathway for the $CH_3OH$ formation unfavorable, but too strong interaction between $CH_3O$ and cobalt causes poisoning of the certain (active) sites. $M_{cov}$ corresponds to the size of the additive metals, and such difference of size might induce modification of the
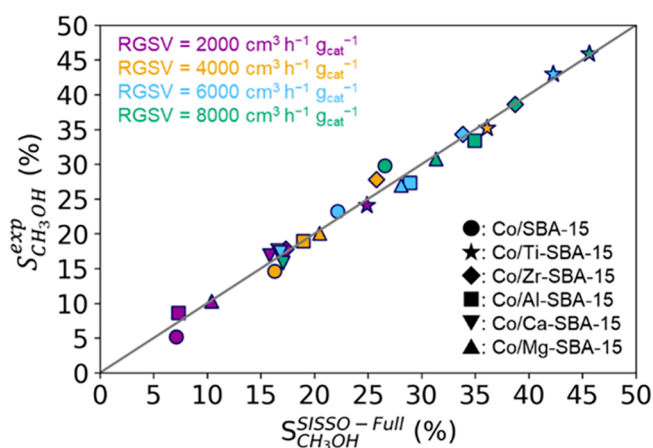
**Figure 6.** Comparison between the experimental selectivity results (*y*-axis) and the description provided by SISSO (*x*-axis). $g_{cat}$ is the weight of the catalyst. The gray line shows the ideal relationship: $S_{CH_3OH}^{exp} = S_{CH_3OH}^{SISSO\text{-}Full}$.



**Figure 7.** Catalyst map for the $CH_3OH$ selectivity focusing on $X_{CO_2}^{RGSV}$ as the *x*-axis. The *y*-axis corresponds to the remaining part of the descriptor ($d_{rem}^{Full}$). White squares are the data points of each catalyst. White lines correspond to contours (constant values) of $CH_3OH$ selectivity. In this figure, we focus on the values at RGSV = 4000 $cm^3$ $h^{-1}$ $g_{cat}^{-1}$. Note that the regions corresponding to $S_{CH_3OH}^{SISSO\text{-}Full}$ below 0% are colored with the same colors as for 0%.

(local) structures of M-SBA-15. The shape of the cobalt particles and/or active sites might be affected by such structure difference of M-SBA-15.

We note that the conversion has been related to the amount of metallic cobalt in ref 26. We found correlations between $X_{CO_2}^{RGSV}$ and the amount of consumed $H_2$ per catalyst mass during TPR measurement ($TPR_{H_2}^{mass}$). The Pearson correlation coefficients between these features are close to the unity (0.95, see Figure S9a). $TPR_{H_2}^{mass}$ represents amount of reduced cobalt oxides by $H_2$ and corresponds to amount of metallic cobalt (see Figure S9b). Namely, it is suggested that Co/SBA-15 has a large amount of the metallic cobalt compared to Co/Ti-SBA-15. These two features also correlate with amount of desorbed $CO_2$ per catalyst mass during TPD measurement ($TPD_{CO_2}^{mass}$), which reflects the amount of adsorbed $CO_2$ on the catalyst.[26] Thus, the metallic cobalt strongly binds $CO_2$, but apparently, it is not a good species for $CH_3OH$ formation. On the other hand, cobalt silicate ($Co-O-SiO_x$) at the interfacial region between cobalt and silica is hardly reducible species and could exist even after the reductive treatment in $H_2$ (Figure S9b). Thus, Co/Ti-SBA-15 that shows a small $TPR_{H_2}^{mass}$ value has relatively large amounts of cobalt silicate. Therefore, those results suggest that cobalt silicate is one of the key species for the high $CH_3OH$ selectivity. A similar experimental observation was also reported by Wang et al.[42] As observed above, the "Theo + Exp + Elem" model gives us the key materials properties that are the good starting points for further investigations toward the understanding of the catalytic mechanisms.

To investigate how the identified materials genes correlate with the $CH_3OH$ selectivity, we show a "catalyst map" in Figure 7. In order to analyze the effect of materials properties and $CO_2$ conversion separately, we chose $X_{CO_2}^{RGSV}$ as the *x*-axis, and the *y*-axis is the remaining part of the descriptor of eq 2, $d_{rem}^{Full} = \frac{M_{cov}}{E_{CH_3O} \times T_{H_2}}$. In this map, RGSV is set to 4000 $cm^3$ $h^{-1}$ $g_{cat}^{-1}$. The catalysts are distributed over the whole map. We also see that the $CH_3OH$ selectivity is inversely proportional to $X_{CO_2}^{RGSV}$: The color in Figure 7 changes from yellow/green to blue/violet from left to right. For example, the catalyst with the
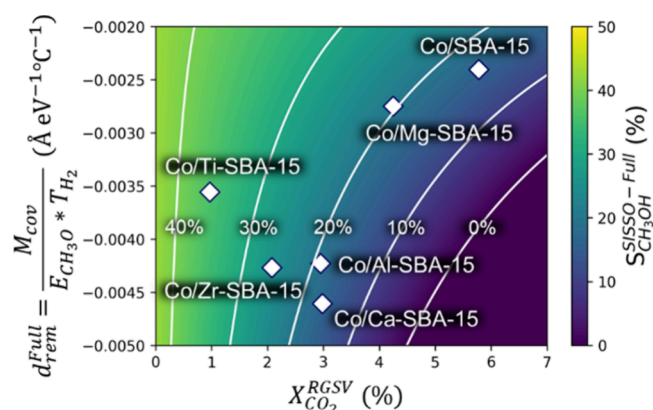
highest $CH_3OH$ selectivity (Co/Ti-SBA-15) shows the lowest $X_{CO_2}^{RGSV}$ and vice versa for Co/SBA-15 (Figures 4 and 7). On the other hand, an increase of $d_{rem}^{Full}$ (the *y*-axis) improves the $CH_3OH$ selectivity. This trend is particularly pronounced in the high $X_{CO_2}^{RGSV}$ region. Thus, not only the $CO_2$ conversion but also $d_{rem}^{Full}$ correlates with the $CH_3OH$ selectivity, and this result reflects the importance of combining different types of features and intricacy of describing the $CH_3OH$ selectivity.

The relevance of these parameters and the underlying processes that they capture would not be identified by a traditional theoretical analysis of the electronic structure via first-principles atomistic simulations or by an AI analysis with experimental data only (refs 22 and 23) independently, but they could be unveiled by the strategy introduced in this paper. In particular, by using theoretical and experimental data with SISSO, we allow that microscopic calculated parameters and experimental mesoscopic parameters are combined in a single analytical expression to model the catalytic performance. This enables capturing the interplay or coupling of processes occurring at different length scales.

### 3.3. Materials Genes for the CH₄ and CO Selectivity.
SISSO models for the measured $CH_4$ and CO selectivity of the Co/M-SBA-15 catalysts were also obtained. Equation 3 shows the expression of the "Theo + Exp + Elem" model for the $CH_4$ selectivity. The optimal complexity identified for this target is $Q = 2$, $D = 1$ (see Figure S13b), with training- and CV-RMSE of 0.90 and 3.56%, respectively.

$$S_{CH_4}^{SISSO\text{-}Full} = c_{32}^{RGSV} + c_{33}^{RGSV} \left\{ \sqrt[3]{X_{CO_2}^{RGSV}} \times \frac{V_{pore}}{M_{cov}} \right\} \tag{3}$$

$X_{CO_2}^{RGSV}$ and $M_{cov}$ are selected as the materials genes for the $CH_4$ selectivity. As discussed in the former section, those reflect the conversion dependence of the selectivity and the modification of the local structures of the catalysts, respectively. $V_{pore}$ is the pore volume of the catalysts, and it can be related to transport of the reactant and/or products. As shown in Figure S13c, the $CH_4$ selectivity increases as RGSV decreases. This

suggests that the transport of the reactant species and the contact time between the reactant gas and the catalysts both affect the $CH_4$ selectivity.

Equation 4 shows the expression of the "Theo + Exp + Elem" model for the CO selectivity. $Q = 2$, $D = 2$ is the optimal complexity for this model (see Figure S13e), with training- and CV-RMSE of 0.12 and 5.44%, respectively.

$$S_{CO}^{SISSO-Full} = c_{34}^{RGSV} + c_{35}^{RGSV}\left\{\left(q_M \times M_{cov}\right) \times \frac{W_{Co}}{E_{M-C}}\right\}$$
$$+ c_{36}^{RGSV}\left\{\left|\left(T_{H_2} \times R_{M/Si}^{B}\right) - \frac{T_{H_2}}{PEN}\right|\right\} \quad (4)$$

In eq 4, $q_M$ and PEN are the Hirshfeld charge and the Pauling electronegativity of the additive metals, respectively. $E_{M-C}$ is the formation energy of a M−C dimer and corresponds to the strength of the interaction between the additive metal and a carbon atom (see more details in section S2.2 in the Supporting Information). These three features can be related to the interaction between the additive metals and reaction intermediates such as CO. For instance, strength of the σ-donation and π-backdonation between CO and M can be correlated with those features. $R_{M/Si}^{B}$ is the bulk M/Si ratio, and it could be correlated with modification of the framework of SBA-15 and charge transfer between cobalt and SBA-15. $W_{Co}$ is the cobalt loading. As discussed above, our SISSO approach can be applied for the different catalytic performances, and the identified materials genes can capture the relevant underlying processes for each considered target. Further details on the analysis of $CH_4$ and CO selectivity are available in section S6 in the Supporting Information.

### 3.4. Exploring Other (New) Additive Metals by Using SISSO Models Trained with the Elemental Features.

As discussed in the previous sections, the model obtained using experimental, theoretical, and elemental features describes the experimental $CH_3OH$ selectivity with reasonable accuracy. It is useful to investigate the details of the catalytic mechanisms. However, the model is not very helpful for predicting the $CH_3OH$ selectivity of new catalysts because we must synthesize such materials to obtain the experimental features needed for the evaluation of eq 2. Conversely, by using the SISSO model only with the elemental features ("Elem" model), exploration of new additive metals that improve the $CH_3OH$ selectivity can be accelerated because these features are readily available for a wide range of elements in the periodic table. Although the distribution of the quartile values of the prediction errors is broader than that of the "Theo + Exp" and "Theo + Exp + Elem" models (Figure 5), CV-RMSEs of the "Elem" model and those two models are comparable (CV-RMSE = 5.78, 5.35, and 6.05% for the "Theo + Exp", "Theo + Exp + Elem", and "Elem" models). The obtained "Elem" model is shown as eq 5.

$$S_{CH_3OH}^{SISSO-Elem} = c_2^{RGSV} + c_3^{RGSV}\left\{\frac{M_{cov}}{E_{M-C}}\right\} + c_4^{RGSV}\left\{\frac{MC_{rad}}{E_{M-C}}\right\} \quad (5)$$

This model is based on a two-dimension descriptor. Similar to the "Theo + Exp + Elem" model, the covalent atomic radius[44] of the additive metals ($M_{cov}$) is also selected in the "Elem" model, suggesting that similar phenomena are being captured by this model. However, different materials genes are also identified by eq 5: $MC_{rad}$ is the radius of the additive metal cations ($M^{1+}$)[45]

and $E_{M-C}$. $MC_{rad}$ correlates with redox properties of the additive metals and might be related with the modification of SBA-15 structure explained in the previous section. $E_{M-C}$ could correlate with stabilities of intermediates adsorbed (interacted) with the additive metal sites as also discussed in the former section.

By using eq 5, the $CH_3OH$ selectivity of the cobalt catalysts with new additive metals is calculated. In this study, we focus on metals in the groups 2−15 up to Bi excluding lanthanoids. As shown in Figure 8, Co/Co-SBA-15 and Co/Ba-SBA-15 (yellow
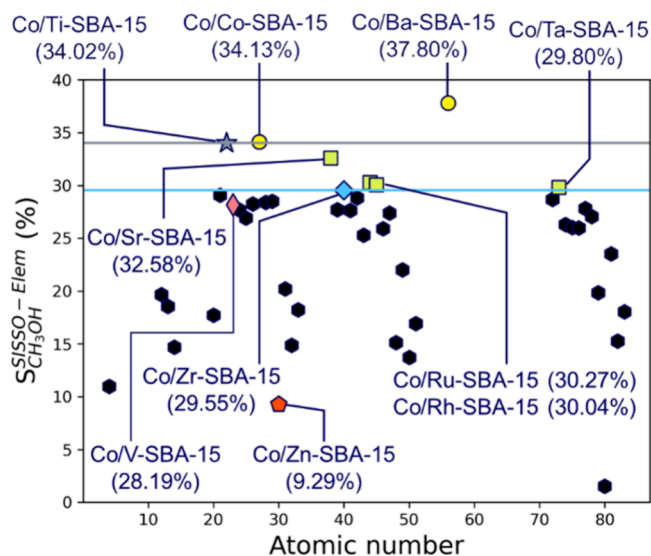


**Figure 8.** Predicted $CH_3OH$ selectivity by the "Elem" model. The x-axis represents atomic number of the additive metals. The gray and blue horizontal lines represent the $CH_3OH$ selectivity of Co/Ti-SBA-15 (34.02%) and Co/Zr-SBA-15 (29.55%) catalysts calculated by eq 5, respectively. In this figure, we focus on the $CH_3OH$ selectivity at RGSV = 4000 $cm^3$ $h^{-1}$ $g_{cat}^{-1}$.

circles) show higher $CH_3OH$ selectivity than that of Co/Ti-SBA-15 (gray star), which shows the highest $CH_3OH$ selectivity in the original experimental data set. Interestingly, cobalt itself is predicted as one of the best additive metals for improving the $CH_3OH$ selectivity. This result is in agreement with the discussion about the active sites for the $CH_3OH$ formation in the previous section, which is the cobalt species strongly interacted with silica (Co−O−SiO$_x$). Additionally, we also find that Co/Sr-, Ru-, Rh-, and Ta-SBA-15 (green squares in Figure 8) show higher $CH_3OH$ selectivity than that of Co/Zr-SBA-15 (blue diamond), which shows the second highest $CH_3OH$ selectivity in the training data. As observed above, we predict the candidate additive metals that can improve the $CH_3OH$ selectivity by using the SISSO model only with the elemental features. We also create a catalyst map based on the "Elem" model. As shown in Figure 9, Co/Sr- and Ba-SBA-15 are located far from the other candidate additive metals that show the high selectivity. Thus, those catalysts possibly show unique mechanisms, and for instance, similarities between Co/Sr-SBA-15 and Co/Ca-SBA-15 are suggested by the map. On the other hand, Co/Ru-, Rh-, Co-, and Ta-SBA-15 are close to Co/Ti- and Zr-SBA-15, which are the high-performance catalysts in the original training data set.

From all the elements considered in our screening, we focus on Co/V- and Zn-SBA-15 because synthesis procedures of those support materials are reported.[53−55] For these materials, we estimated the uncertainty of the prediction on the $CH_3OH$
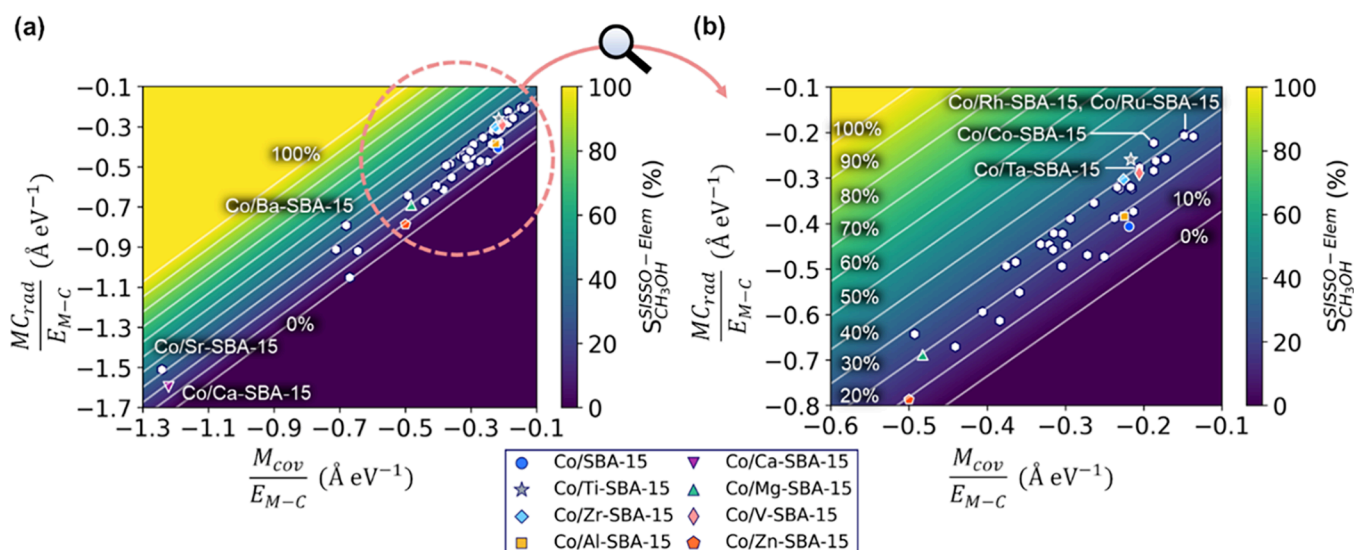
**Figure 9.** (a) Catalyst map for the CH$_3$OH selectivity based on the "Elem" model. (b) The map focusing on the upper-right region (surrounded by a pink circle in panel (a)). The x- and y-axis of the maps are the descriptor components of the "Elem" model. Note that the regions corresponding to $S_{CH_3OH}^{SISSO-Elem}$ above 100% and below 0% are colored with the same colors as for 100 and 0%.

selectivity by using SISSO models for LOMO−CV (hereafter CV models) that are trained without one materials in the original training data set (i.e., SISSO models w/o M$_{1-4}$ in Figure S1). In Figure S10, the CH$_3$OH selectivity of Co/V- and Zn-SBA-15 predicted by each of the CV models is shown. The distribution of the predicted values for Co/V-SBA-15 is narrower than that for Co/Zn-SBA-15 (standard deviation is 3.89 and 20.56%, respectively). More details are discussed in section S4 in the Supporting Information. We note that only six predicted values are considered in our uncertainty estimation. Ideally, we should estimate the uncertainty from a larger ensemble of models. However, we could only deal with the limited number of models via LOMO−CV due to the small size of the data set.

To investigate the prediction accuracy of the "Elem" model, we synthesized Co/V- and Zn-SBA-15 catalysts. We successfully incorporated V and Zn into SBA-15 by almost comparable M/Si ratio (∼0.1) with that of the other catalysts in the original training data set. Additionally, several experimental characterizations, such as N$_2$ physisorption and SEM-EDX, also show reasonable results, and more experimental details about the synthesis are shown in section S7 in the Supporting Information. Note that we also tried to synthesize the Co/Ba-SBA-15 catalyst based on a previous work.[56] However, we could only synthesize the Ba-SBA-15 support that has a small M/Si ratio (0.02) compared with the other catalysts. The CH$_3$OH selectivity of Co/V- and Zn-SBA-15 at each RGSV is shown in Figure S12c together with that of the other catalysts, and the selectivity at RGSV = 4000 cm$^3$ h$^{-1}$ g$_{cat}^{-1}$ is 32.5 and 16.9% for Co/V- and Zn-SBA-15, respectively. The Co/V-SBA-15 shows the second highest selectivity among all eight tested materials. The predicted CH$_3$OH selectivity by the "Elem" model is 28.19% for Co/V-SBA-15 and 9.29% for Co/Zn-SBA-15. Thus, the prediction errors of the "Elem" model for Co/V- and Zn-SBA-15 are 4.31 and 7.61%, respectively. The prediction error for Co/V-SBA-15 is within the range of the prediction accuracy of the "Elem" model estimated by LOMO−CV (Figure 5). However, the prediction error for the Co/Zn-SBA-15 is higher than the estimated accuracy. Those differences on the prediction errors between Co/V- and Zn-SBA-15 are indeed expected from the uncertainty estimated by the CV models. As discussed above and

in section S4, the uncertainty for Co/Zn-SBA-15 is larger. Those results suggest that, for the present data set, the uncertainty of the SISSO models can be estimated by using the CV models, and the prediction errors for new materials can be estimated by the computed uncertainty before synthesizing the materials.

By adding the new experimental CH$_3$OH selectivity and the elemental features of Co/V- and Zn-SBA-15 to the data set, we retrained and updated the "Elem" SISSO model. The updated "Elem" model (here after "ElemUp" model) is shown as eq 6.

$$S_{CH_3OH}^{SISSO-ElemUP} = c_5^{RGSV} + c_6^{RGSV}\{(E_{M-H})^6\} + c_7^{RGSV}\{\frac{PEN}{E_{M-C}}\}$$

(6)

The optimal complexity of the "ElemUp" model is also $Q = 1$, $D = 2$ (see Figure S11). $E_{M-H}$ is the interaction energy between a hydrogen atom and the additive metal. PEN is the Pauling electronegativity of the additive metals. The training- and CV-RMSEs are calculated to be 3.75 and 5.71%, respectively. As shown in Figure S12a, $E_{M-H}$ is strongly correlated with $E_{M-C}$, and PEN is negatively correlated with $M_{cov}$ and MC$_{rad}$. Thus, the materials genes newly selected in the "ElemUp" model and those in the "Elem" model possibly have similar roles on the descriptions for the CH$_3$OH selectivity. To investigate the effect by the difference of the selected materials genes between the "Elem" and "ElemUp" models, a catalyst map of the "ElemUp" model is also investigated (Figure S12b). Compared with the map of the "Elem" model (Figure 9), Co/Zn-SBA-15 moves to be close to Co/Ca-SBA-15, and those catalysts are located away from the other catalysts. Interestingly, the CH$_3$OH selectivity of those catalysts is almost independent of RGSV, and it is a unique trend in the given data set (Figure S12c). This can be an indication for a different reaction pathway compared to the other catalysts. The low number of training materials that display the RGSV dependency of the selectivity similar to that of the Co/Zn-SBA-15 material might contribute to the high prediction error on Co/Zn-SBA-15. On the other hand, in both maps, Co/V-SBA-15 is close to Co/Ti-SBA-15, and both catalysts show high CH$_3$OH selectivity. This analysis illustrates

how the description of the SISSO models can be updated by adding the new data points.

## 4. CONCLUSIONS

To identify the materials genes describing the $CO_2$ hydrogenation catalysis, we employed the SISSO AI approach with primary features from experiments and computations. We focused on the $CH_3OH$ selectivity of the Co/M-SBA-15 catalysts, with different additive metals "M" in the support material. The prediction accuracy estimated by LOMO−CV is studied for different primary feature sets, and we obtain a model that well represents the experimental $CH_3OH$ selectivity by combining experimental + theoretical + elemental primary features. The selected primary features reflect the reducibility of cobalt species, the adsorption strength of reaction intermediates, and the chemical nature of the additive metal as important factors for the selectivity. To accelerate the exploration of new additive metals that improve the $CH_3OH$ selectivity, we also built a SISSO model only with the elemental features, which have low acquisition costs. Based on the predicted values by this model, we find new candidate additive metals that should have high $CH_3OH$ selectivity. We also synthesized cobalt catalysts with new additive metals, V and Zn. The SISSO model was then updated by adding the new experimental data of the Co/V- and Zn-SBA-15 catalysts. The differences of the selected materials genes in the original and the updated models were investigated by using the catalyst maps. Those results by our SISSO approach can be utilized not only to understand the mechanisms of the organic molecule generation at the early Earth but also to design $CO_2$ conversion catalysts. We note that the generality of AI models is limited by the training data set. This also holds for SISSO though the interpretability of SISSO models and their physically meaningful primary features provide some hope for a good description. Importantly, by exploiting predicted metals and adding more data points, the generality of the SISSO models can be systematically improved.

## ASSOCIATED CONTENT

### ⓈⒾ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.3c12984.

> Details of the SISSO approach and the DFT-RPBE calculations, correlation between the amount of consumed $H_2$ and metallic cobalt, details of the "Elem" and "ElemUp" models, details of the SISSO models for the $CH_4$ and CO selectivity, and the experimental details (PDF)

## AUTHOR INFORMATION

### Corresponding Authors

**Ray Miyazaki** − *The NOMAD Laboratory at the Fritz-Haber-Institut der Max-Planck-Gesellschaft and IRIS-Adlershof of the Humboldt-Universität zu Berlin, Berlin 14195, Germany;* Present Address: Institute for Catalysis, Hokkaido University, Sapporo 001−0021 Hokkaido, Japan; ● orcid.org/0000-0001-7210-6646; Email: miyazaki@fhi-berlin.mpg.de

**Lucas Foppa** − *The NOMAD Laboratory at the Fritz-Haber-Institut der Max-Planck-Gesellschaft and IRIS-Adlershof of the Humboldt-Universität zu Berlin, Berlin 14195, Germany;* ● orcid.org/0000-0003-3002-062X; Email: foppa@fhi-berlin.mpg.de

### Authors

**Kendra S Belthle** − *Max-Planck-Institut für Kohlenforschung, Mülheim an der Ruhr 45470, Germany*

**Harun Tüysüz** − *Max-Planck-Institut für Kohlenforschung, Mülheim an der Ruhr 45470, Germany;* ● orcid.org/0000-0001-8552-7028

**Matthias Scheffler** − *The NOMAD Laboratory at the Fritz-Haber-Institut der Max-Planck-Gesellschaft and IRIS-Adlershof of the Humboldt-Universität zu Berlin, Berlin 14195, Germany*

Complete contact information is available at:
https://pubs.acs.org/10.1021/jacs.3c12984

## ■ REFERENCES

(1) Schlögl, R. Heterogeneous Catalysis. *Angew. Chem., Int. Ed.* **2015**, *54* (11), 3465−3520.

(2) Freund, H.-J.; Meijer, G.; Scheffler, M.; Schlögl, R.; Wolf, M. CO Oxidation as a Prototypical Reaction for Heterogeneous Processes. *Angew. Chem., Int. Ed.* **2011**, *50* (43), 10064−10094.

(3) Bowker, M.; DeBeer, S.; Dummer, N. F.; Hutchings, G. J.; Scheffler, M.; Schüth, F.; Taylor, S. H.; Tüysüz, H. Advancing Critical Chemical Processes for a Sustainable Future: Challenges for Industry and the Max Planck−Cardiff Centre on the Fundamentals of Heterogeneous Catalysis (FUNCAT). *Angew. Chem., Int. Ed.* **2022**, *61* (50), No. e202209016.

(4) Docherty, S. R.; Copéret, C. Deciphering Metal−Oxide and Metal−Metal Interplay via Surface Organometallic Chemistry: A Case Study with $CO_2$ Hydrogenation to Methanol. *J. Am. Chem. Soc.* **2021**, *143* (18), 6767−6780.

(5) Vogt, C.; Weckhuysen, B. M. The concept of active site in heterogeneous catalysis. *Nat. Rev. Chem.* **2022**, *6* (2), 89−111.

(6) Liu, L.; Corma, A. Metal Catalysts for Heterogeneous Catalysis: From Single Atoms to Nanoclusters and Nanoparticles. *Chem. Rev.* **2018**, *118* (10), 4981−5079.

(7) Campbell, C. T. Electronic perturbations. *Nat. Chem.* **2012**, *4* (8), 597−598.

(8) Pacchioni, G.; Freund, H.-J. Controlling the charge state of supported nanoparticles in catalysis: lessons from model systems. *Chem. Soc. Rev.* **2018**, *47* (22), 8474−8502.

(9) Sankar, M.; He, Q.; Engel, R. V.; Sainna, M. A.; Logsdail, A. J.; Roldan, A.; Willock, D. J.; Agarwal, N.; Kiely, C. J.; Hutchings, G. J. Role of the Support in Gold-Containing Nanoparticles as Heterogeneous Catalysts. *Chem. Rev.* **2020**, *120* (8), 3890−3938.

(10) De Graaf, R.; De Decker, Y.; Sojo, V.; Hudson, R. Quantifying Catalysis at the Origin of Life. *Chem.—Eur. J.* **2023**, *29* (53), No. e202301447.

(11) Preiner, M.; Xavier, J. C.; Sousa, F. L.; Zimorski, V.; Neubeck, A.; Lang, S. Q.; Greenwell, H. C.; Kleinermanns, K.; Tüysüz, H.;

McCollom, T. M.; Holm, N. G.; Martin, W. F. Serpentinization: Connecting Geochemistry, Ancient Metabolism and Industrial Hydrogenation. *Life* **2018**, *8* (4), 41.

(12) Urakawa, A. Mind the gaps in $CO_2$-to-methanol. *Nat. Catal.* **2021**, *4* (6), 447−448.

(13) Schlexer Lamoureux, P.; Winther, K. T.; Garrido Torres, J. A.; Streibel, V.; Zhao, M.; Bajdich, M.; Abild-Pedersen, F.; Bligaard, T. Machine Learning for Computational Heterogeneous Catalysis. *ChemCatChem.* **2019**, *11* (16), 3581−3601.

(14) Takahashi, K.; Takahashi, L.; Miyazato, I.; Fujima, J.; Tanaka, Y.; Uno, T.; Satoh, H.; Ohno, K.; Nishida, M.; Hirai, K.; Ohyama, J.; Nguyen, T. N.; Nishimura, S.; Taniike, T. The Rise of Catalyst Informatics: Towards Catalyst Genomics. *ChemCatChem.* **2019**, *11* (4), 1146−1152.

(15) Toyao, T.; Maeno, Z.; Takakusagi, S.; Kamachi, T.; Takigawa, I.; Shimizu, K.-i. Machine Learning for Catalysis Informatics: Recent Applications and Prospects. *ACS Catal.* **2020**, *10* (3), 2260−2297.

(16) Esterhuizen, J. A.; Goldsmith, B. R.; Linic, S. Interpretable machine learning for knowledge generation in heterogeneous catalysis. *Nat. Catal.* **2022**, *5* (3), 175−184.

(17) Mou, T.; Pillai, H. S.; Wang, S.; Wan, M.; Han, X.; Schweitzer, N. M.; Che, F.; Xin, H. Bridging the complexity gap in computational heterogeneous catalysis with machine learning. *Nat. Catal.* **2023**, *6* (2), 122−136.

(18) Foppa, L.; Sutton, C.; Ghiringhelli, L. M.; De, S.; Löser, P.; Schunk, S. A.; Schäfer, A.; Scheffler, M. Learning Design Rules for Selective Oxidation Catalysts from High-Throughput Experimentation and Artificial Intelligence. *ACS Catal.* **2022**, *12* (4), 2223−2232.

(19) Chanussot, L.; Das, A.; Goyal, S.; Lavril, T.; Shuaibi, M.; Riviere, M.; Tran, K.; Heras-Domingo, J.; Ho, C.; Hu, W.; Palizhati, A.; Sriram, A.; Wood, B.; Yoon, J.; Parikh, D.; Zitnick, C. L.; Ulissi, Z. Open Catalyst 2020 (OC20) Dataset and Community Challenges. *ACS Catal.* **2021**, *11* (10), 6059−6072.

(20) Ouyang, R.; Ahmetcik, E.; Carbogno, C.; Scheffler, M.; Ghiringhelli, L. M. Simultaneous learning of several materials properties from incomplete databases with multi-task SISSO. *J. Phys.: Mater.* **2019**, *2* (2), No. 024002.

(21) Ouyang, R. H.; 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* (8), No. 083802.

(22) Foppa, L.; Ghiringhelli, L. M.; Girgsdies, F.; Hashagen, M.; Kube, P.; Hävecker, M.; Carey, S. J.; Tarasov, A.; Kraus, P.; Rosowski, F.; Schlögl, R.; Trunschke, A.; Scheffler, M. Materials genes of heterogeneous catalysis from clean experiments and artificial intelligence. *MRS Bull.* **2021**, *46* (11), 1016−1026.

(23) Foppa, L.; Rüther, F.; Geske, M.; Koch, G.; Girgsdies, F.; Kube, P.; Carey, S. J.; Hävecker, M.; Timpe, O.; Tarasov, A. V.; Scheffler, M.; Rosowski, F.; Schlögl, R.; Trunschke, A. Data-Centric Heterogeneous Catalysis: Identifying Rules and Materials Genes of Alkane Selective Oxidation. *J. Am. Chem. Soc.* **2023**, *145* (6), 3427−3442.

(24) Khatamirad, M.; Fako, E.; Boscagli, C.; Müller, M.; Ebert, F.; d'Alnoncourt, R. N.; Schaefer, A.; Schunk, S. A.; Jevtovikj, I.; Rosowski, F.; De, S. A data-driven high-throughput workflow applied to promoted In-oxide catalysts for $CO_2$ hydrogenation to methanol. *Catal. Sci. Technol.* **2023**, *13*, 2656.

(25) Andersen, M.; Levchenko, S. V.; Scheffler, M.; Reuter, K. Beyond Scaling Relations for the Description of Catalytic Materials. *ACS Catal.* **2019**, *9* (4), 2752−2759.

(26) Belthle, K. S.; Beyazay, T.; Ochoa-Hernández, C.; Miyazaki, R.; Foppa, L.; Martin, W. F.; Tüysüz, H. Effects of Silica Modification (Mg, Al, Ca, Ti, and Zr) on Supported Cobalt Catalysts for $H_2$-Dependent $CO_2$ Reduction to Metabolic Intermediates. *J. Am. Chem. Soc.* **2022**, *144* (46), 21232−21243.

(27) do Nascimento Vieira, A.; Kleinermanns, K.; Martin, W. F.; Preiner, M. The ambivalent role of water at the origins of life. *FEBS Lett.* **2020**, *594* (17), 2717−2733.

(28) Jia, J.; Qian, C.; Dong, Y.; Li, Y. F.; Wang, H.; Ghoussoub, M.; Butler, K. T.; Walsh, A.; Ozin, G. A. Heterogeneous catalytic

hydrogenation of $CO_2$ by metal oxides: defect engineering − perfecting imperfection. *Chem. Soc. Rev.* **2017**, *46* (15), 4631−4644.

(29) Bergmann, A.; Roldan Cuenya, B. Operando Insights into Nanoparticle Transformations during Catalysis. *ACS Catal.* **2019**, *9* (11), 10020−10043.

(30) Puga, A. V. On the nature of active phases and sites in CO and $CO_2$ hydrogenation catalysts. *Catal. Sci. Technol.* **2018**, *8* (22), 5681−5707.

(31) Fan, T.; Liu, H.; Shao, S.; Gong, Y.; Li, G.; Tang, Z. Cobalt Catalysts Enable Selective Hydrogenation of $CO_2$ toward Diverse Products: Recent Progress and Perspective. *J. Phys. Chem. Lett.* **2021**, *12* (43), 10486−10496.

(32) Podrojková, N.; Sans, V.; Oriňak, A.; Oriňaková, R. Recent Developments in the Modelling of Heterogeneous Catalysts for $CO_2$ Conversion to Chemicals. *ChemCatChem.* **2020**, *12* (7), 1802−1825.

(33) Bowker, M. Methanol Synthesis from $CO_2$ Hydrogenation. *ChemCatChem.* **2019**, *11* (17), 4238−4246.

(34) Álvarez, A.; Bansode, A.; Urakawa, A.; Bavykina, A. V.; Wezendonk, T. A.; Makkee, M.; Gascon, J.; Kapteijn, F. Challenges in the Greener Production of Formates/Formic Acid, Methanol, and DME by Heterogeneously Catalyzed $CO_2$ Hydrogenation Processes. *Chem. Rev.* **2017**, *117* (14), 9804−9838.

(35) Roy, S.; Cherevotan, A.; Peter, S. C. Thermochemical $CO_2$ Hydrogenation to Single Carbon Products: Scientific and Technological Challenges. *ACS Energy Lett.* **2018**, *3* (8), 1938−1966.

(36) Purcell, T. A. R.; Scheffler, M.; Carbogno, C.; Ghiringhelli, L. M. SISSO++: A C++ Implementation of the Sure-Independence Screening and Sparsifying Operator Approach. *J. Open Source Softw.* **2022**, *7* (71), 3960.

(37) Farkaš, B.; de Leeuw, N. H. Towards a morphology of cobalt nanoparticles: size and strain effects. *Nanotechnology* **2020**, *31* (19), No. 195711.

(38) Comas-Vives, A. Amorphous $SiO_2$ surface models: energetics of the dehydroxylation process, strain, ab initio atomistic thermodynamics and IR spectroscopic signatures. *Phys. Chem. Chem. Phys.* **2016**, *18* (10), 7475−7482.

(39) Hammer, B.; Hansen, L. B.; Nørskov, J. K. Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Phys. Rev. B* **1999**, *59* (11), 7413−7421.

(40) Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. Ab initio molecular simulations with numeric atom-centered orbitals. *Comput. Phys. Commun.* **2009**, *180* (11), 2175−2196.

(41) Have, I. C. T.; Kromwijk, J. J. G.; Monai, M.; Ferri, D.; Sterk, E. B.; Meirer, F.; Weckhuysen, B. M. Uncovering the reaction mechanism behind CoO as active phase for $CO_2$ hydrogenation. *Nat. Commun.* **2022**, *13* (1), 324.

(42) Wang, L.; Guan, E.; Wang, Y.; Wang, L.; Gong, Z.; Cui, Y.; Meng, X.; Gates, B. C.; Xiao, F.-S. Silica accelerates the selective hydrogenation of $CO_2$ to methanol on cobalt catalysts. *Nat. Commun.* **2020**, *11* (1), 1033.

(43) Parastaev, A.; Muravev, V.; Osta, E. H.; Kimpel, T. F.; Simons, J. F. M.; van Hoof, A. J. F.; Uslamin, E.; Zhang, L.; Struijs, J. J. C.; Burueva, D. B.; Pokochueva, E. V.; Kovtunov, K. V.; Koptyug, I. V.; Villar-Garcia, I. J.; Escudero, C.; Altantzis, T.; Liu, P.; Béché, A.; Bals, S.; Kosinov, N.; Hensen, E. J. M. Breaking structure sensitivity in $CO_2$ hydrogenation by tuning metal−oxide interfaces in supported cobalt nanoparticles. *Nat. Catal.* **2022**, *5* (11), 1051−1060.

(44) Cordero, B.; Gómez, V.; Platero-Prats, A. E.; Revés, M.; Echeverría, J.; Cremades, E.; Barragán, F.; Alvarez, S. Covalent radii revisited. *Dalton Trans.* **2008**, *21*, 2832−2838.

(45) Rahm, M.; Hoffmann, R.; Ashcroft, N. W. Atomic and Ionic Radii of Elements 1−96. *Chem.—Eur. J.* **2016**, *22* (41), 14625−14632.

(46) Lias, S. G. Ionization Energy Evaluation In *NIST Chemistry WebBook*, NIST Standard Reference Database Number 69. National Institute of Standards and Technology: Gaithersburg MD, 20899.

(47) *WebElements*. https://www.webelements.com.

(48) Peleg, H. S.; Milo, A. Small Data Can Play a Big Role in Chemical Discovery. *Angew. Chem., Int. Ed.* **2023**, *62*, No. e202219070.

(49) Noh, G.; Lam, E.; Bregante, D. T.; Meyet, J.; Šot, P.; Flaherty, D. W.; Copéret, C. Lewis Acid Strength of Interfacial Metal Sites Drives $CH_3OH$ Selectivity and Formation Rates on Cu-Based $CO_2$ Hydrogenation Catalysts. *Angew. Chem., Int. Ed.* **2021**, *60* (17), 9650–9659.

(50) Noh, G.; Lam, E.; Alfke, J. L.; Larmier, K.; Searles, K.; Wolf, P.; Copéret, C. Selective Hydrogenation of $CO_2$ to $CH_3OH$ on Supported Cu Nanoparticles Promoted by Isolated $Ti^{IV}$ Surface Sites on $SiO_2$. *ChemSusChem* **2019**, *12* (5), 968–972.

(51) Noh, G.; Docherty, S. R.; Lam, E.; Huang, X.; Mance, D.; Alfke, J. L.; Copéret, C. $CO_2$ Hydrogenation to $CH_3OH$ on Supported Cu Nanoparticles: Nature and Role of Ti in Bulk Oxides vs Isolated Surface Sites. *J. Phys. Chem. C* **2019**, *123* (51), 31082–31093.

(52) Lam, E.; Larmier, K.; Wolf, P.; Tada, S.; Safonova, O. V.; Copéret, C. Isolated Zr Surface Sites on Silica Promote Hydrogenation of $CO_2$ to $CH_3OH$ in Supported Cu Catalysts. *J. Am. Chem. Soc.* **2018**, *140* (33), 10530–10535.

(53) Kunkel, B.; Kabelitz, A.; Buzanich, A. G.; Wohlrab, S. Increasing the Efficiency of Optimized V-SBA-15 Catalysts in the Selective Oxidation of Methane to Formaldehyde by Artificial Neural Network Modelling. *Catalysts* **2020**, *10* (12), 1411.

(54) Liu, M.; Gao, K.; Liang, L.; Sun, J.; Sheng, L.; Arai, M. Experimental and theoretical insights into binary Zn-SBA-15/KI catalysts for the selective coupling of $CO_2$ and epoxides into cyclic carbonates under mild conditions. *Catal. Sci. Technol.* **2016**, *6* (16), 6406–6416.

(55) Zhang, J.; Cheng, K.; Li, H.; Yin, F.; Wang, Q.; Cui, L.; Yang, S.; Nie, J.; Zhou, D.; Zhu, B. Efficient Synthesis of Structured Phospholipids Containing Short-Chain Fatty Acids over a Sulfonated Zn-SBA-15 Catalyst. *J. Agric. Food. Chem.* **2020**, *68* (44), 12444–12453.

(56) Yu, S.; Cao, X.; Li, L.; Liu, S.; Wu, Q. Catalytic cracking of rubber seed oil using basic mesoporous molecular sieves $K_2O$/MeO-SBA-15 (Me = Ca, Mg or Ba) as heterogeneous catalysts for the production of liquid hydrocarbon fuels. *Catal. Lett.* **2018**, *148* (12), 3787–3796.