Rules Describing CO_2 Activation on Single-Atom Alloys from DFT-meta-GGA Calculations and Artificial Intelligence

Herzain I. Rivera-Arrieta*[1] and Lucas Foppa*[1]
[1]*The NOMAD Laboratory at the Fritz Haber Institute of the Max Planck Society, Faradayweg 4-6, D-14195, Berlin, Germany*

Single-atom alloys (SAAs) arise as a promising concept for the design of improved CO_2 hydrogenation catalysts. However, from the immense number of possible SAA compositions and structures, only a few might display the properties required to be useful catalysts. Thus, the direct, high-throughput screening of materials is inefficient. Here, we use artificial intelligence to derive rules describing surface sites of SAAs that provide an effective CO_2 activation, a crucial initial step to convert the molecule into valuable products. We start by modeling the CO_2 interaction with 780 sites of flat and stepped surfaces of SAAs composed by Cu, Zn, and Pd hosts via high-quality DFT-mBEEF calculations. Then, we apply subgroup discovery to determine constraints on key physical parameters, out of 24 offered candidate descriptive parameters, characterizing subgroups (SGs) of surface sites where chemisorbed CO_2 displays large elongations of its C−O bonds. The key identified parameters are free-atom properties of the elements constituting the surface sites, such as their electron affinity, electronegativity, and radii of the *d*-orbitals. Additionally, the generalized coordination number is selected as a key geometrical parameter. The SG rules are used to identify promising alloys among more than 1,500 possible single-atom and dual-atom alloys. Some of the promising alloys predicted by the SG rules were explicitly evaluated by additional DFT-mBEEF calculations and confirmed to provide a significant CO_2 activation.

I. INTRODUCTION

Carbon dioxide is the main by-product of fossil fuels combustion and the largest contributor to global warming.[1–3] Among greenhouse gases, CO_2 is responsible for approximately 66% of Earth's total radiative forcing.[4] Thus, there is an urgent need for developing chemical processes to capture and transform CO_2 into valuable compounds, such as olefins and alcohols.[5–12] These processes can support a sustainable society if they are combined with the large-scale production of hydrogen from renewable energy sources.[13–15] Because the molecule is rather inert, the efficient conversion of CO_2 requires the use of catalysis. For instance, transition-metal catalysts can enable the hydrogenation of CO_2 to methanol via thermal processes.[16–20] Currently, $Cu/ZnO/Al_2O_3$ is the state-of-the-art catalyst for performing this reaction.[21] However, this catalyst experiences deactivation in the presence of high concentrations of water within the reaction mixture.[20–22] Hence, developing new, water-tolerant catalysts is crucial to achieve industrial CO_2 utilization.

Pd-based CO_2 hydrogenation catalysts are more water-tolerant than those based on Cu. Nevertheless, these systems favor the reverse water-gas shift reaction and produce significant amounts of carbon monoxide.[17,18] It has been shown that the selectivity to the desired methanol product can be increased by alloying Pd with a second metal, such as Zn or Ga.[20,23] Therefore, alloy materials offer a promising strategy for improving the performance of CO_2 hydrogenation catalysts.

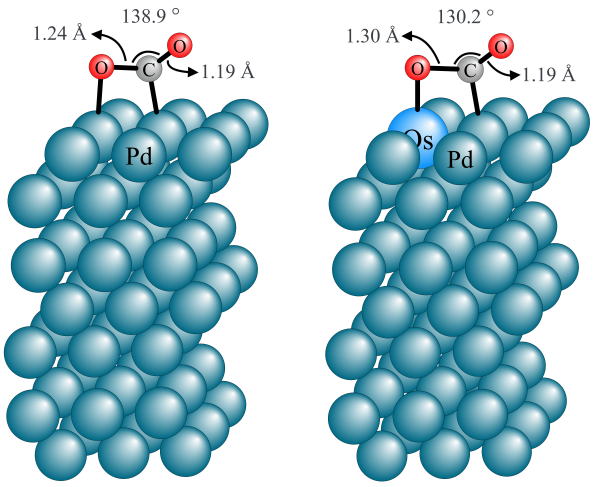
Single-atom catalysts are a frontline approach in catalysis research.[24,25] In these materials, isolated atoms of a metal are dispersed and stabilized on a host material. If the substrates are metals, we identify these materials as single-atom alloys (SAAs). Thus, SAAs are metallic systems in which one single atom of a given chemical element is embedded in a monometallic host surface of a second element.[26–30] We denote these systems hereafter as SA@host. SAAs can display unique electronic properties compared to monometallic systems or stoichiometric alloys.[31–34] These properties can be exploited to design new catalysts. Particularly, the adsorption strength of reactive species with the materials surface can be modulated by the choice of the SA and host elements, and by the geometry of the adsorption sites containing the SA. Several methods for the synthesis of SAAs and applications in catalysis have been reported.[35,36] SAAs are typically designed by embedding a highly active SA element on the surface of a less active host element. This is done to enable the desired reaction while avoiding undesirable side (competing) reactions that would occur in systems of pure, highly active metals. Thus, the selectivity can be improved. Additionally, it has been shown that SAAs can display enhanced stability during time on stream (operation).[27,37] However, due to the practically infinite compositional and structural space, designing SAAs for catalysis is a formidable challenge.

Theoretical approaches like those based on density functional theory (DFT),[38] can elucidate specific processes related to reactivity, such as surface reaction elementary steps.[39] For instance, it is possible to model the CO_2 activation, the key initial step for converting the molecule into valuable products.[40,41] Several DFT studies model the full reaction path for the CO_2 conversion, on SAAs, to different products like methanol and Ethylene.[42,43]

Artificial Intelligence (AI) can accelerate the discovery of promising materials by identifying correlations and patterns in data.[44] Nonetheless, only few SAAs might present the properties required for them to be useful in a specific application. This is often the case in catalysis, where only a handful of compounds are known as efficient catalysts (active, selective, and stable during operation). Global AI models might overlook these exceptional cases, since these models are designed to describe as many materials as possible, *i.e.*, to have the best predictive performance *on average* for most, but not necessarily the useful materials.[45,46] Alternatively, focused AI approaches can provide descriptions of specific regions of inter-

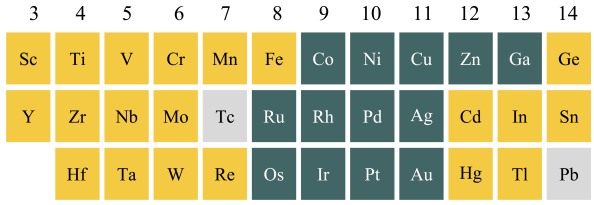
est in the data space (or materials space). For instance, subgroup discovery (SGD) identifies local partitions of the data associated with outstanding distributions of a given target of interest.[47–49] In particular, SGD provides rules as constraints (*e.g.*, inequalities) on the values of the key properties identifying the materials in the subset(s) of interest. Compared with widely used clustering algorithms, SGD has significant differences. Clustering is an unsupervised method that groups data points based solely on similarity, without considering a target variable or providing explicit explanations for the grouping. In contrast, rather than assigning each point to a cluster, SGD is a supervised approach that identifies specific subsets or subgroups (SGs) of interest as well as rules explaining why data points belong to these SGs.

In this work, we combine DFT simulations with the SGD approach to obtain rules describing surface sites on different SAAs able to activate CO_2 effectively (Fig. 1). Relying on the meta-Bayesian-error-estimation functional (mBEEF) for exchange and correlation, a semi-local meta-generalized gradient approximation (meta-GGA),[50] we model the CO_2 interaction with 780 surface sites in several flat and stepped surfaces of 36 SAAs based on Cu, Zn, and Pd host elements. We created a data set containing 24 physicochemical candidate descriptive parameters characterizing the surface sites where the molecule chemisorbs. As the target property for our AI analysis, we use the C−O bond elongation of these chemisorbed CO_2 structures. Then, by applying SGD, we uncover descriptions of surface-site SGs in the SAAs data set resulting in a large elongation of at least one of the C−O bonds. The obtained rules highlight the key electronic and geometric properties of the SAA surface sites associated with CO_2 activation. Based on the obtained SG rules, we efficiently identify promising alloys in a candidate space of more than 1,500 possible SAAs and dual-atom alloys (DAAs).[51,52] Through additional DFT-mBEEF calculations, we confirm the capability of the surface sites in these promising alloys to activate CO_2. Therefore, our approach provides chemical insights into the CO_2 activation on SAAs while enabling the efficient design of new materials.



(a) Pd(111) (b) Os@Pd(111)

FIG. 1. The single-atom-alloy (SAA) concept, *i.e.* embedding a single atom in a host metal, can significantly affect the CO_2 activation. (a) For the Pd(111) surface, the molecule chemisorption is not favorable ($E_{ads}^{CO_2}$ = 0.046 eV), and the C−O bond of chemisorbed CO_2 is slightly elongated with respect to the bond lengths in the isolated (gas-phase) molecule (1.155 Å). (b) In the Os@Pd(111) surface, CO_2 chemisorption is favored ($E_{ads}^{CO_2}$ = −0.337 eV) and the C−O bond is significantly elongated, indicating a stronger activation.



Host atom	Structure	Surfaces
Cu, Pd	FCC[a]	111, 110, 100, 211
Zn	HCP[b]	0001, 0001-step

[a] Face-centered-cubic structure
[b] Hexagonal-closed-packed structure

■ Pool of atoms used in the DFT−mBEEF simulations
■ Additional set of atoms used in the alloy screening

FIG. 2. Single-atoms, hosts, and considered surface terminations in our study of the CO_2 activation on SAAs. The structures containing Co and Ni atoms involved spin-polarized calculations.

II. METHODOLOGY

A. Atomistic Models of the SAA Surfaces and Calculations Settings

Based on previous work on the CO_2 hydrogenation by thermal[16–20] and electrochemical catalysis,[53–55] we choose Cu, Zn, and Pd as host metals in our SAAs. 12 different SAs are combined with each of the three hosts resulting in 36 SAAs. Overall, 120 flat and stepped SAA surfaces and 780 different surface sites were considered (Figure 2). The sites in the pristine surfaces of the three host metals were also included in the study.

The DFT simulations were performed with the FHI-aims package[56] and managed through the Atomic Simulation Environment.[57] We use the mBEEF functional[50] to perform our calculations. Prior benchmark work[58–60] confirms mBEEF as an appropriate choice to model the interaction between CO_2 and the SAA surfaces. The use of the Hartree potential correction of FHI-aims allowed us to speed up the simulations. A detailed description of the slab models, considered surface sites, and the DFT-mBEEF simulations, is provided in the Electronic Supplementary Information (ESI) Section 1.

We evaluate the CO_2 adsorption energy ($E_{ads}^{CO_2}$) to judge whether the interaction with the SAA surface is energetically favorable.

$$E_{ads}^{CO_2} = E_{slab\cdots CO_2} - (E_{slab} + E_{CO_2}). \quad (1)$$

https://doi.org/10.26434/chemrxiv-2024-1dr10 **ORCID:** https://orcid.org/0000-0001-7250-8100

Here, $E_{\mathrm{slab}\cdots\mathrm{CO_2}}$ is the energy of the slab with adsorbed CO_2, E_{slab} is the energy of the clean slab (without the molecule), and $E_{\mathrm{CO_2}}$ is the energy of the isolated CO_2 molecule.

As the physisorption of CO_2 is favorable on the surface of the three host metals, with $E_{\mathrm{ads}}^{\mathrm{CO_2}}$ values between -0.16 and -0.12 eV, we employ two criteria to select the structures for the SGD analysis. First, we only take into account chemisorbed CO_2 structures associated with $E_{\mathrm{ads}}^{\mathrm{CO_2}} < 0$ eV, since metastable configurations with $E_{\mathrm{ads}}^{\mathrm{CO_2}} > 0$ eV will be unlikely compared with the molecule's physisorption. Second, we select only the structures in which CO_2 directly interacts with the SA on the surface (see Fig. 1(b)), as the adsorption on SAA structures that are only composed by atoms of host element is similar to the adsorption on the sites of monometallic Cu, Zn, and Pd surfaces, which are included in our analysis (see Table S2 in the ESI and related discussion).

We consider the C−O bond elongation in the chemisorbed CO_2 molecule ($\Delta \mathrm{d}_{\mathrm{max}}^{\mathrm{C-O}}$) as an indicator of CO_2 activation.[61] This quantity is defined as:

$$\Delta \mathrm{d}_{\mathrm{max}}^{\mathrm{C-O}} = \mathrm{d}_{\mathrm{chem}}^{\mathrm{C-O}} - \mathrm{d}_{\mathrm{equil}}^{\mathrm{C-O}}, \tag{2}$$

were $\mathrm{d}_{\mathrm{chem}}^{\mathrm{C-O}}$ is the largest distance between the two C−O bond distances in the chemisorbed CO_2 molecule, and $\mathrm{d}_{\mathrm{equil}}^{\mathrm{C-O}}$ refers to the distance between C and O in an optimized gas-phase CO_2 molecule evaluated with DFT-mBEEF (1.155 Å).

B. Subgroup discovery

Starting with a data set of population $\widetilde{P}$, containing N physicochemical candidate descriptive parameters (φ_i) and a target quantity of interest Y, *e.g.* a material's property, SGD searches along the φ_i-space and identifies subsets of data or subgroups (SGs) with outstanding distributions of Y.[47,48] These SGs are identified by selectors (σ_i) or "rules" which can be used to obtain physical insights and to screen for new materials outside the training data set. The rules typically have the form:

$$\sigma_i = \pi_1 \ \wedge \ \pi_2 \ \wedge \ \ldots \tag{3}$$

were the π_i are propositions constraining the values of each φ_i to some minimum ($\pi_1 \equiv \varphi_1 > a$) or maximum ($\pi_2 \equiv \varphi_2 < b$) values to be determined. When a sample in the data set follows all the propositions in a given set of rules, the sample belongs to the SG. The identification of the SGs is based on the maximization of a quality function Q:

$$Q(SG,\widetilde{P}) = \frac{s(SG)}{s(\widetilde{P})} \cdot u(SG,\widetilde{P}). \tag{4}$$

The first term, known as the coverage, is the ratio between the size of the subgroup $s(SG)$ and the size of the data set $s(\widetilde{P})$. This term prevents the selection of too small SGs. The second term is the utility function $u(SG,\widetilde{P})$ and provides a metric for the usefulness of the SGs. The target in our analysis is $\Delta \mathrm{d}_{\mathrm{max}}^{\mathrm{C-O}}$ as defined in Eq. 2 and we aim at identifying rules describing structures with large $\Delta \mathrm{d}_{\mathrm{max}}^{\mathrm{C-O}}$ values, Thus, we used the normalized positive mean shift function. This utility function favors SGs with high mean values of $\Delta \mathrm{d}_{\mathrm{max}}^{\mathrm{C-O}}$. Hence, we should identify SGs of SAA surface sites that can significantly activate CO_2, and weaken at least one of its bonds.

TABLE I. Candidate descriptive parameters (φ_i) used in the subgroup-discovery analysis. Including a large number of properties increases the chances of AI to find which properties of the alloy determine an efficient CO_2 activation.

Type	Symbol	Unit	Description
Host	$\mathrm{PE_h}$	-	Host Pauling electronegativity
	$\mathrm{IP_h}$	eV	Host ionization potential
	$\mathrm{EA_h}$		Host electron affinity
	$\mathrm{r}_{s\text{-h}}$	Å	Host *s*-orbital atomic radius
	$\mathrm{r}_{p\text{-h}}$		Host *p*-orbital atomic radius
	$\mathrm{r}_{d\text{-h}}$		Host *d*-orbital atomic radius
	$\mathrm{r}_{val\text{-h}}$		Host valence radius
	$\mathrm{Bulk_{h\text{-}nnd}}$		Neighbor distance in host bulk
Single-atom	$\mathrm{PE_{SA}}$	-	SA Pauling electronegativity
	$\mathrm{IP_{SA}}$	eV	SA Ionization potential
	$\mathrm{EA_{SA}}$		SA electron affinity
	$\mathrm{r}_{s\text{-SA}}$	Å	SA *s*-orbital atomic radius
	$\mathrm{r}_{p\text{-SA}}$		SA *p*-orbital atomic radius
	$\mathrm{r}_{d\text{-SA}}$		SA *d*-orbital atomic radius
	$\mathrm{r}_{val\text{-SA}}$		SA valence radius
Surface site	$\mathrm{PE_{site}}$	-	Surface site PE[a]
	$\mathrm{IP_{site}}$	eV	Surface site IP[a]
	$\mathrm{EA_{site}}$		Surface site EA[a]
	$\mathrm{Site_{no}}$	# atoms	Atoms in the surface site
Surface site + first neighbors	$\mathrm{PE_{snn}}$	-	Surface site and first neighbors PE[a]
	$\mathrm{IP_{snn}}$	eV	Surface site and first neighbors IP[a]
	$\mathrm{EA_{snn}}$		Surface site and first neighbors EA[a]
	CN	# atoms	Surface site coordination number
	gen-CN		Generalized CN

[a] Average of all atoms in the ensemble.

As candidate descriptive parameters, we collected 24 physicochemical properties characterizing the SAAs and the surface sites where CO_2 chemisorbs (Table I). These parameters are constructed considering (free-atom) properties of the elements in the alloys, the monometallic bulk systems and of the SAA surface sites. The elemental properties are Pauling electronegativity (PE), ionization potential (IP), electron affinity (EA), and *s*-, *p*-, *d*-, and valence-orbital radii. We define four types of candidate descriptive parameters: host, SA, surface site, and surface site + first nearest neighbors. The host and SA parameters are defined as the properties of the elements associated with hosts and SAs. Additionally, the bulk interatomic distance is included within the host's candidate parameter. The surface site and surface site + first nearest neighbors parameters are defined as the average of the elemental properties corresponding to all atoms in the site (or ensemble). In other words, if the adsorption site is composed by more than one atom, the property average of all atoms in the ensemble is the value included in the data set. Two geometrical parameters characterizing the surface sites are included: coordination number (CN) and the generalized-CN (gen-CN).[62] By including basic candidate descriptive param-

eters, we aim to derive rules capable of screening new materials efficiently, without the need of further sophisticated DFT simulations. Additional details on the SGD AI approach, the utility function, and the candidate descriptive parameters evaluation are provided in the ESI Section 3.

III. RESULTS AND DISCUSSION

A. DFT-mBEEF Simulations and CO_2 interaction with SAAs

Overall, we performed approximately 2200 simulations modeling the interaction of the CO_2 molecule with SAAs based on the hosts Cu, Zn, and Pd. The monometallic surfaces of the hosts were also included in our analysis. Most of our simulations identified structures where CO_2 is physisorbed. The physisorption is characterized by the molecule's linear configuration. Besides, in the physisorbed systems, the molecule-surface distance is typically around 3 Å. Nevertheless, some simulations resulted in chemisorbed CO_2 structures. We found 199 structures where chemisorb CO_2 satisfies the criteria in Sec. II A. Three kinds of interactions are identified among the structures with chemisorbed CO_2: a) only the carbon atom of CO_2 bonds to the surface (η^1), b) the carbon and one oxygen atom of CO_2 bond to the surface (η^2) as in Fig. 1, and c) the three atoms of the molecule bond to the surface (η^3). The η^1 and η^3 geometries are illustrated in ESI Section 2.1. We analyze the correlation between Δd_{max}^{C-O} and $E_{ads}^{CO_2}$ considering the overall trends observed on these 199 systems. In particular, we focus on the host element, the surface structure, and the SA influences.

Fig. 3 (a) shows the adsorption energy and the C−O bond elongation values associated with the structures containing chemisorbed CO_2. The $E_{ads}^{CO_2}$ and Δd_{max}^{C-O} values are in the ranges [−1.2, 0 eV] and [0.04, 0.25 Å], respectively. These wide ranges show that the surface-adsorbate adsorption energy can be largely tuned by the choice of host and SA elements, and by the structure of the adsorption site. The SAAs displaying the strongest binding with the adsorbate, *i.e.*, the lowest $E_{ads}^{CO_2}$ values, are based on the Cu host. No clear pattern can be identified among the SAAs associated with large Δd_{max}^{C-O} values. Fig. 3 (a) also shows that low $E_{ads}^{CO_2}$ translates to large C−O bond elongations in some systems, implying that these quantities might be inversely correlated. However, the systems presenting the largest C−O bond elongations ($\Delta d_{max}^{C-O} > 0.18$ Å) display rather high CO_2 adsorption energy ($E_{ads}^{CO_2} > -0.30$ eV). Thus, large C−O bond elongations are not necessarily correlated with low CO_2 adsorption energies. The analysis of the charge transfer from the SAAs surfaces to chemisorbed CO_2 shows no clear correlation with the adsorption energy (see ESI Section 2.3).

We have also verified the effect of the surface structure on the CO_2 chemisorption. Here, we focus the discussion on surface terminations of the SAAs based on the Pd host (Fig. 3(b)). The Pd(110) and Pd(211) terminations tend to provide SAAs with stronger CO_2 chemisorption compared to Pd(111) and Pd(100). The Pd(110) and Pd(211) surfaces are also associated with some of the largest Δd_{max}^{C-O} values. These observations can be related to the fact that surface adsorption sites in (110) and (211) terminations display atoms with lower coordination (lower gen-CN) compared to the (111) and (100) terminations. Thus, these more unsaturated surface atoms can bind stronger to CO_2, activating the molecule. However, we note that some surface sites of (110) and (211) terminations are also associated with weak adsorption energies and small Δd_{max}^{C-O} values. Thus, these trends do not hold for all situations and the SA element also plays an important role.

Fig. 3(c) displays the influence of the SA on the $E_{ads}^{CO_2}$ and Δd_{max}^{C-O}. For this analysis, we focus on bridge sites of SAAs with the Pd host and corresponding to the surfaces (111), (100), (110), and (211). Firstly, we note that the presence of the SA typically favors the CO_2 chemisorption compared to the case of monometallic surfaces.[63] Moreover, we observe a decrease in the $E_{ads}^{CO_2}$ as the group of the SA increases within a given period in the periodic table. For instance, the average $E_{ads}^{CO_2}$ decrease between Ru and Rh bridge sites in Fig. 3(c) is 0.104 eV. In general, Δd_{max}^{C-O} seems to be mostly dictated by the SA element. Variations in the C−O elongation among the different surface terminations are small and do not show a clear pattern. The analysis for bridge sites of SAAs based on Cu and Zn hosts (see ESI Section 2.4) highlights similar trends to those for SAAs based on the Pd host.

Through the analysis of the data in Figs. 3(a), (b), and (c), it becomes clear that the host metal, the SA element, as well as the structure of the surface sites all impact CO_2 activation on the SAAs. Thus, it is challenging to establish simple (*e.g.* linear) correlations describing an effective CO_2 activation. In particular, the above discussion highlights that interesting scenarios might escape the overall (global) trends. To obtain such correlations, termed rules, we collected the 24 candidate descriptive parameters described in Table I for each of the 199 structures and used this information as input for the SGD AI analysis.

B. Identifying rules describing large C−O bond elongation in SAAs

The conversion of CO_2 into valuable products requires breaking at least one of its bonds. Hence, large values of Δd_{max}^{C-O} can reflect high reactivity of the SAA surface sites. For this reason, Δd_{max}^{C-O} is chosen as the target in our SGD study. The histogram showing the distribution of Δd_{max}^{C-O} in the data set of 199 structures (Fig. 3(d), in grey) highlights that the mean bond elongation is 0.103 Å. As only a few structures provide bond elongations larger than 0.15 Å, it is clear that the strong activation of CO_2 is an exceptional situation among the different SAA adsorption sites.

We used the SGD approach to identify descriptions of surface sites of SAAs presenting large values of the target Δd_{max}^{C-O}. For this purpose, we use the normalized positive-mean-shift utility function (see ESI Eq. S9). Table I shows the candidate descriptive parameters used in the SGD studies. After sectioning the φ_i-space, SGD analyzed half a million SGs and identi-

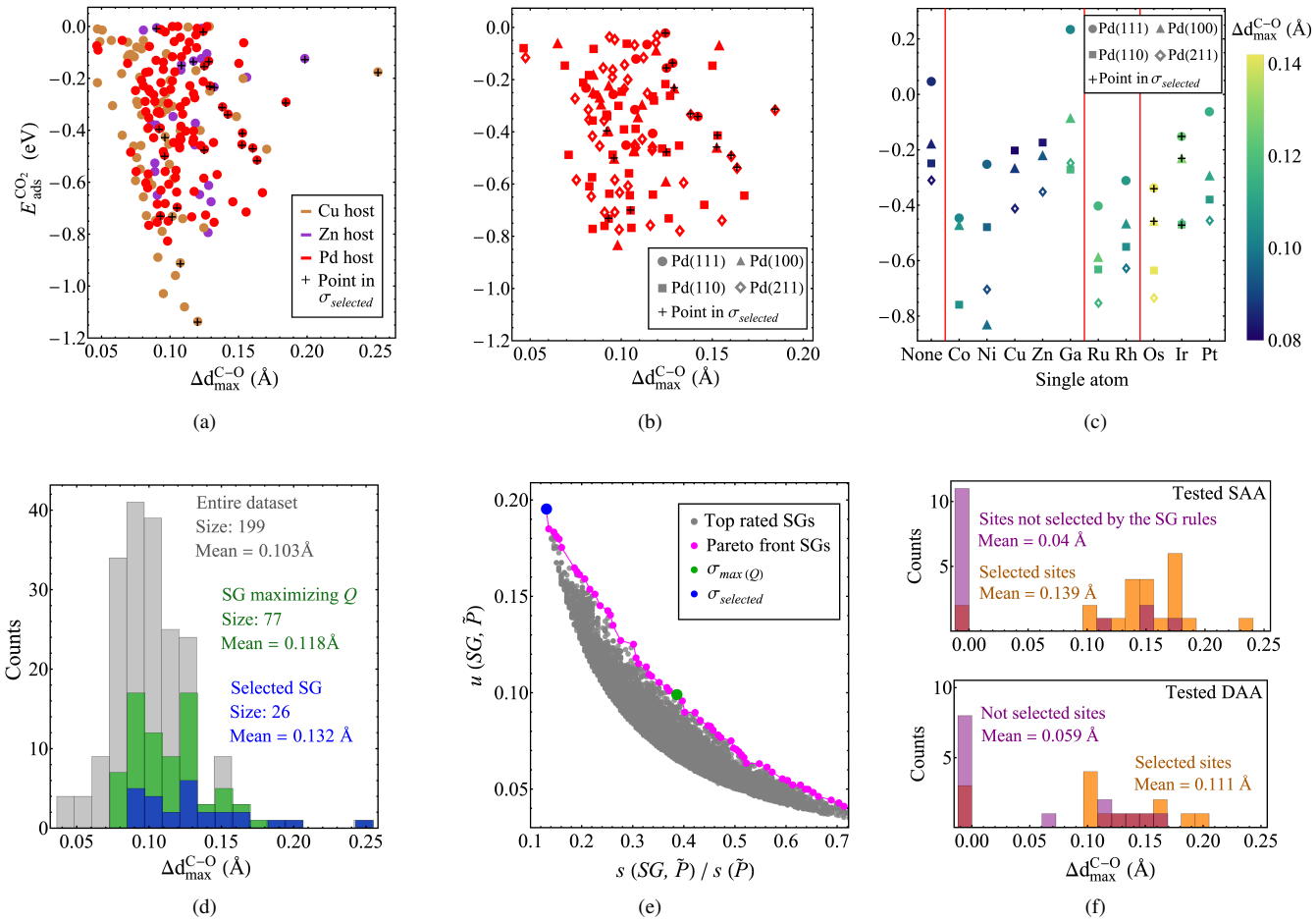


FIG. 3. Analysis of CO_2 chemisorption and activation on surface sites of single-atom alloys (SAAs) via DFT-mBEEF calculations and subgroup discovery (SGD) using 199 data points. (a): Relationship between adsorption energy ($E_{ads}^{CO_2}$) and largest C−O distance in chemisorbed CO_2 (Δd_{max}^{C-O}). (b): Surface-sensitivity of $E_{ads}^{CO_2}$ and Δd_{max}^{C-O} for SAAs based on the Pd host. (c): SA influence on $E_{ads}^{CO_2}$ and Δd_{max}^{C-O} for SAAs based on the Pd host. (d): Distribution of Δd_{max}^{C-O} in the entire data set and in the identified SGs $\sigma_{max(Q)}$ and $\sigma_{selected}$. (e): Top-rated SGD solutions with respect to the quality function of Eq. 4 and Pareto front of optimal solutions with respect to coverage (x axis) and normalized-positive-mean-shift utility function (y axis, see ESI Eq. S9). (f): Results of new DFT-mBEEF simulations used to validate the SG rules (see details of considered systems in Table III).

fies those with high $Q(SG,\widetilde{P})$ values. In Figure 3(e), we show the 15,000 top-rated SGD solutions with respect to the quality function (Eq. 4). These solutions are shown in a utility-function ($u(SG,\widetilde{P})$) vs. coverage plot. The SG maximizing $Q(SG,\widetilde{P})$, denoted $\sigma_{max(Q)}$, has a Δd_{max}^{C-O} mean of 0.118 Å, a coverage of 0.387, and is displayed in green in Fig. 4(d,e). We introduce the rules associated with $\sigma_{max(Q)}$ and the corresponding surface sites in the ESI Section 3.3. This SG contains surface sites associated with larger bond elongation compared to the entire data set. Nevertheless, $\sigma_{max(Q)}$ includes a significant fraction of systems with relatively small values of Δd_{max}^{C-O}.

In order to identify SGs focusing on the large Δd_{max}^{C-O}, *i.e.*, associated with more outstanding distributions of the target, we analyzed the Pareto front of SGD solutions with respect to the objectives coverage and utility function (magenta points in Fig. 3(e)).[64] In multi-objective optimization, a Pareto front is defined as the set of solutions for which no single objective can be improved without deteriorating at least one other objective. Thus, the solutions in the Pareto front reflect an optimal tradeoff between competing objectives. This analysis allows us to take into account multiple tradeoffs between the two conflicting objectives of SGD: coverage and utility. We focus on the SG with the highest utility value in the Pareto front. Denoted as $\sigma_{selected}$, this SG coverage is 0.131 and shows a Δd_{max}^{C-O} mean of 0.132 Å. The target distribution in $\sigma_{selected}$ is shown in blue in Fig. 3(d, e). This SG is more focused on the outstanding situation compared to the SG that maximizes the quality function. From the 24 offered candidate descriptive parameters, SGD identifies four parameters as key to describe the capability of the surface sites in SAAs to strongly elongate CO_2 bonds: the electron affinity of the adsorption site (EA_{site}), the Pauling electronegativity of the adsorption site (PE_{site}), the generalized coordination number

of the adsorption site (gen-CN), and the radius of the SA *d*-orbital ($r_{d\text{-SA}}$). The rules associated with the SG $\sigma_{selected}$ are the following:

$$\sigma_{selected}: \quad \mathrm{EA_{site}} \leq 1.28\,\mathrm{eV} \ \wedge \ \mathrm{PE_{site}} \leq 2.22 \ \wedge \ \text{gen-CN} \geq 5.87 \ \wedge \ r_{d\text{-SA}} > 0.647\,\text{Å} \quad (5)$$

$\mathrm{EA_{site}}$, $\mathrm{PE_{site}}$, and $r_{d\text{-SA}}$ highlight the importance of the electronic properties of the SA and host elements for achieving CO_2 activation. In particular, the rules establish that the SA should display $r_{d\text{-SA}} > 0.647$ Å. This value is larger than the hosts' *d*-orbital radii equal to 0.319 Å, 0.300 Å, and 0.581 Å, for Cu, Zn, and Pd, respectively. The trend for the *d*-orbital radius of the elements within a given period decreases as the group number increases. Thus, there is a connection between the uncovered rules and the patterns observed in Fig. 3(c), where the SAs with larger *d*-orbital radius are also the ones favoring larger Δd_{max}^{C-O} (Os > Ir ≈ Pt, and Ru > Rh). The gen-CN, in turn, denotes the importance of the geometric environment of the surface site as another key factor for CO_2 activation. The rule on gen-CN constrains the values of this parameter to a minimum threshold. This excludes adsorption sites that present unsaturated atoms, such as the ontop and the bridge2-step (see Fig. S1) sites of the (211) surfaces, and ontop and short bridge sites of the (110) surfaces. No explicit properties of the host are listed in the rules. However, $\mathrm{EA_{site}}$ and $\mathrm{PE_{site}}$ introduce a host dependence as these quantities are averages among all the atoms that are part of the adsorption sites (SA + host).

The 26 surface sites that are part of $\sigma_{selected}$ are presented in Table II and highlighted by crosses in Figure 3(a-c). The selected SG contains surface sites of SAAs associated with the three considered hosts. This means that SAAs based on Cu, Zn, and Pd hosts can activate CO_2 effectively. The geometries of the surface sites in $\sigma_{selected}$ for the Cu host are bridge, hollow, and long bridge and they are present in the (111), (100), and (110) surfaces. In the case of the Zn host, ontop and bridge surface sites of the (0001) and (0001)-step surfaces are part of $\sigma_{selected}$. Finally, for the Pd host, the sites in $\sigma_{selected}$ are bridge, ontop, long bridge, 4-fold hollow, and fcc-s. These sites are present in the Pd surfaces (111), (100), (110), and (211). Each of these sites is displayed in the ESI Figure S2. Only two SA chemical elements are contained in $\sigma_{selected}$, namely Os and Ir. Overall, the analysis of the systems in $\sigma_{selected}$ reveals that the electronic and geometric environment mostly provided by bridge sites with Ir and Os SAs are key for CO_2 activation.

Clearly, the rules uncovered by SGD provide physical insights by identifying the key parameters favoring the effective CO_2 activation on SAAs. Nonetheless, each of the constraints identified by SGD does not necessarily have a specific physical meaning, as only the combination of all constraints describes CO_2 activation. In addition to the physical insights, the SG rules provide an efficient way to search for new SAAs containing different metals. This is because the parameters entering the rules in Expression 5 can be evaluated for a large number of surface sites in alloys. In particular, it is desirable to identify SAAs that can activate CO_2 and contain more earth-abundant elements than those in the identified SG (Os and Ir). In the following section, we perform the screening of SAAs considering a broader number of candidate elements.

TABLE II. SAAs and surfaces belonging to the selected SG. From the 26 sites in the SG, SAAs based on Cu and Zn hosts have 5 each, the remaining 16 belong to the Pd host.

Host	Surface	Geometry of sites	SAs
Cu	111	Bridge	Os
	100	Bridge, hollow	
	110	Long bridge	
Zn	0001	Ontop,[a] bridge	Os, Ir
	0001-step	Bridge1-s, fcc-s	
Pd	111	Bridge	Os, Ir
	100	Ontop,[a] bridge	
	110	Long bridge, 4-fold hollow	
	211	Bridge1-s, bridge3-s, fcc-s	

[a] Chemisorbed CO_2 displays an η^3 configuration.

C. Exploiting the SG rules to identify candidate alloys

We used the SG rules to identify SAAs in a larger *candidate space* of SAAs compared to the training set. The chosen candidate space contains 20 SA chemical elements (yellow elements in Fig. 2). The surface terminations, and adsorption sites considered in this analysis are shown in Table III. We focused on bridge adsorption sites, as this site geometry is present in all of the surface terminations within $\sigma_{selected}$. More specifically, three bridge sites were taken into account: the long bridge in Cu(110), and the bridge sites of Zn(0001) and Pd(100). In total, 60 surface sites are taken into account in this screening.

From the 60 SAA surface sites considered in the screening, 23 fulfill the propositions in $\sigma_{selected}$. We performed new DFT-mBEEF calculations to confirm whether the SAAs selected by the SG rules indeed provide significant C−O bond elongation in chemisorbed CO_2. The computational settings for these geometry optimizations were similar to the ones described in Sec. II A. We evaluated the 23 systems selected by the SG rules. These results are displayed in the upper half of Figure 3(f). In this figure, the orange bins correspond to the distribution of Δd_{max}^{C-O} values for the 23 systems. 21 of the 23 tested adsorption sites activate CO_2. These systems contain the SA elements Y, Zr, Nb, Mo, Hf, Ta, W, and Re on the Cu(110) surface, the SA elements Y, Nb, Mo, Ta, W, and Re on the Zn(0001) surface, and the SA elements Y, Zr, Nb, Mo, Hf, Ta, and Re on the Pd(100) surface. The average Δd_{max}^{C-O} among these 23 sites is equal to 0.139 Å. The system displaying the largest elongation among the tested sites is Hf@Cu(110) ($\Delta d_{max}^{C-O} = 0.235$ Å). This large bond elongation is closer to the maximum value within the training data set, of 0.252 Å, which is associated to a long bridge in an Os@Cu(110) surface. Two sites selected by the SG rules do not activate CO_2 (Zr@Zn(0001) and Hf@Zn(0001)). Additionally, we also use DFT-mBEEF to evaluate CO_2 activation

TABLE III. Host surfaces and bridge sites considered in the screening of SA and DA alloys suitable for CO_2 activation. In the case of the SAAs, we used the 20 SAs highlighted in yellow in Fig. 2. For DAAs, we used the set of 33 atoms highlighted in Fig. 2 (green and yellow), and $r_{d\text{-SA}}$ in $\sigma_{selected}$ is taken as the average of the two substituted atoms. Applying the rules allows a reduction of 62% and 75% of the candidate surface sites in the considered SA and DA alloys, respectively. Moreover, additional DFT-mBEEF calculations confirmed the rules' capability to predict an effective CO_2 activation.

Materials	Host surfaces	Studied site	Candidate sites	Sites selected by the SG rules	Selected sites		Not selected sites	
					Tested	Confirmed	Tested	Confirmed
SAAs	Cu(110)	Long bridge	20	8	8	8	5	4
	Zn(0001)	Bridge	20	8	8	6	5	4
	Pd(100)	Bridge	20	7	7	7	5	3
DAAs	Cu(100)	Bridge	496	133	5	5	5	2
	Pd(111)	Bridge	496	125	5	4	5	5
	Pd(100)	Bridge	496	125	5	3	5	2

on 15 sites that were not selected by the SG rules. The distribution of Δd_{max}^{C-O} values corresponding to these 15 systems is shown in 3(f) as purple bins. With a mean bond elongation of 0.04 Å, 11 of 15 tested systems did not activate CO_2. The largest value of Δd_{max}^{C-O} among these 15 systems is 0.178 Å, lower than the corresponding value among the selected SAA surface sites (0.235 Å). Overall, $\sigma_{selected}$ provides good criteria to identify sites where CO_2 can be activated in SAAs.

In addition to SAAs, we have also considered dual-atom alloys (DAAs) in our screening. DAAs are composed of a host and two different chemical elements A and B in close proximity, denoted (A, B)@host.[51,52] DAAs have recently received attention since the synergy between the two atoms in these systems can be exploited to efficiently activate molecules, such as ethanol,[51] and CO_2.[52] The design of DAAs is more challenging compared to the design of SAAs, as the number of possible combinations of elements and surface sites increases substantially. Hence, the SG rules could accelerate the discovery of DAA capable of activating CO_2. For the DAA screening, we considered bridge sites of the following hosts and surface terminations Cu(100), Pd(111), and Pd(100). The 33 SA chemical elements highlighted in green and yellow in Figure 2 were included in this screening. The number of DA bridge sites that can be constructed based on these SA and host surfaces is 1,488. The parameter $r_{d\text{-SA}}$ in Expression 5 is defined based on one SA element. Thus, to apply the constraint on $r_{d\text{-SA}}$ to the screening of surface sites in DAAs, we extended the definition of $r_{d\text{-SA}}$ by considering the average of the d-orbital radii of the two elements A and B in the DAA systems. The remaining parameters in 5 can be evaluated for DAAs using the definitions discussed previously.

By applying the SG rules with such extended definition for $r_{d\text{-SA}}$, we identify 383 surface sites of DAAs likely to provide large C−O bond elongations. We selected 15 out of these 383 systems, corresponding to five DAA sites per surface, and performed new DFT-mBEEF calculations. These results are summarized in the lower half of 3(f). The orange bins show the distribution of Δd_{max}^{C-O} values corresponding to these 15 systems. 12 of the tested bridge sites are able to activate CO_2. These sites are, for each host surface, the following: (Cd, W), (Mo, Pd), (Sc, Hf), (Zr, Ag), (Y, Pd)@Cu(100); (Nb, Mo), (Re, Pt),(Rh, Hf),(Sc, W)@Pd(111); (Mo, Hf), (Ti, Ta),(Ta, Pt)@Pd(100). The average Δd_{max}^{C-O} among these 15 sites is equal to 0.111 Å. The systems displaying the largest C−O bond elongation among the tested sites are (Re, Pt)@Pd(111) and (Ta, Pt)@Pd(100), with Δd_{max}^{C-O} values of 0.190 and 0.187 Å, respectively. Three of the selected DAAs, (Zr, Ag)@Pd(111) and (Nb, Hg), (Y, Ag)@Pd(100), did not present CO_2 activation. Finally, we evaluated CO_2 activation on 15 of the DAA surface sites that were not selected by the SG rules. The purple bins of the lower panel in Fig. 3 (f) show the distribution of Δd_{max}^{C-O} among these 15 systems. 8 sites did not show any activation and the average Δd_{max}^{C-O} value across the 15 sites is 0.059 Å. Even though the SG rules missed some DAA systems providing relatively large Δd_{max}^{C-O} and were not as effective as they were for SAAs, they correctly identified the DAA systems presenting the larger Δd_{max}^{C-O} values among all tested DAAs. This result is remarkable given that the rules were only trained on the much simpler SAA systems. We stress that the SGD rules can be systematically improved by retraining with more data (*e.g.*, the data related to DAAs), providing rules that better describe an effective CO_2 activation on both SA and DA alloy systems.

In addition to the surface reactivity, the synthesizability, and stability of a catalyst during the operation are crucial design criteria in heterogeneous catalysis. Indeed, the migration of metal atoms and the segregation of different metal phases might occur in SAAs and DAAs.[65] As a proxy for the synthesizability and stability of the considered alloys, we evaluated the formation energy of the SAAs used for training SGD by using DFT-mBEEF calculations. The details are presented in Section 4 of the ESI. The formation energy reflects the thermodynamic stability of the SAA system compared to the pure-metal phases. Our results show that the formation energies strongly depend on the host and SA. SAAs with favorable as well as unfavorable formation energies are present in our data set. In general, large C−O bond elongation are associated with unfavorable formation energies, *i.e.*, the most unstable SAAs are also the most reactive towards CO_2 activation. Nonetheless, relatively stable alloys can also achieve moderate C−O bond elongations, making them promising candidates for further investigation.

An active catalyst for CO_2 hydrogenation likely displays a good capability of activating CO_2 and H_2. However, a selective catalyst should also favor the formation of the desired

products by interacting with hydrogen and other reaction intermediates in an appropriate manner. The use of SGD described here can be extended to other key steps and intermediates along CO_2 conversion. For instance, the stability of formate intermediate in the molecule's hydrogenation to methanol has been proposed to be crucial.[11,39] Such analysis will be the subject of a separate contribution.

Finally, we note that the reaction conditions were not taken into account in this work. Nonetheless, if the applied reaction conditions (*e.g.*, temperature, pressure) are harsh, the surface of the catalyst might restructure. Such restructuring is typically unknown. Previously, the blending of high-quality theoretical and experimental data has shown that AI can identify correlations that take into account the experimental reaction conditions.[66,67] These approaches enabled the design of new materials confirmed experimentally. Thus, the incorporation of experimental data into SGD studies describing CO_2 conversion on SA and DA alloys is a promising route for addressing the impact of reaction conditions on the catalytic performance. This will require, however, systematic and rigorous experimental procedures[68] for the characterization and the evaluation of stability and reactivity of series of SAAs and DAAs.

IV. CONCLUSIONS

In this work, we used high-quality data generated by DFT-mBEEF calculations and applied the SGD AI approach to study the CO_2 activation in 36 SAAs based on Cu, Zn, and Pd hosts. From 24 easily accessible candidate descriptive parameters, SGD selected four parameters as key to characterize a SG of SAA surface sites capable of strongly elongating the C−O bonds of the chemisorbed molecule. These key properties highlight the importance of the SA nature (radius of its *d*-orbitals) and the surface site's electronic (the electron affinity and Pauling electronegativity of the surface site) and geometric factors in achieving an effective CO_2 activation. The rules provided by SGD connect these key parameters to the desired bond elongation and can be used for materials design. Indeed, from a total of 1,548 candidate systems, the rules enabled a fast screening and prediction of SA and DA alloys of new potential members of the outstanding SG. Through new DFT-mBEEF simulations, we tested and confirmed the rules' capability to correctly predict the strong CO_2 activation on SA and DA alloys beyond the chemical space used to train SGD. We hope this work will encourage the use of high-quality data with focused AI approaches, like SGD, to accelerate the design of materials for catalysis. Crucially, having access to experimental data capturing phenomena that are hard to include in calculations, such as surface reconstruction and mass/energy transport, will boost the capabilities of data-centric AI approaches.

ACKNOWLEDGMENTS

This project was supported by the Max Planck Society's Research Network on Big-Data-Driven Materials Science (BigMax). H.I. Rivera-Arrieta acknowledges Secretaría de Educación, Ciencia, Tecnología e Innovación de la Ciudad de México (SECTEI), for the valuable support through their Postdoctoral Fellowship Program (Agreement SECTEI/107/2022). L. Foppa acknowledges 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). The authors thank Matthias Scheffler and colleagues from the Max-Planck-Cardiff Centre on the Fundamentals of Heterogeneous Catalysis (FUNCAT), Igor Kowalec, Lara Kabalan, Zhongwei Lu, David Willock, Andrew Logsdail, Richard Catlow, Michael Bowker, and Graham Hutchings for helpful discussions.

DATA AVAILABILITY

The input and output files for the calculations supporting this study's findings are available in the NOMAD repository at `http://doi.org/10.17172/NOMAD/2024.05.10-1`.

AUTHOR INFORMATION

Corresponding authors

Herzain I. Rivera-Arrieta
Email: rarrieta@fhi-berlin.mpg.de

Lucas Foppa
Email: foppa@fhi-berlin.mpg.de

REFERENCES

[1] J. H. Mercer, "West Antarctic ice sheet and CO_2 greenhouse effect: a threat of disaster," Nature **271**, 321–325 (1978).
[2] A. A. Lacis, G. A. Schmidt, D. Rind, and R. A. Ruedy, "Atmospheric CO_2: Principal Control Knob Governing Earth's Temperature," Science **330**, 356–359 (2010).
[3] T. R. Anderson, E. Hawkins, and P. D. Jones, "CO_2, the greenhouse effect and global warming: from the pioneering work of Arrhenius and Callendar to today's Earth System Models," Endeavour **40**, 178–187 (2016).
[4] S. Sonwani and P. Saxena, *Greenhouse Gases: Sources, Sinks and Mitigation* (Springer Nature Singapore Pte Ltd, 2022).
[5] J. Wang, Z. You, Q. Zhang, W. Deng, and Y. Wang, "Synthesis of lower olefins by hydrogenation of carbon dioxide over supported iron catalysts," Catalysis today **215**, 186–193 (2013).
[6] Z. Li, J. Wang, Y. Qu, H. Liu, C. Tang, S. Miao, Z. Feng, H. An, and C. Li, "Highly Selective Conversion of Carbon Dioxide to Lower Olefins," ACS Catalysis **7**, 8544–8548 (2017).
[7] S. Navarro-Jaén, M. Virginie, J. Bonin, M. Robert, R. Wojcieszak, and A. Y. Khodakov, "Highlights and challenges in the selective reduction of carbon dioxide to methanol," Nature Reviews Chemistry **5**, 564–579 (2021).
[8] Á. Morales-García, F. Viñes, J. R. Gomes, and F. Illas, "Concepts, models, and methods in computational heterogeneous catalysis illustrated through CO_2 conversion," Wiley Interdisciplinary Reviews: Computational Molecular Science **11**, e1530 (2021).

[9] B. An, Z. Li, Y. Song, J. Zhang, L. Zeng, C. Wang, and W. Lin, "Cooperative copper centres in a metal–organic framework for selective conversion of CO_2 to ethanol," Nature Catalysis **2**, 709–717 (2019).

[10] X. Wang, Z. Wang, F. P. García de Arquer, C.-T. Dinh, A. Ozden, Y. C. Li, D.-H. Nam, J. Li, Y.-S. Liu, J. Wicks, Z. Chen, M. Chi, B. Chen, Y. Wang, J. Tam, J. Y. Howe, A. Proppe, P. Todorović, F. Li, T.-T. Zhuang, C. M. Gabardo, A. R. Kirmani, C. McCallum, S.-F. Hung, Y. Lum, M. Luo, Y. Min, A. Xu, C. P. O'Brien, B. Stephen, B. Sun, A. H. Ip, L. J. Richter, S. O. Kelley, D. Sinton, and E. H. Sargent, "Efficient electrically powered CO_2-to-ethanol via suppression of deoxygenation," Nature Energy **5**, 478–486 (2020).

[11] M. Bowker, "Methanol Synthesis from CO_2 Hydrogenation," ChemCatChem **11**, 4238–4246 (2019).

[12] R. Cheula, T. A. M. Q. Tran, and M. Andersen, "Unraveling the Effect of Dopants in Zirconia-Based Catalysts for CO_2 Hydrogenation to Methanol," ACS Catalysis **14**, 13126–13135 (2024).

[13] G. Squadrito, G. Maggio, and A. Nicita, "The green hydrogen revolution," Renewable Energy **216**, 119041 (2023).

[14] M. Bertau, H. Offermanns, L. Plass, F. Schmidt, and H.-J. Wernicke, *Methanol: The Basic Chemical and Energy Feedstock of the Future* (Springer-Verlag, Berlin, Heidelberg, 2014).

[15] M. Bowker, S. DeBeer, N. F. Dummer, G. J. Hutchings, M. Scheffler, F. Schüth, S. H. Taylor, and H. Tüysüz, "Advancing Critical Chemical Processes for a Sustainable Future: Challenges for Industry and the Max Planck–Cardiff Centre on the Fundamentals of Heterogeneous Catalysis (FUNCAT)," Angewandte Chemie **134**, e202209016 (2022).

[16] M. Behrens, F. Studt, I. Kasatkin, S. Kühl, M. Hävecker, F. Abild-Pedersen, S. Zander, F. Girgsdies, P. Kurr, B.-L. Kniep, M. Tovar, R. W. Fischer, J. K. Nørskov, and R. Schögl, "The Active Site of Methanol Synthesis over $Cu/ZnO/Al_2O_3$ Industrial Catalysts," Science **336**, 893–897 (2012).

[17] H. Bahruji, M. Bowker, G. Hutchings, N. Dimitratos, P. Wells, E. Gibson, W. Jones, C. Brookes, D. Morgan, and G. Lalev, "Pd/ZnO catalysts for direct CO_2 hydrogenation to methanol," Journal of Catalysis **343**, 133–146 (2016).

[18] H. Bahruji, J. R. Esquius, M. Bowker, G. Hutchings, R. D. Armstrong, and W. Jones, "Solvent Free Synthesis of $PdZn/TiO_2$ Catalysts for the Hydrogenation of CO_2 to Methanol," Topics in catalysis **61**, 144–153 (2018).

[19] J. Díez-Ramírez, J. Díaz, P. Sánchez, and F. Dorado, "Optimization of the Pd/Cu ratio in Pd-Cu-Zn/SiC catalysts for the CO_2 hydrogenation to methanol at atmospheric pressure," Journal of CO_2 Utilization **22**, 71–80 (2017).

[20] M. Bowker, N. Lawes, I. Gow, J. Hayward, J. R. Esquius, N. Richards, L. R. Smith, T. J. Slater, T. E. Davies, N. F. Dummer, L. Kabalan, A. Logsdail, R. C. Catlow, S. Taylor, and G. J. Hutchings, "The Critical Role of βPdZn Alloy in Pd/ZnO Catalysts for the Hydrogenation of Carbon Dioxide to Methanol," ACS Catalysis **12**, 5371–5379 (2022).

[21] G. Pacchioni, "From CO_2 to Methanol on $Cu/ZnO/Al_2O_3$ Industrial Catalyst. What Do We Know about the Active Phase and the Reaction Mechanism?" ACS Catalysis **14**, 2730–2745 (2024).

[22] M. B. Fichtl, D. Schlereth, N. Jacobsen, I. Kasatkin, J. Schumann, M. Behrens, R. Schlögl, and O. Hinrichsen, "Kinetics of deactivation on $Cu/ZnO/Al_2O_3$ methanol synthesis catalysts," Applied Catalysis A: General **502**, 262–270 (2015).

[23] S. R. Docherty, N. Phongprueksathat, E. Lam, G. Noh, O. V. Safonova, A. Urakawa, and C. Copéret, "Silica-Supported PdGa Nanoparticles: Metal Synergy for Highly Active and Selective CO_2-to-CH_3OH hydrogenation," JACS Au **1**, 450–458 (2021).

[24] G. Di Liberto and G. Pacchioni, "Modeling Single-Atom Catalysis," Advanced Materials **35**, 2307150 (2023).

[25] G. Di Liberto, L. Giordano, and G. Pacchioni, "Predicting the Stability of Single-Atom Catalysts in Electrochemical Reactions," ACS Catalysis **14**, 45–55 (2024).

[26] R. T. Hannagan, G. Giannakakis, M. Flytzani-Stephanopoulos, and E. C. H. Sykes, "Single-Atom Alloy Catalysis," Chemical Reviews **120**, 12044–12088 (2020).

[27] R. T. Hannagan, G. Giannakakis, R. Réocreux, J. Schumann, J. Finzel, Y. Wang, A. Michaelides, P. Deshlahra, P. Christopher, M. Flytzani-Stephanopoulos, M. Stamatakis, and E. C. H. Sykes, "First-principles design of a single-atom–alloy propane dehydrogenation catalyst," Science **372**, 1444–1447 (2021).

[28] J. Schumann, Y. Bao, R. T. Hannagan, E. C. H. Sykes, M. Stamatakis, and A. Michaelides, "Periodic Trends in Adsorption Energies around Single-Atom Alloy Active Sites," The Journal of Physical Chemistry Letters **12**, 10060–10067 (2021).

[29] Z.-K. Han, D. Sarker, R. Ouyang, A. Mazheika, Y. Gao, and S. V. Levchenko, "Single-atom alloy catalysts designed by first-principles calculations and artificial intelligence," Nature Communications **12**, 1833 (2021).

[30] M. Andersen, S. V. Levchenko, M. Scheffler, and K. Reuter, "Beyond Scaling Relations for the Description of Catalytic Materials," ACS Catalysis **9**, 2752–2759 (2019).

[31] M. T. Darby, R. Réocreux, E. C. H. Sykes, A. Michaelides, and M. Stamatakis, "Elucidating the Stability and Reactivity of Surface Intermediates on Single-Atom Alloy Catalysts," ACS Catalysis **8**, 5038–5050 (2018).

[32] R. Réocreux and M. Stamatakis, "One Decade of Computational Studies on Single-Atom Alloys: Is In Silico Design within Reach?" Accounts of Chemical Research **55**, 87–97 (2021).

[33] W. Li, S. E. Madan, R. Réocreux, and M. Stamatakis, "Elucidating the Reactivity of Oxygenates on Single-Atom Alloy Catalysts," ACS Catalysis **13**, 15851–15868 (2023).

[34] Z. Chen and P. Zhang, "Electronic Structure of Single-Atom Alloys and Its Impact on The Catalytic Activities," ACS omega **7**, 1585–1594 (2022).

[35] Z. Xu, Z. Ao, M. Yang, and S. Wang, "Recent progress in single-atom alloys: Synthesis, properties, and applications in environmental catalysis," Journal of Hazardous Materials **424**, 127427 (2022).

[36] Q. Gao, X. Han, Y. Liu, and H. Zhu, "Electrifying Energy and Chemical Transformations with Single-Atom Alloy Nanoparticle catalysts," ACS catalysis **14**, 6045–6061 (2024).

[37] J. Liu, F. R. Lucci, M. Yang, S. Lee, M. D. Marcinkowski, A. J. Therrien, C. T. Williams, E. C. H. Sykes, and M. Flytzani-Stephanopoulos, "Tackling CO Poisoning with Single-Atom Alloy Catalysts," Journal of the American Chemical Society **138**, 6396–6399 (2016).

[38] A. D. Becke, "Perspective: Fifty years of density-functional theory in chemical physics," J. Chem. Phys. **140**, 18A301 (2014).

[39] I. Kowalec, L. Kabalan, C. R. A. Catlow, and A. J. Logsdail, "A computational study of direct CO_2 hydrogenation to methanol on Pd surfaces," Physical Chemistry Chemical Physics **24**, 9360–9373 (2022).

[40] H.-J. Freund and M. W. Roberts, "Surface chemistry of carbon dioxide," Surface Science Reports **25**, 225–273 (1996).

[41] A. Álvarez, M. Borges, J. J. Corral-Pérez, J. G. Olcina, L. Hu, D. Cornu, R. Huang, D. Stoian, and A. Urakawa, "CO_2 Activation over Catalytic Surfaces," ChemPhysChem **18**, 3135–3141 (2017).

[42] Z. Zhao and G. Lu, "Cu-Based Single-Atom Catalysts Boost Electroreduction of CO_2 to CH_3OH: First-Principles Predictions," The Journal of Physical Chemistry C **123**, 4380–4387 (2019).

[43] N. Zhang, Y. Si, X. Chen, X. Wang, and J. Yao, "Bimetallic Synergy in Single-Atom Alloy Nanocatalysts for CO_2 Reduction to Ethylene," ACS Applied Nano Materials **6**, 2394–2402 (2023).

[44] C. P. Gomes, B. Selman, and J. M. Gregoire, "Artificial intelligence for materials discovery," MRS Bulletin **44**, 538–544 (2019).

[45] Y. Hong, B. Hou, H. Jiang, and J. Zhang, "Machine learning and artificial neural network accelerated computational discoveries in materials science," Wiley Interdisciplinary Reviews: Computational Molecular Science **10**, e1450 (2020).

[46] P. Reiser, M. Neubert, A. Eberhard, L. Torresi, C. Zhou, C. Shao, H. Metni, C. van Hoesel, H. Schopmans, T. Sommer, and P. Friederich, "Graph neural networks for materials science and chemistry," Communications Materials **3**, 93 (2022).

[47] S. Wrobel, "An Algorithm for Multi-relational Discovery of Subgroups," in *European symposium on principles of data mining and knowledge discovery* (Springer, 1997) pp. 78–87.

[48] B. R. Goldsmith, M. Boley, J. Vreeken, M. Scheffler, and L. M. Ghiringhelli, "Uncovering structure-property relationships of materials by subgroup discovery," New Journal of Physics **19**, 013031 (2017).

[49] L. Foppa and L. M. Ghiringhelli, "Identifying Outstanding Transition-Metal-Alloy Heterogeneous Catalysts for the Oxygen Reduction and Evolution Reactions via Subgroup Discovery," Topics in Catalysis **65**, 196–206 (2022).

[50] J. Wellendorff, K. T. Lundgaard, K. W. Jacobsen, and T. Bligaard, "mBEEF: An accurate semi-local Bayesian error estimation density func-

tional," The Journal of Chemical Physics **140** (2014), 10.1063/1.4870397.
[51]D. Behrendt, S. Banerjee, C. Clark, and A. M. Rappe, "High-Throughput Computational Screening of Bioinspired Dual-Atom Alloys for CO_2 Activation," Journal of the American Chemical Society **145**, 4730–4735 (2023).
[52]P. L. Kress, S. Zhang, Y. Wang, V. Çınar, C. M. Friend, E. C. H. Sykes, and M. M. Montemore, "A Priori Design of Dual-Atom Alloy Sites and Experimental Demonstration of Ethanol Dehydrogenation and Dehydration on PtCrAg," Journal of the American Chemical Society **145**, 8401–8407 (2023).
[53]M. Ma, K. Djanashvili, and W. A. Smith, "Controllable Hydrocarbon Formation from the Electrochemical Reduction of CO_2 over Cu Nanowire Arrays," Angewandte Chemie International Edition **55**, 6680–6684 (2016).
[54]A. Bagger, W. Ju, A. S. Varela, P. Strasser, and J. Rossmeisl, "Electrochemical CO_2 reduction: classifying Cu facets," ACS Catalysis **9**, 7894–7899 (2019).
[55]S. Nitopi, E. Bertheussen, S. B. Scott, X. Liu, A. K. Engstfeld, S. Horch, B. Seger, I. E. Stephens, K. Chan, C. Hahn, J. K. Nørskov, T. F. Jaramillo, and I. Chorkendorff, "Progress and Perspectives of Electrochemical CO_2 Reduction on Copper in Aqueous Electrolyte," Chemical Reviews **119**, 7610–7672 (2019).
[56]V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, "Ab initio molecular simulations with numeric atom-centered orbitals," Computer Physics Communications **180**, 2175–2196 (2009).
[57]A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, , E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schött, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, "The atomic simulation environment—a Python library for working with atoms," Journal of Physics: Condensed Matter **29**, 273002 (2017).
[58]L. Kabalan, I. Kowalec, C. R. A. Catlow, and A. J. Logsdail, "A computational study of the properties of low-and high-index Pd, Cu and Zn surfaces," Physical Chemistry Chemical Physics **23**, 14649–14661 (2021).
[59]P. S. Schmidt and K. S. Thygesen, "Benchmark Database of Transition Metal Surface and Adsorption Energies from Many-Body Perturbation Theory," The Journal of Physical Chemistry C **122**, 4381–4390 (2018).
[60]R. Miyazaki, S. Faraji, S. V. Levchenko, L. Foppa, and M. Scheffler, "Vibrational frequencies utilized for the assessment of exchange-correlation functionals in the description of metal–adsorbate systems: C_2H_2 and C_2H_4 on transition-metal surfaces," Catal. Sci. Technol. , – (2024).
[61]A. Mazheika, Y.-G. Wang, R. Valero, F. Viñes, F. Illas, L. M. Ghiringhelli, S. V. Levchenko, and M. Scheffler, "Artificial-intelligence-driven discovery of catalyst genes with application to CO_2 activation on semiconductor oxides," Nature Communications **13**, 419 (2022).
[62]F. Calle-Vallejo, J. I. Martínez, J. M. García-Lastra, P. Sautet, and D. Loffreda, "Fast Prediction of Adsorption Properties for Platinum Nanocatalysts with Generalized Coordination Numbers," Angewandte Chemie International Edition **53**, 8316–8319 (2014).
[63]H. Li, J. Zhao, L. Luo, J. Du, and J. Zeng, "Symmetry-Breaking Sites for Activating Linear Carbon Dioxide Molecules," Accounts of Chemical Research **54**, 1454–1464 (2021).
[64]L. Foppa and M. Scheffler, "Coherent Collections of Rules Describing Exceptional Materials Identified with a Multi-Objective Optimization of Subgroups," arXiv preprint arXiv:2403.18437 (2024), 10.48550/arXiv.2403.18437.
[65]A. Ruban, H. L. Skriver, and J. K. Nørskov, "Surface segregation energies in transition-metal alloys," Physical Review B **59**, 15990 (1999).
[66]L. Foppa, C. Sutton, L. M. Ghiringhelli, S. De, P. Löser, S. A. Schunk, A. Schäfer, and M. Scheffler, "Learning Design Rules for Selective Oxidation Catalysts from High-Throughput Experimentation and Artificial Intelligence," ACS Catalysis **12**, 2223–2232 (2022).
[67]R. Miyazaki, K. S. Belthle, H. Tüysüz, L. Foppa, and M. Scheffler, "Materials Genes of CO_2 Hydrogenation on Supported Cobalt Catalysts: An Artificial Intelligence Approach Integrating Theoretical and Experimental Data," Journal of the American Chemical Society **146**, 5433–5444 (2024).
[68]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, and S. Wrabetz, "Towards Experimental Handbooks in Catalysis," Topics in Catalysis **63**, 1683–1699 (2020).