Insights into Supported Subnanometer Catalysts Exposed to CO via Machine-Learning-Enabled Multiscale Modeling

Yifan Wang,[1,2†] Ya-Qiong Su,[3,4†] Emiel J.M. Hensen,[4*] and Dionisios G. Vlachos[1,2*]

[1]Department of Chemical and Biomolecular Engineering, 150 Academy St., University of Delaware, Newark, Delaware 19716, United States

[2]Catalysis Center for Energy Innovation, RAPID Manufacturing Institute, and Delaware Energy Institute (DEI), 221 Academy St., University of Delaware, Newark, Delaware 19716, United States

[3]School of Chemistry, Xi'an Key Laboratory of Sustainable Energy Materials Chemistry, State Key Laboratory of Electrical Insulation and Power Equipment, Xi'an Jiaotong University, Xi'an 710049, China

[4]Laboratory of Inorganic Materials and Catalysis, Department of Chemical Engineering and Chemistry, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

*Corresponding authors: e.j.m.hensen@tue.nl, vlachos@udel.edu

†These authors contributed equally

ABSTRACT

Subnanometer catalysts offer high noble metal utilization and superior performance for several reactions. Yet, understanding their structures and properties on an atomic scale under working conditions is challenging due to the large configurational space. Here we introduce an efficient multiscale framework to predict their stability exposed to an adsorbate. The framework integrates a comprehensive toolset including density functional theory (DFT) calculations, cluster expansion, machine learning, and structure optimization. The end-to-end machine learning workflow guides DFT data generation and enables significant computational acceleration. We demonstrate the approach for CO-adsorbed Pd_n (n =1−55) clusters on $CeO_2(111)$. Simulation results reveal that CO can facilitate restructuring by stabilizing smaller planar structures and bilayer structures of specific intermediate sizes, consistent with experimental reports. Metal-support interactions, preferential CO adsorption, and metal nuclearity and structure control catalyst stability. The framework allows automatic discovery of stable catalyst structures and a systematic strategy to exploit properties in the subnanometer scale.

INTRODUCTION

Single-atom catalysts (SACs) and subnanometer clusters have recently emerged as an exciting field due to high noble metal utilization and superior performance in several chemistries.[1–4] Yet, little is known about the stability and structure of catalysts in this size regime. It is generally believed that SACs are prone to sintering, forming subnanometer clusters and nanoparticles,[5] and the actual catalyst state (single atoms vs. clusters of a few atoms ($n \leq 30$) vs. nanoparticles, or combinations) is often debated and subject to change under catalyst pretreatment or characterization and reaction.[6] This ambiguity stems from multiple experimental challenges, e.g., the inability to perform comprehensive *operando* characterization for most reactions, the low experimental signal-to-noise ratio at low catalyst loadings, the catalyst heterogeneity, and the sample sensitivity to the electron beam in microscopy experiments, to mention a few. Adsorbates, such as CO, could further complicate the structures and properties. CO is a reactant or a product of industrial relevant chemistries,[7] and often used as a probe molecule in characterization, such as chemisorption (site counting by assuming a specific stoichiometry, i.e., number of COs per metal atom) and infrared (IR) spectroscopy.[8] Due to its strong adsorption on noble metals, CO can block catalyst sites and cause catalyst restructuring.[9,10]

Many of these challenges can be overcome by multiscale simulations. The continuum Wulff construction fails in the subnanometer regime where structures are not well defined, support/catalyst interactions can be critical, and catalyst restructuring maybe the rule rather than the exception. Unlike previous gas-phase cluster calculations,[8,11] density functional theory (DFT) of supported catalysts can naturally account for all the above phenomena. Determination of stable structures via DFT is though challenging. First, DFT performs local structure optimization at zero Kelvin. It can miss catalytically relevant structures, among the numerous possible conformers, induced by restructuring unless coupled with a finite temperature optimization, e.g., a Metropolis algorithm. Even then, this approach is computationally demanding due to needing numerous energy evaluations for optimization. Second, supported nanoparticles require a massive DFT supercell that renders calculations impractical. Common generalized force fields lack charge transfer for interfacial phenomena as happens in supported catalysts. Not surprisingly, only a single cluster size at a time has been studied, e.g., Pt_8 on alumina exposed to H_2,[12] Au_8 on $CeO_2(111)$ exposed to CO,[13] and Pt_8 on $CeO_2(111)$ exposed to O_2.[14,15] Despite providing insights, understanding the most stable structure and size along with the effect of adsorbates over a large range of catalyst sizes has remained elusive.

Addressing these issues requires efficient and accurate energy mappings to replace expensive DFT calculations. The cluster expansion (CE) technique[16] can represent energies in terms of cluster patterns on a lattice, as demonstrated for alloys[17] and adsorbate layers on surfaces[18–21]. It allows to sample the configuration space and locate the lowest-energy (stable) structures using a global optimization algorithm. Modern machine learning (ML) models can further improve the accuracy of the CE and reduce overfitting.[22] We have also seen applications of ML on the approximation of adsorption energies.[23–25]

Here, combining CE, global optimization, and ML, we introduce an efficient multiscale framework to model the stability of single atoms, clusters of several atoms, and nanoparticles up to ~1.8 nm. We chose Pd ($n = 1$-55) on $CeO_2(111)$ in a CO atmosphere as a case study. We demonstrate that CO induces clustering and possible restructuring, and in contrast to common belief, stabilizes intermediate-size clusters. The most stable size changes nonlinearly with operating conditions and both growth and redispersion are seen. Various size clusters can co-exist. We expose profound variation of the stoichiometry (CO per Pd atom) and the adsorption energy with size for small clusters. Importantly, site heterogeneity is crucial and entails not only the adsorption type but also the degree of site unsaturation and a site-dependent electronic support-effect.

COMPUTATIONAL METHODS

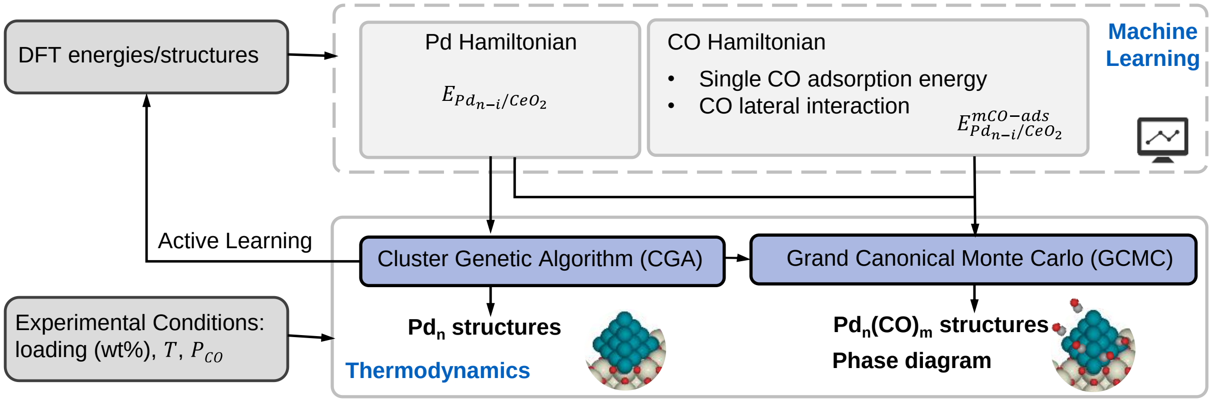

Figure 1. Modeling framework for subnanometer catalyst stability using first-principles multiscale modeling and machine learning. Color code: gray, C; red: O; light yellow, Ce; cyan, Pd.

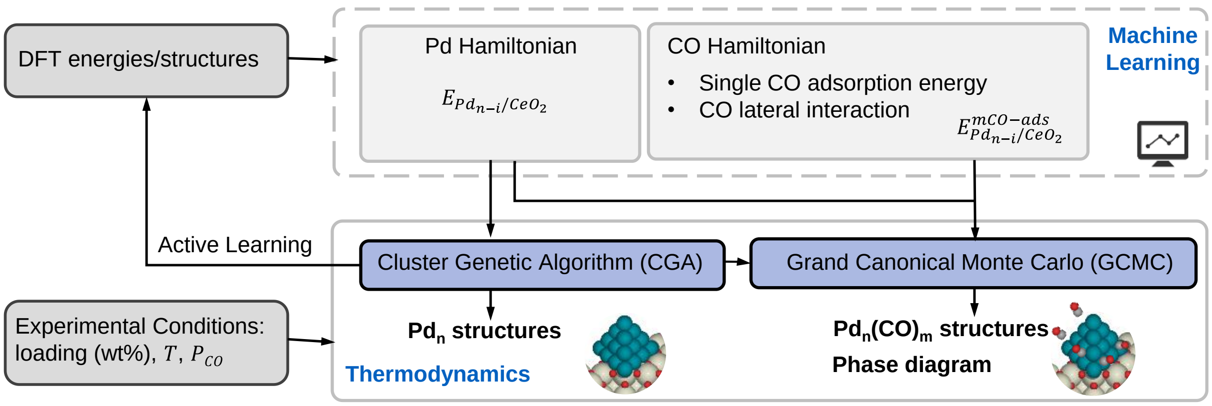

Figure 1 shows the machine-learning-enabled multiscale modeling framework. DFT calculations of supported metal clusters supply the ground truth structures and energies (details in the the Supporting Information (SI) Section 1). We mapped these structures to 3D lattice configurations for efficiency (SI Section 2) and trained machine learning energy models to replace DFT and enable structure optimization at finite temperatures. Two lattices were constructed based on the Pd atom locations and the CO adsorption sites, respectively. The first Hamiltonian described the bare Pd_n cluster energies given atom locations, and a cluster genetic algorithm (CGA) predicts the low energy structures.[26] A second Hamiltonian described CO adsorption. The adsorption energies of single CO molecules were approximated with a random-forest model. Both Hamiltonians captured interactions between nearest neighbors of Pd atoms (or CO adsorbates) using the CE approach. The models were improved iteratively using active learning. Notably, we optimized low-energy structures, passed them for new DFT calculations, and retrained the model until sufficiently accurate. Next, the GCMC simulations accelerated the sampling in the configurational space with temperature and CO pressure dependencies introduced via the chemical potential of CO. For each size and CO coverage, GCMC provided an ensemble ranging from the most stable structure (the global energy minimum) to higher energy minima (metastable structures). Experimental structures may be metastable or the global minimum, depending on the synthesis and the ability (the free energy barrier) of a cluster to "hop" from one energy minimum to another. This hoping entails complex, cooperative atom reorganization. Finally, we computed experimentally relevant information, i.e., surface phase diagrams, structural characterization, and calorimetry information.

RESULTS AND DISCUSSION

Single CO Adsorption on Clusters via a Descriptor-based Machine Learning Model

First, we developed a Hamiltonian of a single CO molecule adsorption. The CO adsorption energy $E_{Pd_{n-i}/CeO_2}^{CO-ads^{(x)}}$ at site x of a structure (the i^{th} isomer) of the $Pd_{n\text{-}i}$ cluster of n Pd atoms is

$$E_{Pd_{n-i}/CeO_2}^{CO-ads^{(x)}} = E_{Pd_{n-i}(CO)^{(x)}/CeO_2} - (E_{Pd_{n-i}/CeO_2} + E_{CO}) \qquad (1)$$

where $E_{Pd_{n-i}(CO)^{(x)}/CeO_2}, E_{Pd_{n-i}/CeO_2}, E_{CO}$ are the electronic energies of the supported cluster with CO-adsorbed at site x, a bare $Pd_{n\text{-}i}$ cluster, and a CO molecule in the gas phase, respectively, all obtained from DFT. Specifically, we performed DFT calculations of a single CO on a diverse set of 69 clusters of varying size ($n = 1 - 21$) and isomers, selected among low energy structures. The CO adsorption energy varies significantly with size (Figure 2e) and by as much as 1.5 eV on the same type of site (Figure 2f), depending on size and structure. Adsorption is generally less preferred on top sites. The adsorption energy exhibits modest variation with size for n >10 but some even for n~20. This dataset clearly shows the complexity in describing energies, adsorption sites, and structures of supported clusters compared to large nanoparticles of well-defined facets and the simplest case of single crystals. It is an enormous computational task to determine CO adsorption energies on all possible sites of each possible structure of a cluster. A descriptor-based approach[27,28] can be used to cope with this combinatorial problem. Most prior works exploited extended surfaces[29–33] rather than subnanometer clusters whose intimate contact with the support makes modeling more complicated.

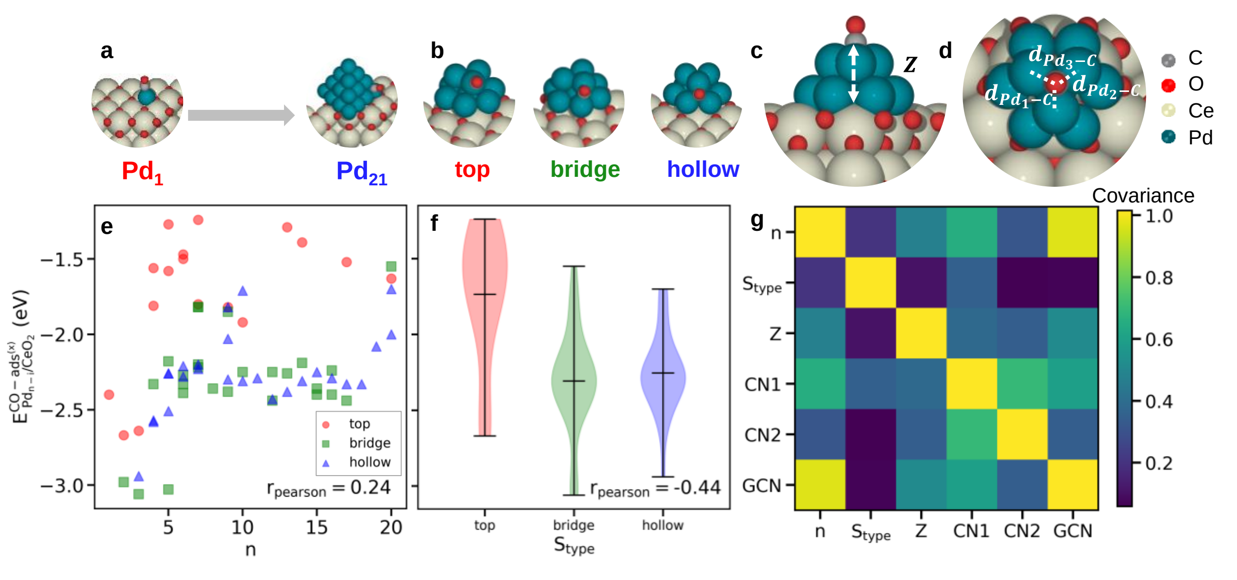

Figure 2. Descriptors in CO adsorption energy prediction. a-d, Example structures illustrating various geometric descriptors: (**a**) cluster size n, (**b**) site type $\boldsymbol{S_{type}}$, (**c**) the distance $\boldsymbol{Z}$ of the adsorption site from the support, and (**d**) the Pd-C bond lengths for a hollow site on a Pd_{10} cluster. **CN1** and **CN2** are the weighted average of the number of 1st or 2nd nearest neighbor Pd atoms to the Pd binding site; the weights are the reciprocals of the Pd-C bond lengths. **e**, DFT CO adsorption energy *vs*. n with the Pearson correlation coefficient ($\boldsymbol{r_{pearson}}$) shown. **f**, DFT CO adsorption energy *vs*. $\boldsymbol{S_{type}}$. The distributions are approximated

as Gaussian kernel density estimations (KDE). The maximum, the arithmetic mean, and the minimum are marked as short horizontal lines. **g**, The descriptor correlation matrix, showing the absolute covariance between descriptor pairs. The brighter colors indicate stronger correlations (higher covariances) between descriptor pairs.

We chose geometric descriptors (Figure 2a-d) to capture the local environment and enable thousands of on-the-fly energy calculations during structure optimization. Inspired by previous work,[33] we explored six descriptors: the cluster size n, the site type S_{type}, the coordination numbers (CNs: CN1, CN2 and GCN), and the distance of each site from the support Z. S_{type} is a categorical variable denoting top, bridge, and hollow sites that possess different spectroscopic and adsorption characteristics (see an overview in reference[8]). The CNs reflect the number of Pd atoms being 1st and 2nd nearest neighbors to the Pd binding site and its degree of unsaturation (2nd nearest neighbors are sufficient due to the small size of clusters). CN1 and CN2 are calculated here as the weighted averages of Pd neighbors to all binding Pd atoms, with weights being the reciprocals of the Pd-C bond length, obtained from DFT; the longer the Pd-C bond, the weaker the Pd-CO interaction and the lower the weight (SI Section 3). This allows us to account for cluster relaxations in the model. The Pearson correlation coefficient ($\boldsymbol{r_{pearson}}$), measuring the strength of correlation between variables, indicates that no single descriptor can capture CO adsorption energy solely (absolute $r_{pearson}$ <0.5, Figure S1). CNs, commonly employed for nanoparticles and extended surfaces, are insufficient for subnanometer clusters. The distance from the support, Z, and the Bader charge are strongly correlated ($r_{pearson} = -0.67$, Figure S2), implying that Z reflects electronic effects from the support.

Next, we identified and removed the linearly dependent descriptors by computing the correlation matrix (Figure 2g). The correlation coefficient between n and GCN is 0.967; GCN was eliminated. The final five descriptors include n, S_{type}, CN1, CN2 and Z. The adsorption energy at site x is expressed as $E_{Pd_{n-i}/CeO_2}^{CO-ads^{(x)}} = f(n, S_{type}^{(x)}, CN1^{(x)}, CN2^{(x)}, Z^{(x)})$. The effect of cluster structure is implicitly accounted for in the descriptors of its sites.

We trained a Random Forest (RF) model for the CO adsorption energy at a given site, with an RMSE of 0.23 eV (testing set) and an R^2 value of 0.92 (all data; Figure 3a), indicating a good fit. Figure 3b ranks the importance of descriptors and discovers the site type and the distance from the support as the most important ones, capturing 70% of the total data variability. We provided the model training details in the SI Section 5. Unlike extended systems, the metal-support interactions play a significant role in adsorption. Metal nuclearity and the degree of unsaturation of sites are secondary but also essential descriptors. We illustrated the adsorption sites and predicted energies for single CO molecules on a Pd_{20} cluster, which exposes facets, in Figure 3c,d. CO adsorption is stronger on bridge sites near the support and weaker on top corner sites. At low CO doses in IR and chemisorption and experiments invoking low CO coverages, these near the support sites will be occupied and dictate spectroscopic, calorimetric, and reactivity signatures.

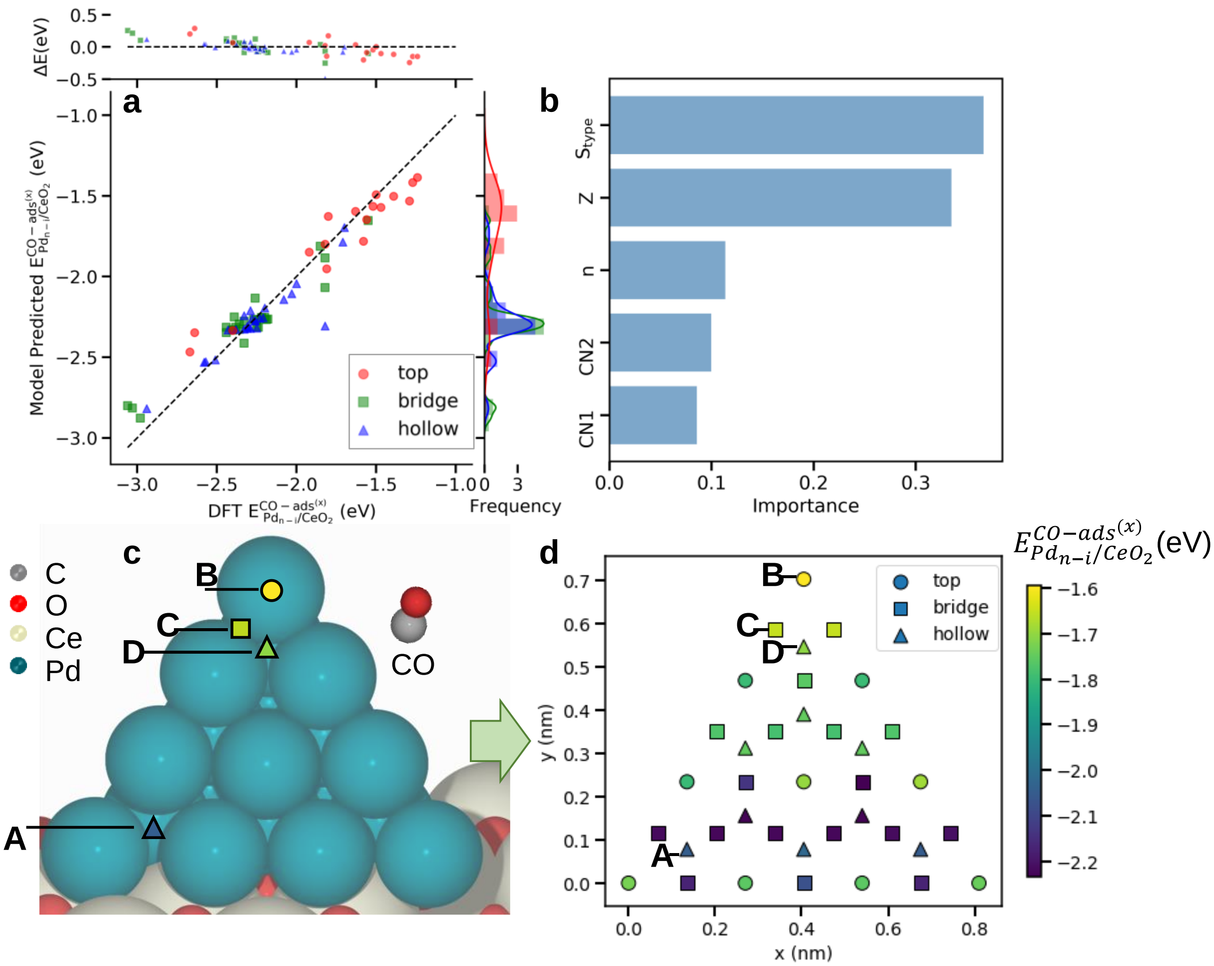

Figure 3. Random forest (RF) model for single CO adsorption energy on Pd_n clusters. a, Parity (dashed line) and DFT data (circles (top), squares (bridge), and triangles (hollow)). Points at the top represent the prediction error for all data. The histogram at the right indicates the density of data points in 0.1 eV bins of energy, normalized to the total number of data points; solid lines represent the distributions fitted with Gaussian kernel density estimates. b, Relative (decreasing) descriptor importance from top to bottom. c, Structure of a Pd_{20} cluster on $CeO_2(111)$. At the sites labelled by A (hollow), B (top), C (bridge), and D (hollow), the DFT CO adsorption energies are -2, -1.63, -1.55 and -1.7 eV, respectively. Color code: gray, C; red: O; light yellow, Ce; cyan, Pd. d, CO adsorption sites and energies (for a single CO molecule) projected on a 2D plane of a Pd_{20} cluster. The x axis is parallel to the support surface, whereas the y axis is perpendicular to the surface. The color of the site reflects the adsorption energy (right color bar). A, B, C, D correspond to the sites in panel c.

CO-CO Lateral Interactions

The RF model describes the adsorption energies of a single CO molecule per cluster. Under experimental conditions, the CO coverage could be high. A CE model of lateral (CO-CO) interactions was added to the single-molecule Hamiltonian. We considered the 2-body lateral interactions among CO-CO adsorbates on top, bridge, and hollow sites. As a result, the total adsorption energy is written as a lattice-gas Hamiltonian representing the energy of adsorption of m independent CO molecules (single molecule coverage limit) plus the lateral interaction between the CO molecules (see the SI Section 6). The final

RF+CE model gives the total adlayer energy $E_{Pd_{n-i}/CeO_2}^{mCO-ads}(\boldsymbol{\sigma})$ for a given adlayer configuration $\boldsymbol{\sigma}$**.** Throughout the data generation and model training, we followed the best ML practice for chemistry.[34] A complete checklist can be found in Table S1.

Optimization of Adsorbed Layer Structure using Grand Canonical Monte Carlo Simulations
Referenced to a CO reservoir, the free energy with m COs at a specific state can be written as (see derivation in the SI Section 8)

$$G(T, P_{CO}, \boldsymbol{\sigma}) = E_{Pd_{n-i}/CeO_2} + E_{Pd_{n-i}/CeO_2}^{mCO-ads}(\boldsymbol{\sigma}) - m[\Delta\mu_{CO}(T,\ P_0) + k_B T ln\left(\frac{P_{CO}}{P_0}\right)] \quad (2)$$

We systematically determined the most stable adsorbed layer structure at temperature T and partial pressure of CO P_{CO} by grand canonical Monte Carlo (GCMC) simulations.[35,36] We redesigned the traditional GCMC to improve its efficiency, resulting in a rejection-free "n-fold" algorithm (details in the SI Section 9). To avoid simultaneous optimization of structure and adlayer, we performed GCMC simulations on many precalculated low energy structures of Pd_{n-i} over a wide temperature (200 – 875 K) and a CO partial pressure range (10^{-11} – 10 bar) of relevance to systems that involve CO as a probe molecule, reactant, or product. We computed the average (overbar denotes average per Pd atom) free energy $\bar{G}$ and the CO loading $\bar{m}_{CO}$ (also known as stoichiometry) and compared the relative stability (see example isobars and isotherms in Figure S9).

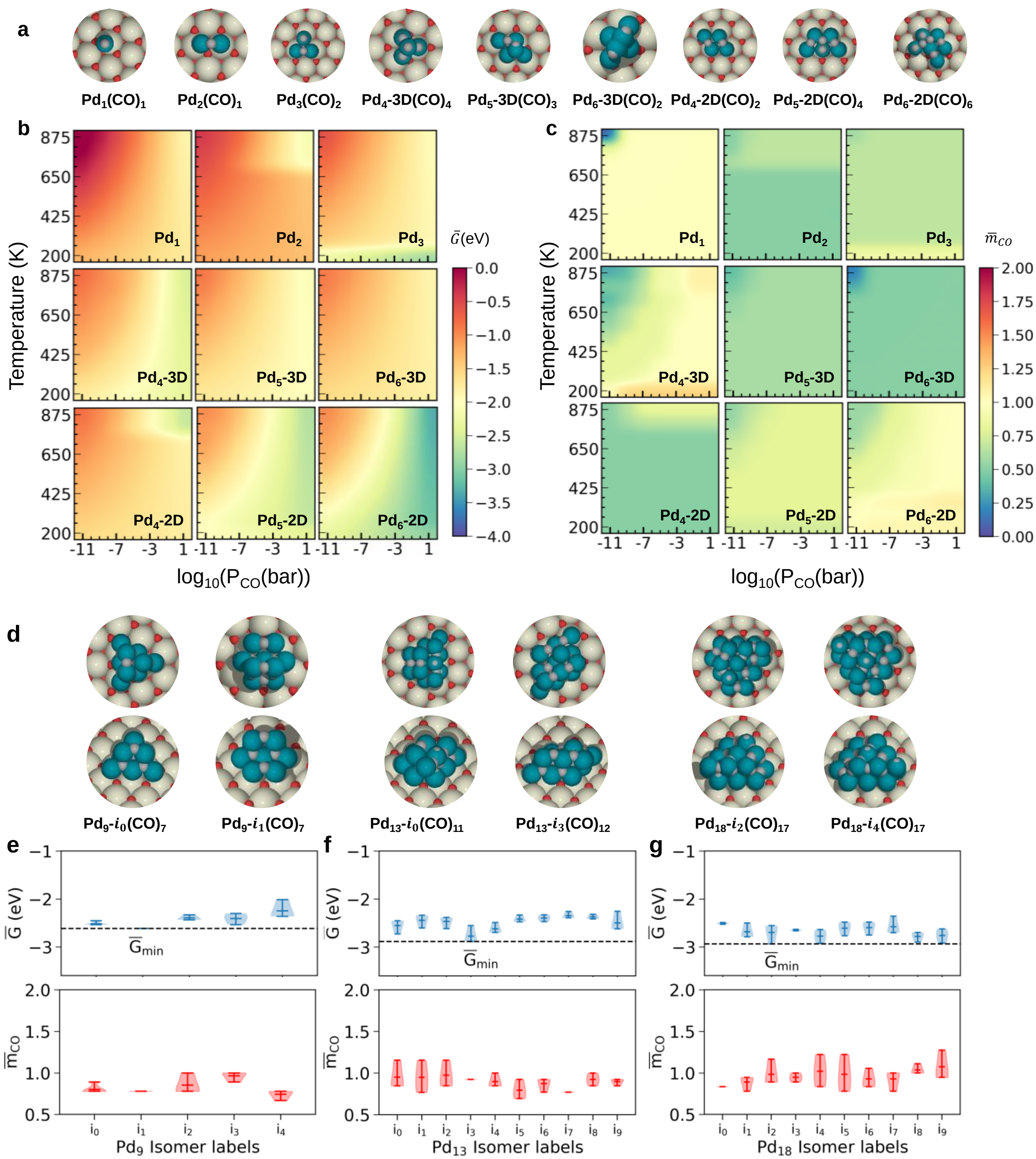

Figure 4. Structures, free energies, and stoichiometry of $E^{mCO-ads}_{Pd_{n-i}/CeO_2}(\sigma)$. a, The most stable structures ($n \leq 6$) predicted by GCMC at 300 K, CO pressure of 0.1 bar (only carbon atoms of CO at the adsorption sites are shown). For Pd4-6, the dimensionality (2D *vs*. 3D) is also indicated. **b**, Free energy per Pd atom $\overline{G}$(eV) as a function of temperature and CO pressure. Red/yellow colors indicate low absolute value (low stability), whereas green/blue colors indicate high stability. **c**, Optimum (the most stable free energy) adsorbed CO per Pd atom $\overline{m}_{CO}$ as a function of temperature and CO pressure. Red/yellow colors indicate more COs, and green/blue colors indicate fewer COs. **d**, Select stable (marked by $\overline{G}_{min}$ in (e)) and metastable structures ($n > 6$) predicted by GCMC at 300 K, CO pressure of 0.1 bar ($n = 9, 13, 18$;

top view (top row) and side view (bottom row)). **e-g**, Free energy and stoichiometry distributions of (e) Pd_9, (f) Pd_{13}, and (g) Pd_{18} isomers. The index $\boldsymbol{i}$ ($\boldsymbol{i} = \boldsymbol{0}, \boldsymbol{1}, \boldsymbol{2}, \ldots$) indicates the stability of bare isomers in descending order ($\boldsymbol{i_0} > \boldsymbol{i_1} > \boldsymbol{i_2} \ldots$). The lowest free energy isomer at each size is labeled as $\overline{\boldsymbol{G}}_{\boldsymbol{min}}$. The distributions are computed from multiple GCMC trajectories using a Gaussian kernel density estimation (KDE). The maximum, the arithmetic mean, and the minimum are marked as short horizontal lines.

Phase Behavior

For small clusters ($n \leq 6$), the $\bar{G}$ and $\bar{m}_{CO}$ are shown in Figure 4b, c, respectively. At high temperatures and a low CO pressure, bare clusters are the most stable for some sizes. Consistent with other computational work,[37] the clusters contain at least one CO under ultra-high vacuum conditions (10^{-11} bar) at lower temperatures (T < 700 K). Generally, low temperature favors more CO adsorption, but the CO loadings also depend on the cluster structure and the adsorption energy at each site. We visualized the cluster structures with the optimum $\bar{m}_{CO}$ and their positions at 300 K, 0.1 bar (typical low-temperature CO oxidation conditions) in Figure 4a. $\bar{m}_{CO}$ is smaller or equal to 1. $Pd_1(CO)$ is favored over a bare Pd. Unlike Rh or Pt, the DFT calculation shows the bicarbonyl $Pd_1(CO)_2$ on ceria is not stable. Pd_2 is stable with one adsorbed CO at the bridge site. Consistently, the DFT calculations revealed that the adsorption energy of CO is lower on the bridge site (E_{ads} = -2.98 eV) than the top site (E_{ads} = -2.67 eV). For Pd_3, CO adsorbs at one top (E_{ads} = -2.64 eV) and one bridge site (E_{ads} = -3.06 eV) instead of a hollow site (E_{ads} = -2.94 eV) or other more packed combinations due to the increased CO-CO repulsion. Pd_4-3D is interesting. Four COs can simultaneously adsorb on top Pd sites, due to the stretched-out atomic positions and the lower repulsion between CO molecules. The planar structures of bare Pd_5 and Pd_6 clusters (Pd_5-2D, $E_{Pd_{5-2D}}$ = -3.99 eV; Pd_6-2D, $E_{Pd_{6-2D}}$ = -4.49 eV) are less stable than the 3D structures (Pd_5-3D, $E_{Pd_{5-3D}}$ = -4.49 eV; Pd_6-3D, $E_{Pd_{5-3D}}$ = -4.95 eV). CO adsorption, however, stabilizes the planar structures compared to their 3D counterparts under moderate to high CO pressure ($P_{CO} > 10^{-3}$ bar); see the green region on the right side of Pd_5-2D and Pd_6-2D free energy heatmaps (Figure 4b). This is the first evidence for CO-induced restructuring.

Small clusters ($n \leq 6$) exhibit a limited number of structures that can be enumerated manually, whereas larger ones possess an exponential number. We explored the low energy structures of $n = 7 - 21$ (diameters below 1.2 nm). Due to the vast configurational space, we obtain an ensemble of isomers within 0.5 eV from the most stable isomer. This energy span corresponds to a huge difference in the probability of observing an isomer (2.5×10^8 at 300 K and 1400 at 800 K) and ensures that we can capture relevant isomers. For small clusters (e.g., n = 5 or 6), 2D or 3D structures can be stable depending on conditions, whereas for $n > 6$ only 3D structures are stable under most conditions. Figure 4d shows the two most stable isomers at 300 K, 0.1 bar CO pressure at select sizes (n = 9, 13, 18). Figure 4e-g shows the corresponding free energy and CO loading distributions of each isomer from multiple GCMC trajectories. The most stable i_0 bare isomers may restructure upon CO adsorption, as found for Pd_9-i_2, Pd_{13}-i_3 and Pd_{18}-i_2. The structures are more rod-like with Pd atoms arranged in parallel lines compared to the usually symmetric bare clusters. Our calculations provided clear evidence that CO induces cluster restructuring by lowering the free energy. For Pd_{18} (Figure 4g), several isomers (i_4, i_8, and i_9) have comparable low free energy (within 0.1 eV/atom) to the most stable isomer i_2, i.e., are nearly isoenergetic. This isoenergetic behavior implies that one structure could become more stable by adding or removing a single CO. The large variance in stoichiometry for some isomers, e.g., from ~0.7 to 1.25 on Pd_{18}-i_5, arising from site heterogeneity in a cluster is remarkable. CO favors bridge sites on the second layer or the edges in between the first two layers (followed by hollow and then top sites). For example, 7 CO molecules are adsorbed on Pd_9-i_1 on bridge sites, 12 on Pd_{13}-i_3 and 10 out of 17 on Pd_{18}-i_2.

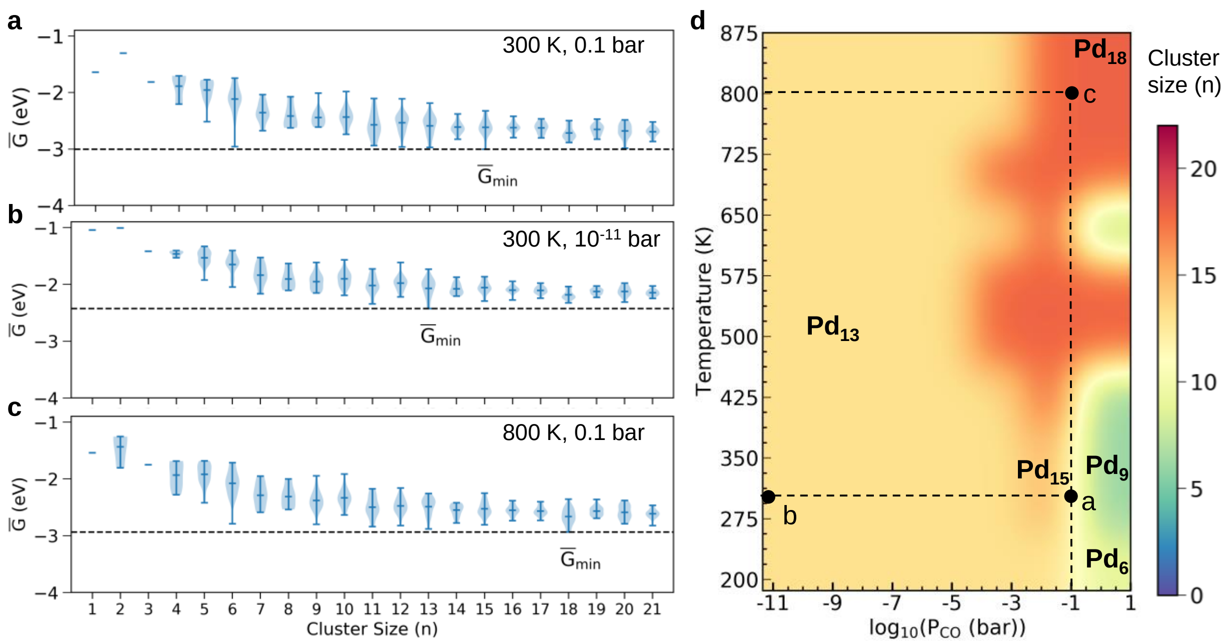

Figure 5. Predicted cluster stability. a-c, Free energy distributions *vs.* cluster size from GCMC simulations at: (a) 300 K, P_{CO} = 0.1 bar, (b) 300 K, $P_{CO} = 10^{-11}$ bar, and (c) 800 K, P_{CO} = 0.1 bar. The distributions are computed from the isomers at each size using a Gaussian kernel density estimation (KDE). The maximum, the mean, and the minimum are marked as short horizontal lines. The cluster size with the lowest free energy is labeled as $\overline{\boldsymbol{G}}_{\boldsymbol{min}}$. **d**, The most stable cluster size with conditions in (a)-(c) marked. Red/yellow colors indicate larger cluster sizes, and green/blue colors indicate smaller cluster sizes.

Figure 5a-c shows the relative stability *vs.* size at three conditions. First, we identified the lowest free energy isomers *vs.* size n. These isomers contribute to the low energy hull. Next, we selected the lowest $\bar{G}$ cluster size (marked by $\bar{G}_{min}$ in Figure 5a-c) at each T and P_{CO} (Figure S10 shows the corresponding stoichiometry distributions) and construct a phase diagram (Figure 5d). At high CO pressures, temperature plays a crucial role in determining stability: small clusters, e.g., Pd_6 and Pd_9, are the most stable at low temperatures, whereas larger clusters (Pd_{18}) are most stable at high temperatures. Pd_{13}, with a bilayer structure, is the most stable at low CO pressures ($P_{CO} < 10^{-3}$ bar) and dominates vacuum conditions. At low temperature/high CO pressure (condition a), several sizes, including Pd_6, Pd_9, Pd_{13}, and Pd_{15}, have comparable stability and should co-exist. Furthermore, redispersion from Pd_{13} to Pd_9 could occur upon CO pressure increase; growth and redispersion could alternate upon temperature increase at high P_{CO}. Previously, we observed the energy per atom decreases monotonically with increasing cluster size for bare Pd clusters on $CeO_2(111)$.[26] The variability in energy is higher for small clusters than larger clusters since more isoenergetic isomer structures exist for larger clusters. For example, small perturbations of a single atom in a cluster result in different isomers with nearly the same energy. In contrast, in small clusters, the energy levels are more distinct and the potential energy surface is more corrugated. Here, we observed the monotonically decreasing trend in free energy is broken by the CO adsorption. Surprisingly, we saw some intermediate-sized clusters (such as 13) are the most stable structures. We still expect the lower variability in free energy at larger sizes due to the isoenergetic structures in the low-energy ensemble.

Physical Insights

Experiments showed CO-induced cluster formation of Pd single atoms on ceria and other metal-support combinations in the range of 0.5-1.2 nm. For instance, Maurer *et al*. reported Pt_n clusters with a mean diameter of 1.2 nm and a coordination number of 7 at 150 °C and a CO pressure of 0.01 mbar.[38] Similarly, stable small nanoparticles (< 2 nm) were formed for Pt[39] and Pd[40] single atoms on ceria upon exposure to CO. For Pd single atoms on Fe_3O_4, Parkinson and coworkers determined clusters of 15-19 atoms at low CO pressure 5×10^{-10} mbar and room temperature.[41] For Pt_n on $TiO_2(110)$, after exposing 2D Pt_7 clusters to CO at 520 K, a transition from 2D-to-3D bilayer Pt_n structures with a mean size of ~15 atoms was observed. Intermediate-size clusters in CO oxidation were reported to be more active than single atoms, dimers, trimers, or larger nanoparticles.[42] Gänzler *et al*. proposed 1.4 nm as optimal for Pt_n on CeO_2.[43] Similarly, Heiz suggested Pd_n clusters (n = 14-20) were more active and Pd_{15} had maximum activity on MgO(111).[44] Zhou *et al.* proposed an optimal size for Pd_n on CuO of 0.7-1.1 nm.[45] While these studies manifested similar trends, they could not provide clear underlying reasons.

Our results agree qualitatively with the above experimental observations and underscore that intermediate sizes, rather than single atoms or nanoparticles, possess high thermodynamic stability under CO. The driving force forming larger nanoparticles is either non-existent or rather low. CO clearly induces structure and size changes compared to a fresh or under working conditions catalyst. Under low pressure, Pd_{13} is the most stable cluster (Figure 5c); clustering of single atoms is expected given sufficient thermal energy. We revealed that planar isomers of smaller clusters are more stable under high CO pressure, e.g., CO flattens the 3D-Pd_5 and Pd_6 (Figure 4a-c). On the other hand, bilayer and trilayer structures dominate intermediate sizes and larger nanoparticles, respectively, under most operating conditions due to the strong CO adsorption on the bridge and hollow sites (Figure 4d-g). A magic number-like behavior[46,47] (where certain sizes are stable, e.g., Pd_{13}) is seen, reminiscent of gas-phase clusters, but the behavior is much more complex. We predicted the stable structures of larger systems up to n = 55 (~1.8 nm in size) by extrapolating the model beyond the training sizes at conditions a (SI Section 11). The free energy landscape is relatively flat above a cluster size with some exceptions. The most stable cluster size changes non-monotonically with operating conditions; modest growth or redispersion is observed as operating conditions vary. Interestingly, we predict that under certain conditions, clusters of multiple sizes can co-exist due to comparable thermodynamic stability. Maximizing the number of exposed Pd atoms, while forming suitable adsorption sites, controlled by the size, the isomer, and strong metal-support effects, are all at play in favoring specific cluster and nanoparticle sizes.

The stoichiometry of supported metal clusters varies considerably and non-monotonically with nuclearity, from as low as 1/3, e.g., Pd_6-3D$(CO)_2$, to higher than 1 (Figure 5). Different isomers of the same size can exhibit profoundly different stoichiometry, e.g., n=6 (Figure 4a). Ensemble average properties are often reported in CO chemisorption and calorimetric experiments. In the latter, the molar heat of adsorption is measured and plotted against the coverage.[48] Both properties can be computed from simulations (see the SI Section 10). The ensemble average stoichiometry and the adsorption energy oscillate due to isomerism and site heterogeneity for clusters up to n=21 atoms but converge to size-independent properties (~1 and ~-2.2 eV, respectively, Figure S12d) as the size increases ($n \geq 25$). Descriptor-based machine learning analysis (SI Section 12) reveals the attributes of equilibrated structures: CO prefers to adsorb on hollow and bridge sites near the support. The primary descriptor affecting the stability (free energy) is the adsorption site type followed by the distance of the site from the support and the coordination numbers. Unlike extended surfaces, and expectedly, site heterogeneity (type, coordination, and distance from the support) is clearly manifested. Given the sensitivity on nuclearity, simulations will be essential to complement such experiments given the non-monotonic change of properties. These, in turn, modify characterization, e.g., calorimetric, chemisorption, and potentially IR signatures.

Potential limitations and uncertainties of the approach could arise from the DFT data and the ML training. Our method captures configurational entropy through the Metropolis sampling scheme. The vibrational entropy of the adsorbed CO to the adsorption energies, which is less than 0.03 eV on metal surfaces,[49–52] is left out. The Pd atoms' vibrational contribution can become significant at high temperatures, as pointed out in the literature.[53] To simplify the modeling effort, we focused on the CO pressure effect and leave out Pd-O interactions,[54] such as O vacancy formation[55,56] that could be important at high temperatures. Since the Hamiltonians are additive, the framework allows extensions to new species and sites. Future work should incorporate such effects to predict structures under specific experimental conditions.

CONCLUSIONS

The computational framework we introduced is a generic blueprint. It integrates DFT, molecular modeling, machine learning, and optimization to determine stable and metastable structures of supported metal subnanometer and nanoparticles catalysts in the presence of an adsorbate *vs*. metal nuclearity. The approach differs radically from continuum approximations, such as the Wulff construction, which fail to account for site-dependent metal-support interactions, non-well faceted particles, and stabilization of preferred sites by CO adsorption. We employed an end-to-end modular machine learning framework that integrates data collection, model training, and deployment for prediction. The physics rules-based models are interpretable and transferable to larger nanoparticles. The active learning, the rejection-free n-fold GCMC, and the lattice representations provide significant computational speedup while accounting for subtle effects of structure relaxation and strain. As such, the proposed framework opens vast opportunities to understand and close the gap with characterization data. It applies equally to predicting structure during a chemical reaction when the most abundant surface intermediate species is known.

ASSOCIATED CONTENT

Supporting Information

The supporting information includes (1) additional GCMC simulation results; (2) derivations for the free energy in the Grand Canonical ensemble; (3) details of the machine learning and GCMC methods; and (4) machine learning analysis for stability. The supporting information is available free of charge.

Data Availability

Data of this study are available within the paper and its Supporting Information, or in the associated GitHub repository described in the section on Code Availability. The repository contains 69 DFT structures for single CO adsorption, 21 DFT structures for multiple CO molecules adsorption, the low energy structures obtained from GCMC at select sizes, and all descriptor values of equilibrium structures.

Code Availability

The Python codes for the lattice mapping, cluster expansion, and GCMC simulations are available on GitHub at https://github.com/VlachosGroup/Pdn-CO-Stability. Instructions on how to train and run the models are included in the readme files. The main readme file of the repository describes the directory structure.

Competing interests

The authors declare no competing financial interest.

Acknowledgements

YW's and DGV's work was supported as part of the Catalysis Center for Energy Innovation, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, and Office of Basic Energy Sciences under award number DE-SC0001004. Their software development was supported from Department of Energy's Office of Energy Efficient and Renewable Energy's Advanced Manufacturing Office under Award Number DE-EE0007888-9.5. The Delaware Energy Institute gratefully acknowledges the support and partnership of the State of Delaware toward the RAPID projects. E.J.M.H. and Y.Q.S. acknowledge the financial support from The Netherlands Organization for Scientific Research (NWO) through a Vici grant. Supercomputing facilities were provided by NWO and Hefei Advanced Computing Center and from the European Union's Horizon 2020 research and innovation programme under grant No 686086 (Partial-PGMs). Y.Q.S. acknowledge the "Young Talent Support Plan" Fellowship of Xi'an Jiaotong University.

Author contributions

Y.Q.S. contributed to the DFT calculations. Y.W. contributed to the rest of the modeling framework. E.J.M.H. and D.G.V. provided supervision. The manuscript was written by Y.W. and D.G.V. All authors contributed to analyzing the results and commenting on the manuscript.

References

(1) Qiao, B.; Wang, A.; Yang, X.; Allard, L. F.; Jiang, Z.; Cui, Y.; Liu, J.; Li, J.; Zhang, T. Single-Atom Catalysis of CO Oxidation Using Pt 1 / FeO X. *Nat. Chem.* **2011**, *3*, 634–641.

(2) Peterson, E. J.; DeLaRiva, A. T.; Lin, S.; Johnson, R. S.; Guo, H.; Miller, J. T.; Kwak, J. H.; Peden, C. H. F.; Kiefer, B.; Allard, L. F.; Ribeiro, F. H.; Datye, A. K. Low-Temperature Carbon Monoxide Oxidation Catalysed by Regenerable Atomically Dispersed Palladium on Alumina. *Nat. Commun.* **2014**, *5*, 1–11.

(3) Riley, C.; Zhou, S.; Kunwar, D.; De La Riva, A.; Peterson, E.; Payne, R.; Gao, L.; Lin, S.; Guo, H.; Datye, A. Design of Effective Catalysts for Selective Alkyne Hydrogenation by Doping of Ceria with a Single-Atom Promotor. *J. Am. Chem. Soc.* **2018**, *140*, 12964–12973.

(4) Kyriakou, G.; Boucher, M. B.; Jewell, A. D.; Lewis, E. A.; Lawton, T. J.; Baber, A. E.; Tierney, H. L.; Flytzani-Stephanopoulos, M.; Sykes, E. C. H. Isolated Metal Atom Geometries as a Strategy for Selective Heterogeneous Hydrogenations. *Science* **2012**, *335*, 1209–1212.

(5) Ouyang, R.; Liu, J. X.; Li, W. X. Atomistic Theory of Ostwald Ripening and Disintegration of Supported Metal Particles under Reaction Conditions. *J. Am. Chem. Soc.* **2013**, *135*, 1760–1771.

(6) Tang, Y.; Asokan, C.; Xu, M.; Graham, G. W.; Pan, X.; Christopher, P.; Li, J.; Sautet, P. Rh Single Atoms on TiO2 Dynamically Respond to Reaction Conditions by Adapting Their Site. *Nat. Commun.* **2019**, *10*, 4488.

(7) Stefanidis, G. D.; Vlachos, D. G. Intensification of Steam Reforming of Natural Gas: Choosing Combustible Fuel and Reforming Catalyst. *Chem. Eng. Sci.* **2010**, *65*, 398–404.

(8) Lansford, J. L.; Vlachos, D. G. Infrared Spectroscopy Data- and Physics-Driven Machine Learning for Characterizing Surface Microstructure of Complex Materials. *Nat. Commun.* **2020**, *11*, 1–12.

(9) Wang, Y.; Mei, D.; Glezakou, V.; Li, J.; Rousseau, R. Dynamic Formation of Single-Atom

Catalytic Active Sites on Ceria-Supported Gold Nanoparticles. *Nat. Commun.* **2015**, *6*, 6511.

(10) Liu, J.; Wang, Y.; Li, J. Toward Rational Design of Oxide-Supported Single-Atom Catalysts: Atomic Dispersion of Gold on Ceria. *J. Am. Chem. Soc.* **2017**, *139*, 6190–6199.

(11) Sun, G.; Sautet, P. Metastable Structures in Cluster Catalysis from First-Principles: Structural Ensemble in Reaction Conditions and Metastability Triggered Reactivity. *J. Am. Chem. Soc.* **2018**, *140*, 2812–2820.

(12) Sun, G.; Alexandrova, A. N.; Sautet, P. Pt 8 Cluster on Alumina under a Pressure of Hydrogen: Support-Dependent Reconstruction from First-Principles Global Optimization. *J. Chem. Phys.* **2019**, *151*, 194703.

(13) Chang, M. W.; Zhang, L.; Davids, M.; Filot, I. A. W.; Hensen, E. J. M. Dynamics of Gold Clusters on Ceria during CO Oxidation. *J. Catal.* **2020**, *392*, 39–47.

(14) Liu, J. X.; Su, Y.; Filot, I. A. W.; Hensen, E. J. M. A Linear Scaling Relation for CO Oxidation on CeO2-Supported Pd. *J. Am. Chem. Soc.* **2018**, *140*, 4580–4587.

(15) Wang, H.; Liu, J.-X.; Allard, L. F.; Lee, S.; Liu, J.; Li, H.; Wang, J.; Wang, J.; Oh, S. H.; Li, W.; Flytzani-Stephanopoulos, M.; Shen, M.; Goldsmith, B. R.; Yang, M. Surpassing the Single-Atom Catalytic Activity Limit through Paired Pt-O-Pt Ensemble Built from Isolated Pt1 Atoms. *Nat. Commun.* **2019**, *10*, 1–12.

(16) Sanchez, J. M.; Ducastelle, F.; Gratias, D. Generalized Cluster Description of Multicomponent Systems. *Physica A* **1984**, *128*, 334–350.

(17) Hart, G. L. W.; Blum, V.; Walorski, M. J.; Zunger, A. Evolutionary Approach for Determining First-Principles Hamiltonians. *Nat. Mater.* **2005**, *4*, 391–394.

(18) Drautz, R.; Singer, R.; Fähnle, M. Cluster Expansion Technique: An Efficient Tool to Search for Ground-State Configurations of Adatoms on Plane Surfaces. *Phys. Rev. B* **2003**, *67*, 1–4.

(19) Schmidt, D. J.; Chen, W.; Wolverton, C.; Schneider, W. F. Performance of Cluster Expansions of Coverage-Dependent Adsorption of Atomic Oxygen on Pt(111). *J. Chem. Theory Comput.* **2012**, *8*, 264–273.

(20) Herder, L. M.; Bray, J. M.; Schneider, W. F. Comparison of Cluster Expansion Fitting Algorithms for Interactions at Surfaces. *Surf. Sci.* **2015**, *640*, 104–111.

(21) Miller, S. D.; Kitchin, J. R. Uncertainty and Figure Selection for DFT Based Cluster Expansions for Oxygen Adsorption on Au and Pt (111) Surfaces. *Mol. Simul.* **2009**, *35*, 920–927.

(22) Nelson, L. J.; Ozoliņš, V.; Reese, C. S.; Zhou, F.; Hart, G. L. W. Cluster Expansion Made Easy with Bayesian Compressive Sensing. *Phys. Rev. B* **2013**, *88*, 155105.

(23) Ma, X.; Li, Z.; Achenie, L. E. K.; Xin, H. Machine-Learning-Augmented Chemisorption Model for CO2 Electroreduction Catalyst Screening. *J. Phys. Chem. Lett.* **2015**, *6*, 3528–3533.

(24) Back, S.; Yoon, J.; Tian, N.; Zhong, W.; Tran, K.; Ulissi, Z. W. Convolutional Neural Network of Atomic Surface Structures to Predict Binding Energies for High-Throughput Screening of Catalysts. *J. Phys. Chem. Lett.* **2019**, *10*, 4401–4408.

(25) Ulissi, Z. W.; Tang, M. T.; Xiao, J.; Liu, X.; Torelli, D. A.; Karamad, M.; Cummins, K.; Hahn, C.; Lewis, N. S.; Jaramillo, T. F.; Chan, K.; Nørskov, J. K. Machine-Learning Methods Enable Exhaustive Searches for Active Bimetallic Facets and Reveal Active Site Motifs for CO 2

Reduction. *ACS Catal.* **2017**, *7*, 6600–6608.

(26) Wang, Y.; Su, Y.-Q.; Hensen, E. J. M.; Vlachos, D. G. Finite-Temperature Structures of Supported Subnanometer Catalysts Inferred via Statistical Learning and Genetic Algorithm-Based Optimization. *ACS Nano* **2020**, *14*, 13995–14007.

(27) Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V; Draxl, C.; Scheffler, M. Big Data of Materials Science: Critical Role of the Descriptor. *Phys. Rev. Lett.* **2015**, *114*, 105503.

(28) O'Connor, N. J.; Jonayat, A. S. M.; Janik, M. J.; Senftle, T. P. Interaction Trends between Single Metal Atoms and Oxide Supports Identified with Density Functional Theory and Statistical Learning. *Nat. Catal.* **2018**, *1*, 531–539.

(29) Calle-Vallejo, F.; Tymoczko, J.; Colic, V.; Vu, Q. H.; Pohl, M. D.; Morgenstern, K.; Loffreda, D.; Sautet, P.; Schuhmann, W.; Bandarenka, A. S. Finding Optimal Surface Sites on Heterogeneous Catalysts by Counting Nearest Neighbors. *Science* **2015**, *350*, 185–189.

(30) Guo, W.; Vlachos, D. G. Effect of Local Metal Microstructure on Adsorption on Bimetallic Surfaces: Atomic Nitrogen on NiPt(111). *J. Chem. Phys.* **2013**, *138*.

(31) Mpourmpakis, G.; Andriotis, A. N.; Vlachos, D. G. Identification of Descriptors for the CO Interaction with Metal Nanoparticles. *Nano Lett.* **2010**, *156801*.

(32) Dean, J.; Taylor, M. G.; Mpourmpakis, G. Unfolding Adsorption on Metal Nanoparticles: Connecting Stability with Catalysis. *Sci. Adv.* **2019**, *5*.

(33) Giles, S. A.; Yan, Y.; Vlachos, D. G. Effect of Substitutionally Doped Graphene on the Activity of Metal Nanoparticle Catalysts for the Hydrogen Oxidation Reaction. *ACS Catal.* **2019**, *9*, 1129–1139.

(34) Artrith, N.; Butler, K. T.; Coudert, F.-X.; Han, S.; Isayev, O.; Jain, A.; Walsh, A. Best Practices in Machine Learning for Chemistry. *Nat. Chem.* **2021**, *13*, 505–508.

(35) Frenkel, D.; Smit, B.; Ratner, M. A. *Understanding Molecular Simulation: From Algorithms to Applications*; Elsevier, 2001; Vol. 1, 126-135.

(36) Wexler, R. B.; Qiu, T.; Rappe, A. M. Automatic Prediction of Surface Phase Diagrams Using Ab Initio Grand Canonical Monte Carlo. *J. Phys. Chem. C* **2019**, *123*, 2321–2328.

(37) Kauppinen, M. M.; Melander, M. M.; Honkala, K. First-Principles Insight into CO Hindered Agglomeration of Rh and Pt Single Atoms on m -ZrO 2. *Catal. Sci. Technol.* **2020**, *10*, 5847–5855.

(38) Maurer, F.; Jelic, J.; Wang, J.; Gänzler, A.; Dolcet, P.; Wöll, C.; Wang, Y.; Studt, F.; Casapu, M.; Grunwaldt, J. Tracking the Formation, Fate and Consequence for Catalytic Activity of Pt Single Sites on CeO2. *Nat. Catal.* **2020**, *3*, 824–833.

(39) Jiang, D.; Wan, G.; García-Vargas, C. E.; Li, L.; Pereira-Hernández, X. I.; Wang, C.; Wang, Y. Elucidation of the Active Sites in Single-Atom Pd 1 /CeO 2 Catalysts for Low-Temperature CO Oxidation . *ACS Catal.* **2020**, 11356–11364.

(40) Pereira-Hernández, X. I.; DeLaRiva, A.; Muravev, V.; Kunwar, D.; Xiong, H.; Sudduth, B.; Engelhard, M.; Kovarik, L.; Hensen, E. J. M.; Wang, Y.; Datye, A. K. Tuning Pt-CeO2 Interactions by High-Temperature Vapor-Phase Synthesis for Improved Reducibility of Lattice Oxygen. *Nat. Commun.* **2019**, *10*, 1358.

(41) Parkinson, G. S.; Novotny, Z.; Argentero, G.; Schmid, M.; Pavelec, J.; Kosak, R.; Blaha, P.; Diebold, U. Carbon Monoxide-Induced Adatom Sintering in a Pd-Fe3O4model Catalyst. *Nat. Mater.* **2013**, *12*, 724–728.

(42) Cargnello, M.; Doan-Nguyen, V. V. T.; Gordon, T. R.; Diaz, R. E.; Stach, E. A.; Gorte, R. J.; Fornasiero, P.; Murray, C. B. Control of Metal Nanocrystal Size Reveals Metal-Support Interface Role for Ceria Catalysts. *Science* **2013**, *341*, 771–773.

(43) Gänzler, A. M.; Casapu, M.; Maurer, F.; Störmer, H.; Gerthsen, D.; Ferré, G.; Vernoux, P.; Bornmann, B.; Frahm, R.; Murzin, V.; Nachtegaal, M.; Votsmeier, M.; Grunwaldt, J.-D. Tuning the Pt/CeO 2 Interface by in Situ Variation of the Pt Particle Size. *ACS Catal.* **2018**, *8*, 4800–4811.

(44) Heiz, U.; Sanchez, A.; Abbet, S.; Schneider, W. D. Catalytic Oxidation of Carbon Monoxide on Monodispersed Platinum Clusters: Each Atom Counts. *J. Am. Chem. Soc.* **1999**, *121*, 3214–3217.

(45) Zhou, X.; Yang, W.; Chen, Q.; Geng, Z.; Shao, X.; Li, J.; Wang, Y.; Dai, D.; Chen, W.; Xu, G.; Yang, X.; Wu, K. Stable Pt Single Atoms and Nanoclusters on Ultrathin CuO Film and Their Performances in CO Oxidation. *J. Phys. Chem. C* **2016**, *120*, 1709–1715.

(46) Chen, M.; Straatsma, T. P.; Fang, Z.; Dixon, D. A. Structural and Electronic Property Study of (ZnO)n, n ≤ 168: Transition from Zinc Oxide Molecular Clusters to Ultrasmall Nanoparticles. *J. Phys. Chem. C* **2016**, *120*, 20400–20418.

(47) Nikbin, N.; Austin, N.; Vlachos, D. G.; Stamatakis, M.; Mpourmpakis, G. Catalysis at the Sub-Nanoscale: Complex CO Oxidation Chemistry on a Few Au Atoms. *Catal. Sci. Technol.* **2015**, *5*, 134–141.

(48) Hörtz, P.; Ruff, P.; Schäfer, R. A Temperature Dependent Investigation of the Adsorption of CO on Pt(111) Using Low-Temperature Single Crystal Adsorption Calorimetry. *Surf. Sci.* **2015**, *639*, 66–69.

(49) Beniya, A.; Isomura, N.; Hirata, H.; Watanabe, Y. Low Temperature Adsorption and Site-Conversion Process of CO on the Ni(111) Surface. *Surf. Sci.* **2012**, *606*, 1830–1836.

(50) Ge, Q.; King, D. A. Surface Diffusion Potential Energy Surfaces from First Principles: CO Chemisorbed on Pt{110}. *J. Chem. Phys.* **1999**, *111*, 9461–9464.

(51) Abild-Pedersen, F.; Andersson, M. P. CO Adsorption Energies on Metals with Correction for High Coordination Adsorption Sites - A Density Functional Study. *Surf. Sci.* **2007**, *601*, 1747–1753.

(52) Feibelman, P. J.; Hammer, B.; Nørskov, J. K.; Wagner, F.; Scheffler, M.; Stumpf, R.; Watwe, R.; Dumesic, J. The CO/Pt(111) Puzzle. *J. Phys. Chem. B* **2001**, *105*, 4018–4025.

(53) Craievich, P.; Sanchez, J.; Watson, R.; Weinert, M. Structural Instabilities of Excited Phases. *Phys. Rev. B* **1997**, *55*, 787–797.

(54) Spezzati, G.; Su, Y.; Hofmann, J. P.; Benavidez, A. D.; DeLaRiva, A. T.; McCabe, J.; Datye, A. K.; Hensen, E. J. M. Atomically Dispersed Pd-O Species on CeO2(111) as Highly Active Sites for Low-Temperature CO Oxidation. *ACS Catal.* **2017**, *7*, 6887–6891.

(55) Li, Y.; Kottwitz, M.; Vincent, J. L.; Enright, M. J.; Liu, Z.; Zhang, L.; Huang, J.; Senanayake, S. D.; Yang, W.-C. D.; Crozier, P. A.; Nuzzo, R. G.; Frenkel, A. I. Dynamic Structure of Active Sites in Ceria-Supported Pt Catalysts for the Water Gas Shift Reaction. *Nat. Commun.* **2021**, *12*, 1–9.

(56) Dvořák, F.; Farnesi Camellone, M.; Tovt, A.; Tran, N.-D.; Negreiros, F. R.; Vorokhta, M.; Skála,

T.; Matolínová, I.; Mysliveček, J.; Matolín, V.; Fabris, S. Creating Single-Atom Pt-Ceria Catalysts by Surface Step Decoration. *Nat. Commun.* **2016**, *7*, 10801.

TOC Graphic