# Interaction trends between single metal atoms and oxide supports identified with density functional theory and statistical learning

Nolan J. O'Connor[1,4], A. S. M. Jonayat[2,4], Michael J. Janik[1]* and Thomas P. Senftle [1,3]*

Single-atom catalysts offer high reactivity and selectivity while maximizing utilization of the expensive active metal component. However, they are susceptible to sintering, where single metal atoms agglomerate into thermodynamically stable clusters. Tuning the binding strength between single metal atoms and oxide supports is essential to prevent sintering. We apply density functional theory, together with a statistical learning approach based on least absolute shrinkage and selection operator regression, to identify property descriptors that predict interaction strengths between single metal atoms and oxide supports. Here, we show that interfacial binding is correlated with readily available physical properties of both the supported metal, such as oxophilicity measured by oxide formation energy, and the support, such as reducibility measured by oxygen vacancy formation energy. These properties can be used to empirically screen interaction strengths between metal–support pairs, thus aiding the design of single-atom catalysts that are robust against sintering.

The stability and activity of supported metal catalysts can be significantly influenced by interactions between the metal and the support[1,2]. For single metal atom catalysts supported on metal-oxide surfaces, support interactions impact both the metal's catalytic properties and its resistance to sintering[3]. The nature of active sites exposed on the metal surface can be affected by the size, shape and dispersion of the metal clusters[1–3]. Therefore, understanding metal–support interactions is essential for tuning the activity, selectivity and stability of oxide-supported metal catalysts. Despite the large number of studies investigating metal–support interactions[1,2,4–10], trends that predict interaction strengths between metal–support pairs are relatively unexplored.

Sintering compromises the reactivity of a catalyst by reducing the available catalytic surface area[7–9]. Typically governed by Ostwald ripening[11], sintering of metal catalyst particles has been shown to decrease the reactivity for a number of important reactions, including the water-gas shift reaction, methane oxidation and the selective hydrogenation of nitroarenes[12–14]. However, single atoms anchored to supports can exhibit strong resistance to sintering, and make for longer-lasting, highly active catalysts[8]. The degree to which a metal is anchored to a support is governed by the metal atom's binding energy—metals that exhibit strong exothermic binding to a support are less likely to diffuse across the support and agglomerate[9].

Due to their influence over catalytic performance, strong metal–support interactions (SMSIs) have been extensively studied[15,16]. Recently, SMSIs have taken on a new meaning[4,7,17,18], with numerous studies investigating metal/oxide adsorption energies as a method for improving catalytic performance[4,5,7–9,18–21]. The activity and selectivity of many metal/oxide systems are governed by effects linked to SMSIs[1,2]. Metal/oxide systems that can stabilize single-atom catalysts (SACs) have proven to be effective in many applications[22,23]. The well-defined active site of a SAC system can also provide high selectivity if one can achieve well-controlled uniformity across the catalyst surface[6]. As such, the strength of metal interaction on an oxide support can significantly impact the overall stability and activity of the SAC.

Several notable studies have investigated SMSIs using both experimental and computational techniques, most examining the interactions between one particular metal–oxide pair[19], and some studying trends across different metals[5,8,20,21] and supports. Understanding such trends is critical for screening combinations of metals and supports that can work in conjunction to produce sinter-resistant catalysts. Considering their well-defined morphology as supported metal catalysts, SACs frequently serve as useful experimental model systems for studying SMSIs, even if the technologically relevant catalysts feature metal clusters as opposed to single atoms[24]. Density functional theory (DFT) has also proven to be a powerful tool for investigating the energetic properties of metal/oxide systems, and has complemented experimental efforts in many studies[4,19,20].

Campbell et al.[8] demonstrated that a metal adatom's oxide formation enthalpy can be used to effectively predict a metal's adsorption energy to an oxide support—an approach that has been further validated in recent experimental and computational work[20]. They also suggested some correlation between a metal atom's adsorption energy and a support's oxide reduction energy[10,25].

Herein, we expand these correlations for a range of early and late transition metals and oxide supports. We employ DFT to investigate the properties of single adatoms and oxide supports that influence SMSIs, and we investigate periodic trends and electronic interactions of metal/oxide systems. Finally, we analyse different analytical forms of these properties using statistical learning based on least absolute shrinkage and selection operator (LASSO)[26] regression, which is employed to identify physical descriptors that significantly influence metal/oxide binding energy. Based on this analysis, we propose predictor equations for metal/oxide binding energy by evaluating error predictions characteristic of all possible subsets of various analytical forms of descriptor combinations.

[1]Department of Chemical Engineering, The Pennsylvania State University, University Park, PA, USA. [2]Department of Mechanical and Nuclear Engineering, The Pennsylvania State University, University Park, PA, USA. [3]Department of Chemical and Biomolecular Engineering, Rice University, Houston, TX, USA. [4]These authors contributed equally: Nolan J. O'Connor, A. S. M. Jonayat. *e-mail: mjanik@engr.psu.edu; tsenftle@rice.edu
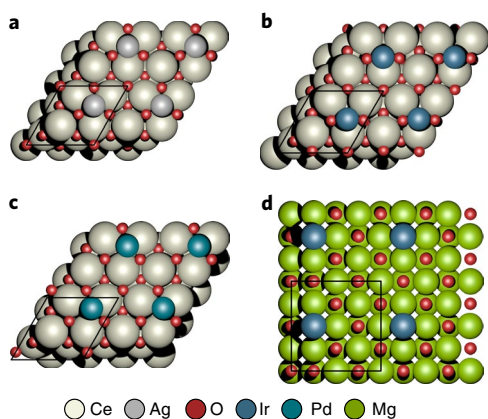
**Fig. 1 | Adsorption geometries of metals on CeO₂(111) and MgO(100).** **a–c**, Top view of adsorbed late transition metals on the CeO₂(111) surface at the threefold hollow site (Ag) (**a**), twofold oxygen bridge site (Ir) (**b**) and oxygen side-bridge site (Pd) (**c**). **d**, Adsorbed metals on MgO(100) all prefer the anionic oxygen site. The black rectangle represents the unit cell used in the study.



**Fig. 2 | Correlation between metal/support adsorption energies and metal adatom's oxide formation enthalpy.** Adsorption energies of transition metals on MgO(100), CeO₂(111), CeO₂(110), TiO₂(011), ZnO(100), TbO₂(111) and α-Al₂O₃(0001) plotted against the formation enthalpy of the metal adatom's most stable oxide. TbO₂(111) adsorption energies are plotted on the right (blue) y axis while the remaining data are plotted on the left y axis (due to scale difference). Vertical grey lines are label guides. Dashed lines represent the best linear fit to the data for each support, with fit equations and quality given in Supplementary Table 2.

## Results

**Trends in strong metal–support adsorption energies.** A range of transition metals (Cu, Ag, Au, Ni, Pd, Pt, Co, Rh, Ir, Fe, Ru, Mn and V) were adsorbed on several reducible and irreducible oxide supports (CeO₂(111), MgO(100), CeO₂(110), TbO₂(111), ZnO(100), TiO₂(011) and α-Al₂O₃(0001) surfaces) to investigate trends in metal adsorption energy among metals and supports. On each support surface, a single metal atom was adsorbed to several high-symmetry sites corresponding to energetically favourable configurations. For example, silver favourably adsorbs at a threefold hollow site (Fig. 1a), consistent with previous DFT studies[5], while iridium favourably adsorbs to CeO₂(111) at an oxygen bridge location (Fig. 1b). Conversely, all atoms adsorb above the anionic oxygen on the MgO(100) surface, which is a well-documented property of this surface[27]. The full range of adsorption sites and energies for each metal atom on all oxide surfaces are reported in Supplementary Table 1, and adsorbed configurations are shown in Supplementary Figs. 1–13. The adsorption sites for all previously studied metal/oxide systems were validated against the existing literature where possible[21,28–30]. Specifications of DFT method and calculation procedures for adsorption energies ($\Delta E_{ads}$) are provided in the Methods.

Campbell and Sellers[8] first proposed the correlation between the binding energy of metal nanoparticles on oxide supports and the oxide formation enthalpy ($\Delta H_{f,ox}$) of the metal adatoms. This descriptor determines the interaction strength between a gas-phase metal atom and oxygen, and is determined from experimental reference data, as explained in the Methods. This oxide formation enthalpy reflects the metal adatom's affinity for oxygen and intuitively correlates with the metal's interaction with oxide surfaces. Previous work by Hosokawa et al.[31] has confirmed the presence of metal–oxygen–metal bonds formed during single-atom adsorption on oxides, demonstrating the importance of metal–oxygen interactions that dictate metal adsorption. However, this descriptor is not independently sufficient to predict metal atom adsorption energy, in part due to possible metal–metal interactions between the adatom metal and the support metal (see below).

In Fig. 2, the adsorption energy of a metal onto an oxide surface is plotted against the adsorbed metal's oxide formation enthalpy. A linear trend for each support suggests that the metal adatom's interactions with surface oxygen atoms have a considerable effect on the strength of its interfacial bond to the oxide. Metals with a weaker affinity for oxygen, such as silver, bind to any oxide significantly weaker than those with high oxygen affinity, such as
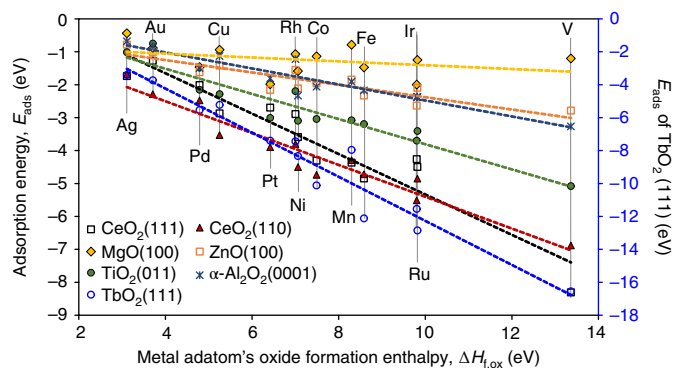
vanadium. Thus, the oxide formation enthalpy of the metal is an effective predictor of that metal's relative binding strength to an oxide support. The slopes, y intercepts, $R^2$ values (coefficient of determination) and mean absolute errors corresponding to each oxide surface are reported in Supplementary Table 2. The $R^2$ values confirm a reasonable degree of correlation for each support, except for MgO(100), which experiences a poor correlation because it binds all metals so weakly that oxidation enthalpy becomes a poor descriptor. The mean absolute error values range from 0.20 to 0.68 eV and are higher for supports that bind metals more strongly, such as TbO₂(111).

The trend of metal adsorption strength on each support surface is described by a unique slope, suggesting that characteristics of the support itself also play a role in determining the overall metal binding strength. Indeed, properties of oxide supports, such as nanocrystal shape and exposed facet, have been shown to strongly influence the catalytic activity of oxide-supported catalysts[1,32]. In particular, the reducibility of a support has been shown to affect the strength with which it adsorbs metal atoms[10,33]. Support reducibility is quantified by the oxygen vacancy formation energy ($\Delta E_{vac}$) of the oxide (see Methods for the computation of $\Delta E_{vac}$). Typically, reducible supports with lower endothermic oxygen vacancy formation energies are formed from parent metals with unoccupied low-energy states that can readily accept transferred electrons following the formation of an oxygen vacancy. The ability of the surface to accept donated electrons influences the adsorption of metal atoms, which in many cases have been shown to oxidize upon adsorption and, accordingly, reduce the metal atoms in the support itself[34].

In Fig. 3, the slope of each surface's trend line in Fig. 2 is correlated with the surface oxygen vacancy formation energies of each support. This reflects a general trend in which reducible supports (that is, supports with lower endothermic vacancy formation energies) tend to have a steep slope featuring more variation in metal atom binding across transition metals. The linear correlation in Fig. 2 over a large range of oxygen vacancy formation energies indicates that the reducibility of each surface is a viable indicator that each support's reducibility reflects its capacity to strongly adsorb metal atoms. MgO(100), an irreducible surface, does not interact strongly with silver or vanadium, which are the atoms with the lowest and highest oxide formation enthalpy, respectively. Thus, the slope of MgO(100) is much shallower than that of TbO₂(111), the most
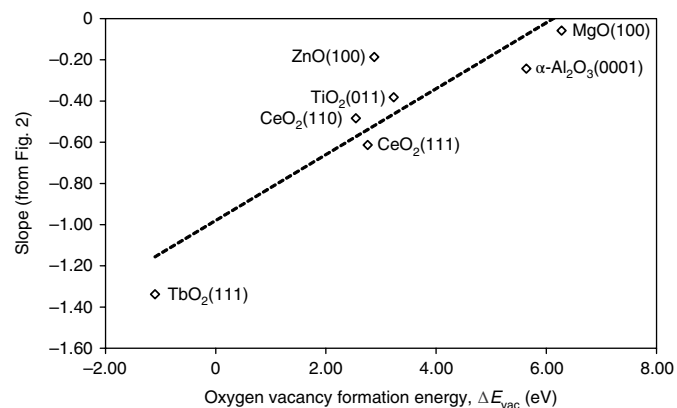
**Fig. 3 | Correlation between the slopes from Fig. 2 and the support's oxygen vacancy formation energy.** Slopes from the lines of best fit for adsorption energy versus metal adatom's oxide formation enthalpy for each surface are plotted against the surface's oxygen vacancy formation energy. The dashed line represents the best linear fit for all data.



**Fig. 4 | Correlation between metal/support adsorption energies and the support's oxygen vacancy formation energy.** Adsorption energies of silver, iridium and palladium atoms to $TbO_2(111)$, $\alpha$-$Al_2O_3(0001)$, $CeO_2(111)$, $CeO_2(110)$, $TiO_2(011)$, $ZnO(100)$ and $MgO(100)$ supports plotted against the corresponding oxygen vacancy formation energies of the support. Vertical grey lines are label guides and the dashed lines represent the best linear fit of the data.

reducible surface, which strongly adsorbs the metal atoms that readily oxidize, such as iridium, ruthenium and vanadium. The notion that the surface's reducibility measures its capacity to strongly adsorb metal atoms also explains the larger variations in vanadium adsorption energies compared with silver adsorption energies.

If surface reducibility, as measured by the oxygen vacancy formation energy ($\Delta E_{vac}$), serves as a reasonable descriptor for the varying slopes in Fig. 2, the oxygen vacancy formation energy in a support should correlate with the adsorption energy of a specific metal. In Fig. 4, the surface adsorption energies of silver, iridium and palladium atoms are plotted against each oxide support's reducibility. As shown in Fig. 4, a linear trend emerges, this time for each metal across varying supports. These linear trends indicate a relationship between the $\Delta E_{vac}$ of a support and the strength to which that support can bind a metal atom. Whereas Fig. 2 shows the characteristics of the metal adatom that influence strong adsorption energies to an oxide support, Fig. 4 shows the characteristics of the oxide support that influence its ability to strongly bind metal atoms. That is, metal atoms with a higher affinity for oxygen bind strongly to oxide supports, and supports that readily release oxygen bind metals more strongly. The equations and mean average errors for the best-fit lines shown in Fig. 4 are reported in Supplementary Table 3.

The adsorption energy and oxygen vacancy formation energy correlations obtained in this study are stronger than those recently reported by Hemmingson and Campbell[10] when attempting to compare experimental adsorption energies with computationally derived oxygen vacancy formation energies. It is likely that the lack of an apparent trend between experimental metal adsorption energies and computational oxygen vacancy energies results from the difficulty in ensuring that the DFT data and experimental data are properly normalized with respect to metal coverage and oxygen vacancy concentration on the surface. The concentration of surface oxygen vacancies is known to significantly influence the oxygen vacancy formation energy[35]. The oxygen vacancy formation energy is highly dependent on the size and shape of the simulation cell, as well as the local configuration of vacancies, making it difficult to cross-compare the experimental and computational data. In this study, the oxygen vacancy formation energy and metal adsorption energy are always computed with simulation cells that are the same size, thus enforcing the proper normalization between metal adatom coverage and oxygen vacancy concentration. The oxygen vacancy and metal adsorption data are therefore directly comparable, leading to the strong correlations seen in Figs. 2 and 4.
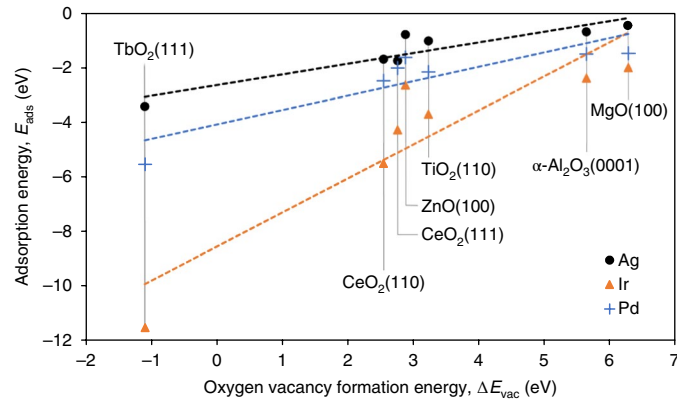
While the trends between adsorption energy and oxygen vacancy formation energy hold for supports over a wide range of surface reducibilities, they do not accurately describe the adsorption energy differences between surfaces with similar oxygen vacancy formation energies (for example, $CeO_2(110)$, $CeO_2(111)$ and $TiO_2(011)$ surfaces). As discussed below, metal–metal binding can also contribute to differences in metal atom adsorption energies and can partially explain differences among supports. Specific binding geometries, as well as surface reconstruction and relaxation, will inevitably contribute noise to the presented correlations. It is also important to consider that DFT has well-known deficiencies when describing highly correlated rare-earth oxides such as $CeO_2$ and $TbO_2$, even with the inclusion of the U correction. Despite such deficiencies, we have included analyses of these systems due to their industrial relevance and importance to the catalysis community—particularly $CeO_2$. Within the context of this study, inaccuracies inherent to DFT probably result in total formation energies and oxygen vacancy formation energies that are overestimated in magnitude on the $CeO_2$ and $TbO_2$ facets considered here. However, we anticipate that the overall trends are still meaningful, as they agree with trends related to oxides for which DFT and DFT + U are known to perform better for these types of analyses.

**Electronic structure analysis.** This section provides electronic structure analyses of the trends identified in the previous section, which serve to elucidate the electronic mechanism behind SMSIs. We focus on two metal atoms and two oxide supports: silver and iridium (representing metals with low and high oxide formation enthalpies, respectively) and $CeO_2(111)$ and $MgO(100)$ surfaces (representing low and high oxygen vacancy formation energies, respectively). These metal atoms and oxide supports are also technologically relevant in many industrial applications, and have been the subject of many studies[4,5,21,27].

The binding energies of silver to $CeO_2(111)$ and $MgO(100)$ are −1.74 and −0.44 eV, respectively. For iridium, the binding energies are −4.26 and −2.00 eV, respectively. These adsorption energy values follow the trends described in the previous section: silver binds weakly to each surface because it has a low affinity for oxygen, and $MgO(100)$ binds both atoms weakly because its oxygen vacancy formation enthalpy is endothermic. The density of states (DOS) plots in Fig. 5 demonstrate clear differences in the adsorption of silver and iridium on the $CeO_2(111)$ surface. The many overlapping
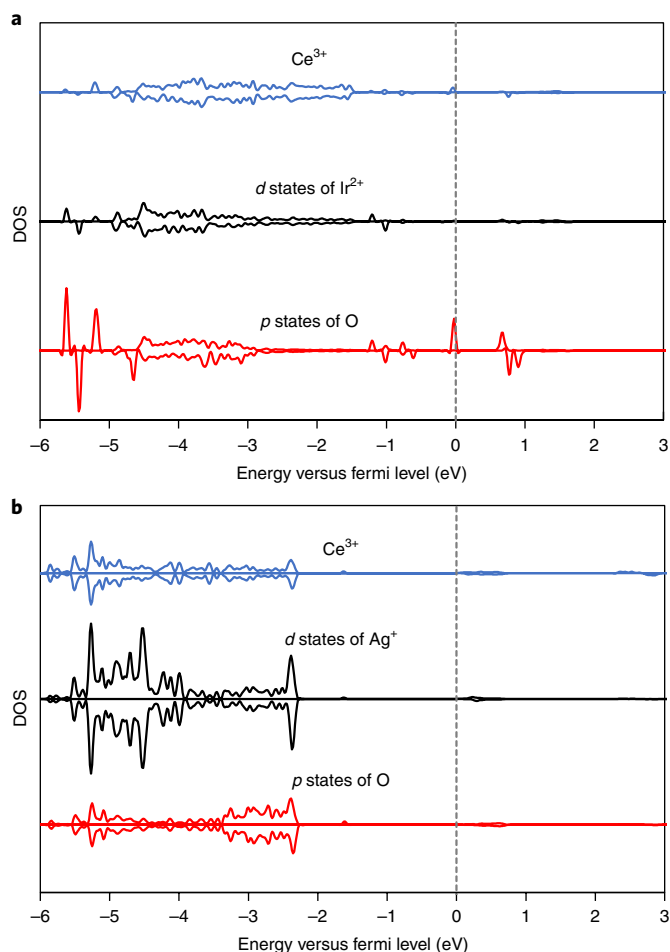
**Fig. 5 | DOS plots of Ir/CeO₂(111) and Ag/CeO₂(111). a,b,** Projected DOS of metal adsorption on $CeO_2(111)$ showing the $d$ states of $Ir^{2+}$, all states of $Ce^{3+}$ and $p$ states of O in the Ir–O–Ce bond on the surface (**a**) and the $d$ states of $Ag^+$, all states of $Ce^{3+}$ and $p$ states of O in the Ag–O–Ce bond on the surface (**b**). The vertical dashed line is the Fermi level.
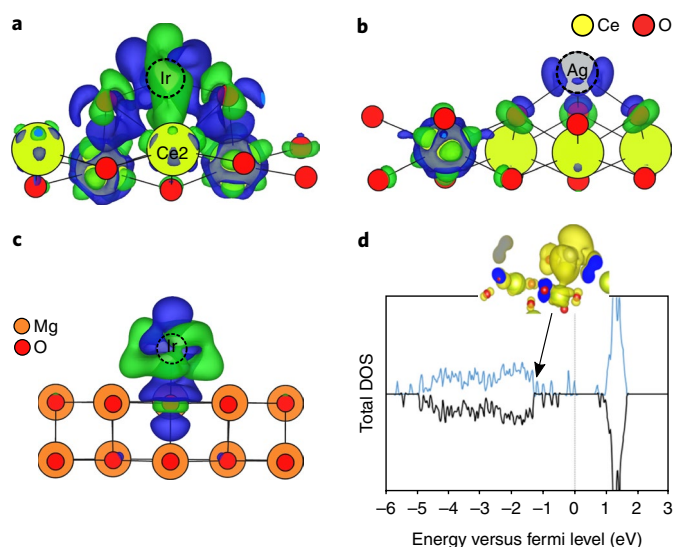


**Fig. 6 | Metal–support interactions on CeO₂(111) and MgO(100) supports. a–c,** Isostructural charge density difference plots of Ir/CeO₂(111) (**a**), Ag/CeO₂(111) (**b**) and Ir/MgO(100) (**c**). Blue denotes depletion of electron density while green represents accumulation. The isosurface level is ±0.005 e Bohr⁻³. **d,** DOS plot of Ir/CeO₂(111) and an orbital density image of the peak, indicated with an arrow.

different metal adatoms and oxide supports, which is adequate for determining physically relevant trends.

While the reduction of surface cerium atoms in both Ir/CeO₂(111) and Ag/CeO₂(111) may be interpreted as stemming from the metal atoms' interaction with oxygen atoms, the adsorbed metal does, in both cases, interact with surface cerium atoms directly. Examination of the charge density difference plot of Ir/CeO₂(111) in Fig. 6a reveals the presence of metal–metal binding between the iridium and cerium atoms (labelled Ce2). The green band linking the two atoms indicates a charge accumulation in a metal–metal hybrid orbital, which corresponds to the peak just

states between $Ce^{3+}$, $O^{2-}$ and $Ir^{6+}$ below the Fermi level in Fig. 5a suggest hybridization between iridium, cerium and oxygen orbitals, reflecting the formation of surface bonds responsible for the strong adsorption of iridium on $CeO_2(111)$. Conversely, the overlap in Fig. 5b of only one state between $Ce^{3+}$, $O^{2-}$ and $Ag^{6+}$ below the Fermi level demonstrates weaker silver adsorption to the $CeO_2(111)$ surface.

A visual representation of charge transfer upon metal adsorption on $CeO_2(111)$ is shown in the isostructural charge density difference plots presented in Fig. 6a,b. Figure 6a demonstrates that iridium adsorption on $CeO_2(111)$ causes significant charge transfer between the iridium adatom and the surface. Charge initially surrounding the iridium adatom redistributes towards neighbouring oxygen and cerium atoms, where two cerium atoms are reduced by the formation of an $Ir^{2+}$ state. This reduction is evident in the charge density difference's distinct resemblance to $f$ orbitals localized on adjacent cerium atoms, and by the Bader charge differences and site-projected magnetic moments reported in Table 1. Cerium atoms with a total magnetic moment approaching one represent $Ce^{3+}$ ions. Indeed, the formation of interfacial $Ce^{3+}$ ions has been reported in several computational and experimental studies[5,34,36]. In Fig. 6b, silver exchanges significantly less charge with $CeO_2(111)$, reducing only one cerium atom forming an $Ag^+$ state. Bader charges are used here to qualitatively compare charge transfer between

**Table 1 | Bader charge differences and total magnetic moments after iridium and silver adsorption on CeO₂(111) and MgO(100) surfaces**

| | Silver adsorption | | Iridium adsorption | |
|---|---|---|---|---|
| | Bader charge difference (e⁻) | Total magnetic moment ($\mu_B$) | Bader charge difference (e⁻) | Total magnetic moment ($\mu_B$) |
| **CeO₂(111)** | | | | |
| Ce1 | 0.25 | 0.84 | 0.07 | 0.02 |
| Ce2 | 0.06 | 0 | 0.28 | 0.96 |
| Ce3 | 0.06 | 0 | 0.13 | 0.02 |
| Ce4 | 0.06 | 0 | 0.26 | 0.93 |
| O1 | 0.02 | 0 | −0.09 | −0.14 |
| O2 | 0.01 | 0 | −0.1 | −0.14 |
| O3 | 0.01 | 0 | 0 | −0.01 |
| Metal | −0.55 | 0 | −0.84 | −0.65 |
| **MgO(100)** | | | | |
| O1 | −0.16 | 0.12 | −0.34 | −0.01 |
| Metal | 0.15 | 0.72 | 0.39 | 0.94 |

The atom labels correspond to the labels shown in Supplementary Figs. 1–13. A positive difference indicates the accumulation of negative charge. $\mu_B$, Bohr magneton units.

above the Ce–O band on the Ir/CeO$_2$(111) DOS plot in Fig. 6d, as indicated by the image of charge density associated with that peak (also showing a band of electron density linking the iridium and cerium atoms). Metal–metal binding between metal adatoms and metal oxides has previously been reported[15,20], but has never been considered influential in SMSIs. This possibility is examined in the following sections.

Electronic structure analyses can also explain the energy differences between CeO$_2$(111) and MgO(100) in the binding of metal adatoms. Since MgO(100) is irreducible, it is unable to readily accept electron density donated by a metal adatom and instead reduces the metal adatom, as indicated by a positive Bader charge difference on the metal adatom (Table 1). Whereas the adsorption of iridium onto CeO$_2$ involves the interaction between neighbouring oxygen and cerium atoms, the metal–support interactions from the adsorption of iridium onto MgO(100) are localized onto one surface oxygen atom. This transfer of electron density from the support to the adsorbed iridium atom is evident in the isostructural charge density difference of Ir/MgO(100) shown in Fig. 6c.

**Physical descriptors for the prediction of binding energy.** Our DFT results have shown correlations between the binding energy of a metal adatom, its oxide formation enthalpy and the support's oxygen vacancy formation energy. Metal–metal binding is also evident from DOS and Bader charge analyses. Although the above trends suggest linear correlations, these primary descriptors can be used (and combined) in various other functional forms to predict the adatom binding energy. We therefore performed a systematic analysis to identify the most important analytical form for each descriptor set. In statistical learning, various shrinkage methods, such as ridge regression, LASSO and least-angle regression[26], can be used to identify within a feature space a subset of descriptors that minimize deviation between the predicted and actual values. These methods are also used in machine learning applications.

Our objective is to determine the descriptors that yield the best prediction for the binding energy, considering those suggested by the DFT analysis in the previous sections as well as other available descriptors of structural and electronic properties. This problem is readily addressed using statistical learning based on compressed sensing[37]. Here, we adopt a systematic analysis method based on compressed sensing described by Ghiringhelli et al.[38,39], where a feature space is generated by combining primary descriptors using analytical formulas. LASSO is then used to identify the most important descriptors via equation (1),

$$\underset{c \in R^M}{\operatorname{argmin}} \sum_{j=1}^{N} \left( P_j - \sum_{k=1}^{M} d_{j,k} c_k \right)^2 + \lambda \sum_{k=1}^{M} |c_k| \qquad (1)$$

where $P$ is a column vector of responses (that is, DFT binding energies) that has been centered, $d$ is a ($N \times M$) dimensional design matrix of descriptors and $c$ is the column vector of coefficients that is to be determined, $j$ and $k$ are indices, and $R$ represents the set of real numbers. $\lambda$ is a penalty parameter that, when increased, decreases the number of non-zero components of vector $c$. For a maximum value of $\lambda$, all elements of $c$ will be zero, and as $\lambda$ is decreased more elements will take non-zero values. Since LASSO is not scale invariant, we standardized[26] the descriptor matrix, $d$, to have zero-mean and unit-variance by subtracting from each column its mean and normalizing by its s.d. This normalization renders all descriptors in the common scale and makes the penalty parameter, $\lambda$, meaningful. After enforcing a cutoff value of $\lambda$, a subsequent exhaustive search using $l_2$ norm minimization over all possible subsets of descriptors with non-zero $c$ values is completed. Due to the linear correlation between descriptors, the best results

may not be achieved solely by employing LASSO. Details of this composite method (LASSO$+l_o$) can be found elsewhere[38,39], and examples of the application of LASSO for material-science problems can be found in refs [40,41].

*Feature space.* We completed two analyses with varying numbers of primary descriptors. For the first analysis, we used a minimal feature space based on the DFT analysis presented above. This first feature space included only the oxide formation enthalpy of the metal adatom ($\Delta H_{f,ox} = \Delta H_{sub} - \Delta H_{f,ox,bulk}$)—where $\Delta H_{sub}$ is the heat of sublimation and $\Delta H_{f,ox,bulk}$ is the oxidation energy of the bulk metal—and the oxygen vacancy energy of the oxide support ($\Delta E_{vac}$) as the two primary descriptors. The secondary descriptors included ratios, differences, summations and squares of the primary descriptors, as well as absolute differences and squares of absolute differences. This generated a total of ten descriptors in the first feature space (Supplementary Table 4).

For the second analysis, the primary feature space was expanded to also include atomic properties of the adatom and support. The chosen properties have been proposed in the literature to have some correlation to metal–metal and metal–oxide interactions and have been previously used in machine learning approaches[39–44]. Atomic properties of the metal adatom (m) and support (s) include the electronegativity of the metal in Pauling and Martynov–Batsanov scales, $(n-1)^{th}$ and $n^{th}$ ionization energies (IE$_n$) with the bulk metal in the $n^+$ oxidation state, electron affinity, highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) of single metal atoms relative to vacuum calculated with DFT, $s$ and $p$ orbital radii[2,45] (Zunger[46] and Weber–Cremer[47]), number of valance electrons and atomic number. Two parameters derived from a previously reported semi-empirical method for predicting metal–metal binding enthalpies were also included[48]. These two parameters, first introduced by Miedema et al.[48], represent the chemical potential of the electrons in the metal ($\phi$), and were combined with parameters that represent the discontinuity in electron density between the two binding metals ($\eta^{1/3}$). $\Delta H_{sub}$ and $\Delta H_{f,ox,bulk}$ of each metal adatom were also included as individual descriptors. For oxide surface properties, the descriptor set includes $\Delta E_{vac}$ and the workfunction and surface energy ($\gamma$) of the support. The coordination number of the metal and oxygen in the support, and bond valance[49] of the metal in the support, were also included. Values used for these primary descriptors, calculation methods and data sources are provided in the Supplementary Information.

The secondary descriptors for this feature space are populated with the absolute value of ratios, differences, summations and multiplications of primary descriptors, followed by taking the inverse, square and square root of all of the generated descriptors, where we only allow analytical functions (summation and subtraction) between descriptors with consistent units. Since the number of features grows rapidly during this process, in some cases, we used physical intuition to remove unphysical (for example, (HOMO$^m$ + LUMO$^s$)) and unimportant (for example, (EA$^m$ + EA$^s$)) analytical forms. In the final step, all primary and secondary descriptors are combined, and another set of descriptors is generated by multiplying each descriptor of this set with the remaining descriptors (avoiding repetition). The combination of these two sets yields a final feature space with 333,932 descriptors. The full list of analytical formations applied to generate the secondary feature space is provided in the Supplementary Information. While a simple least-squares regression is adequate for screening out the best of the ten descriptors in the first feature space, it is computationally intractable for screening the second feature space. Therefore, application of LASSO$+l_o$ is necessary with such a large feature set.

**LASSO$+l_o$ analysis.** The first feature space contains no metal–metal binding parameters. The penalty term, $\lambda$, is logarithmically
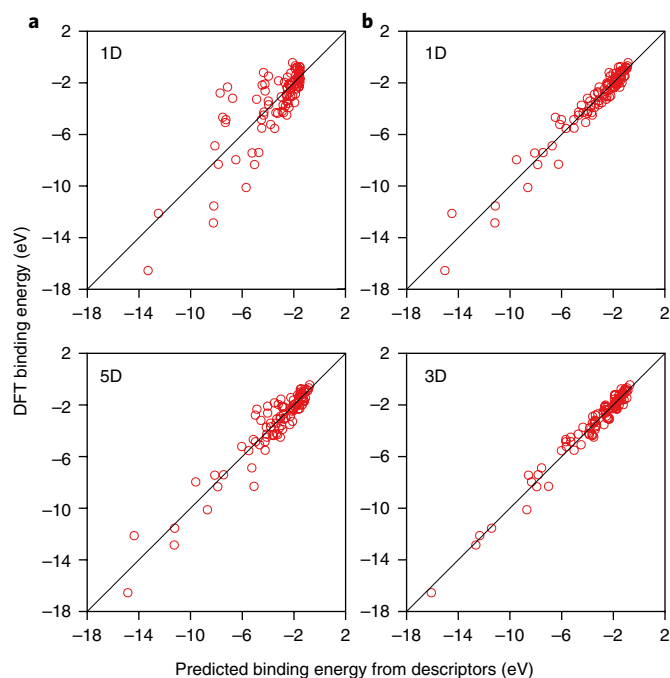
**Fig. 7 | Comparison between descriptor-predicted and DFT binding energies. a**, Feature space 1 for 1D (top) and 5D (bottom). **b**, Feature space 2 for 1D (top) and 3D (bottom).

**Table 2 | Equations for binding energy prediction based on LASSO + $l_o$ analysis using feature space 2**

| | Binding energy predicting equation | RMSE (eV) |
|---|---|---|
| 1D | $-0.1507 \times \left( CN_{bulk}^m \times \left\| \frac{\Delta H_{f,ox}}{\Delta E_{vac}} \right\| \right) - 0.3962$ | 0.6873 |
| 2D | $-0.4839 \times \left( \sqrt{EA^m} \times \left\| \frac{\Delta H_{f,ox}}{\Delta E_{vac}} \right\| \right) - 1.1756$ $\times \left( \left\| \frac{L^s}{L^m} \right\| \times \left\| \frac{\Delta H_{f,ox}}{\Delta E_{vac}} \right\| \right) - 0.0770$ | 0.5586 |
| 3D | $-0.3206 \times \left( \sqrt{EA^m} \times \left\| \frac{\Delta H_{f,ox}}{\Delta E_{vac}} \right\| \right) - 1.0850$ $\times \left( \left\| \frac{L^s}{L^m} \right\| \times \left\| \frac{\Delta H_{f,ox}}{\Delta E_{vac}} \right\| \right)$ $-2.1228 \times \left( \left\| \frac{IE_n^m - IE_n^s}{IE_n^m} \right\| \times \left( \Delta \eta^{\frac{1}{3}} \right)^2 \right) - 0.3279$ | 0.5028 |

$CN_{bulk}^m$ is the coordination number of the support metal in the bulk phase. $\Delta H_{f,ox}$, $\Delta E_{vac}$, electron affinity (EA), LUMO, $L$ and ionization energy (IE) are in eV, and $\Delta \eta^{1/3}$ is in (density unit)$^{1/3}$ (ref. [48]).

decreased from $\lambda_{max} = 2.363$ to $\lambda_{min} = 0.001 \times \lambda_{max}$ following the procedure described in ref. [39] in which all $c_k$ values are zero at $\lambda_{max} = \frac{1}{N} \max_j |d_j, P\rangle |$ [50]. The descriptors with non-zero coefficients in sequence with decreasing $\lambda$ are: (1) $(\Delta H_{f,ox} - \Delta E_{vac})^2$; (2) $\Delta E_{vac}$; (3) $\Delta H_{f,ox}/\Delta E_{vac}$; (4) $\Delta H_{f,ox} - \Delta E_{vac}$; and (5) $\Delta E_{vac}^2$, where the employed cutoff is $\lambda = 0.0044$. To check the consistency of these descriptors, we conducted several trials in which we randomly left out 10% of the DFT-predicted binding energies from the analysis. The same descriptors appeared in the same sequence for all trials. Next, an exhaustive search using $l_2$ norm minimization with all possible combinations of descriptors with one (1D) to five (5D) elements was performed with all data points. The descriptor combinations with the lowest root mean square error (RMSE) for 1D to 5D are 1.65, 1.16, 1.03, 0.96 and 0.91 eV, respectively. The predictor equations are given in Supplementary Table 5. Figure 7a demonstrates that, intuitively, the prediction improves with increasing numbers of descriptors, where five descriptors (5D) yields the lowest RMSE. Even at 5D, the RMSE remains high (0.91 eV), suggesting that our descriptor set is incomplete. With such a small feature space, we could have reached the same predictor equations using direct $l_2$ norm minimization over the ten descriptors, but have detailed the complete procedure used here as it was repeated for the larger second feature space.

For the second feature space, all 333,932 descriptors were screened using LASSO, with $\lambda$ logarithmically decreased from $\lambda_{max} = 2.7921$ to $\lambda_{min} = 0.001 \times \lambda_{max}$. A total of 75 descriptors were identified with the cutoff value reduced to $\lambda = 0.0602$. An exhaustive search using $l_2$ norm minimization yields 1D to 5D sets that have significantly lower RMSE values compared with feature space 1. The 1D descriptor set of the second feature space outperforms the 5D descriptor set of the first feature space (Fig. 7b). Table 2 reports the 1D–3D descriptor equations with RMSE, and Fig. 7b shows the comparison between predicted and DFT binding energies. Here, all data points were used for training. The 4D and 5D descriptors are reported in Supplementary Table 6. The optimal 5D descriptor set reduces the

RMSE to 0.4096 eV. To check the robustness of these descriptors, we left out 10% of the data (randomly selected) and used the remaining data for our LASSO + $l_o$ analysis. We repeated the procedure 50 times, standardizing the remaining 90% of the training data each time before the LASSO step. This confirmed that no information from the test set was transferred to the randomly selected training data. For 1D and 2D, the same descriptors were predicted in 74 and 68% of the trials, respectively. For 3D–5D, the variations in the selected descriptor set were larger. However, similar RMSE values for different trials indicate that in higher dimensions there are more combinations of similar primary descriptors that can equally describe the randomly selected training data. For example, when all data points are used for training, the RMSE varies by 0.03 eV within the top five 1D descriptors, and for 3D it only varies by 0.006 eV. For the reported coefficients in Table 2, we used all available data points and did not keep a separate set of data for testing due to the limited availability of data (91 points). However, for 50 repeated 'leave 10% out' analyses, the average RMSEs when applying the 1D and 2D descriptors to predict the test set with the left-out data were $0.6969 \pm 0.2473$ eV and $0.5948 \pm 0.1632$ eV, respectively, which supports their applicability to external data points.

In agreement with our DFT results, the top five 1D descriptors contain both the metal adatom's oxidation enthalpy and the support's oxygen vacancy formation energy (Supplementary Table 7). The ratio of these values (that is, $|\Delta H_{f,ox}/\Delta E_{vac}|$) always appears in the 1D descriptor multiplied by a second term. Metal atoms bind more strongly to surfaces when they have increased oxide formation enthalpies (that is, when they form more stable oxide) and bind less strongly to surfaces that have increased surface oxygen vacancy formation energies (that is, surfaces that are less reducible). Hence, $|\Delta H_{f,ox}/\Delta E_{vac}|$ is a descriptor that captures this trend. TbO$_2$(111) is the only surface with a negative oxygen vacancy formation energy ($-1.1$ eV). Although the absolute value of the ratio does not capture the negative $\Delta E_{vac}$, we are still able to model this surface since the oxygen vacancy formation energy of TbO$_2$(111) is much smaller than all other surfaces.

The lowest RMSE is obtained when the second term is the coordination number of the surface metal atom in the bulk phase ($CN_{bulk}^m$), suggesting again that the surface metal's ability to form bonds plays an important role in metal/support binding. A similar RMSE is obtained with this second term in the

1D descriptor replaced by the LUMO energy (or $\sqrt{\text{LUMO}}$) of the surface metal atom, or difference of $n^{\text{th}}$ ionization energy of the adatom and surface metal atom weighted by the $n^{\text{th}}$ ionization energy of the surface metal. A term involving $(p + s)$ orbital radius (Zunger) also gives similar RMSE. These descriptors are all highly correlated, as demonstrated by the calculated Pearson correlation coefficient (Supplementary Table 7). The additional term in the 1D descriptors is either a characteristic of the support metal atom or a combined property of both the adatom and the support metal, suggesting that properties related to metal–metal interactions are required to better predict adatom binding. For a 2D descriptor set, both elements of the top five descriptor pairs (Supplementary Table 8) contain the $|\Delta H_{\text{f,ox}}/\Delta E_{\text{vac}}|$ term. Among the top five descriptor pairs, at least one of the multiplication terms is a combined property of the adatom and surface metal atom. These terms involve their LUMO/HOMO ratio, electron affinity, ionization energy or multiplication of $(p + s)$ orbital radius terms. This indicates that the metal–metal interaction is of secondary importance compared with the metal–oxygen parameters, yet must be included. Similar terms are also present in the 3D descriptor set. The last descriptor includes $(\Delta \eta^{1/3})^2$, which is a known form of the Miedema parameters[48] that represents metal–metal interaction. In 4D and 5D, we find that surface properties are introduced, but our 'leave 10% out' analysis shows that the possible combinations with lowest RMSE will vary greatly. Hence, 1D and 2D descriptors are more robust.

Our LASSO + $l_0$ analysis agrees well with our DFT analyses, indicating that neither $\Delta H_{\text{f,ox}}$ nor $E_{\text{vac}}$ can alone predict the binding energy. We conclude that their ratio is a better predictor when multiplied by either a surface metal property or combined property of both the adatom and the surface metal atom. Higher accuracy in binding energy prediction is achieved when metal–metal interactions are captured in descriptor terms that include properties of both the adatom and the surface metal in the ≥2D descriptor sets.

## Conclusion

Adsorption energies obtained by DFT were reported for a range of early and late transition metals when adsorbed onto several reducible and irreducible oxide supports. These adsorption energies are correlated with properties of both the metal and the support; namely, the metal oxidation enthalpy, support reducibility and enthalpy of metal–metal interactions. To explain the trends in SMSIs, electronic structures for metal/oxide systems that exhibit both strong and weak interactions were analysed. Metal–metal interactions between adatom and support were observed for every metal/oxide system except for those supported on MgO(100). The parameters that influence strong metal–support interactions were used in conjunction with LASSO + $l_0$ to develop a predictive model for screening metal/support combinations that produce strongly adsorbed SACs.

## Methods

**DFT specifications.** Spin-polarized DFT calculations were employed to investigate the adsorption of single transition metal atoms on oxide supports. The calculations were performed using the Vienna Ab initio Simulation Package[51], using the Perdew–Wang (PW91)[52] version of the generalized gradient approximation as the exchange-correlation functional, with the projector augmented wave[53] approximation representing the atomic core regions. An atomic force convergence criterion of 0.05 eV Å$^{-1}$ was used to identify optimized geometries. Plane wave basis sets were truncated at a kinetic energy cutoff of 450 eV. Valences for each atom type are reported in Supplementary Table 9.

A range of non-reducible and reducible low-index oxide supports, including MgO(100), CeO$_2$(111), CeO$_2$(110), TbO$_2$(111), ZnO(100), rutile TiO$_2$(011) (referred to as TiO$_2$(011) in this paper) and α-Al$_2$O$_3$(0001), were studied in their stoichiometric states. Each surface facet was cleaved from an optimized bulk unit cell and then expanded such that each surface had at least four oxygen atoms in the outermost oxygen layer, except for α-Al$_2$O$_3$(0001), which had three oxygen atoms.

A surface normal lattice vector of 30 Å gave ample vacuum space between periodic slabs. The lattice constants for optimized surface structures and Monkhorst–Pack[54] $k$-point sampling used for each facet are reported in Supplementary Table 10. DFT fails to accurately represent localized $d$ and $f$ orbitals, and as such we employed the Dudarev DFT + U[55] formalism to treat the $f$ states of TbO$_2$ and CeO$_2$ and the $d$ states of TiO$_2$, using literature-derived U values of 6 eV[56], 5 eV[57] and 4.2 eV[19], respectively. DFT + U with the PW91 functional yields charge transfer and adsorption energy values comparable to the hybrid functional HSE06 for metal/oxide systems (for example, see ref. [58] and the references therein), verifying that it is appropriate for studying the metal/oxide systems investigated here. Various single metal atoms were adsorbed on the surface models, including both late transition metals (Cu, Ag, Au, Ni, Pd, Pt, Co, Rh and Ir) and early transition metals (Fe, Ru, Mn and V). All high-symmetry adsorption sites of each surface (for example, hollow, bridge and atop) were tested to identify the most favourable adsorption geometry for each metal–oxide pair. A dipole correction[59] was used to correct the unphysical interaction between dipoles in adjacent images along the surface normal for asymmetric slab models. Viable adsorption sites for CeO$_2$(111) and MgO(100), which were the primary focus in the later sections of this work, are shown in Fig. 1. Detailed descriptions of all adsorption sites can be found in Supplementary Table 1.

**Calculation of adsorption energies and correlation parameters.** Adsorption energies were calculated using equation (2), where $E_{\text{metal}-(\text{g})}$ is the total DFT energy of the metal atom in the gas phase, is the energy of the stoichiometric oxide support, and $E_{\text{support}}$ is the energy of the total metal/support system.

$$\Delta E_{\text{ads}} = E_{\text{metal/support}} - E_{\text{support}} - E_{\text{metal}-(\text{g})} \tag{2}$$

The reducibility of each oxide support, quantified by the support's surface oxygen vacancy formation energy, was investigated as a descriptor for the strength of the metal–support interaction. The oxygen vacancy formation energy was calculated using equation (3),

$$\Delta E_{\text{vac}} = E_{\text{M}_n\text{O}_{x-1}} + \frac{E_{\text{O}_2}}{2} - E_{\text{M}_n\text{O}_x} \tag{3}$$

where is the energy of the reduced support, $E_{\text{M}_n\text{O}_{x-1}}$ is the energy of gas-phase molecular oxygen and $E_{\text{M}_n\text{O}_x}$ is the energy of the stoichiometric support. Note that the vibrational energy and entropy of gas-phase O$_2$ are not included, as these corrections are uniform across all oxides and therefore do not impact relative trends between supports.

To predict the binding strength of single metal atoms, we use the metal atom's oxide formation enthalpy ($\Delta H_{\text{f,ox}}$), a descriptor first proposed by Campbell and Sellers[8]. This oxide formation enthalpy is determined from experimental reference data using the relation presented in equation (4),

$$\Delta H_{\text{f,ox}} = \Delta H_{\text{sub}} - \Delta H_{\text{f,ox,bulk}} \tag{4}$$

where $\Delta H_{\text{sub}}$ is the sublimation energy of the metal atom (that is, the experimentally determined cohesion energy of the bulk metal) and $\Delta H_{\text{f,ox,bulk}}$ is the formation enthalpy of the metal's most stable oxide relative to the bulk metal and O$_2$ (per metal atom). With the inclusion of $\Delta H_{\text{sub}}$, $\Delta H_{\text{f,ox}}$ is related to the energy of a single atom rather than the bulk metal phase. Experimental $\Delta H_{\text{sub}}$ and values used for each metal type are provided in Supplementary Table 11.

**Code availability.** The MATLAB code written to perform the LASSO + $l_0$ analysis is available under Massachusetts Institute of Technology licence in the following online repository: https://github.com/tsenftle/Metal-Oxide-LASSO-lo.git.

**Data availability.** The data that support the plots within this paper and other findings of this study are available from the corresponding author upon reasonable request. The following files are available in the Supplementary Information: adsorption sites and energies for each combination of metal adatom and oxide; cell configuration, size, number of surface layers, lattice constants, exchange-correlation functionals, $k$-points and U correction for each DFT-optimized surface; valence configurations for each metal adatom and oxide; oxide formation enthalpies for each metal adatom; slopes, $y$ intercepts, $R^2$ values and mean absolute errors for each trend in Figs. 2 and 4; values of primary descriptors; the procedure for feature space generation; and RMSEs of different descriptor combinations.

## References

1. Haruta, M. Catalysis of gold nanoparticles deposited on metal oxides. *CATTECH* **6**, 102–115 (2002).

2. Valden, M., Lai, X. & Goodman, D. W. Onset of catalytic activity of gold clusters on titania with the appearance of nonmetallic properties. *Science* **281**, 1647–1650 (1998).

3. Fiedorow, R. M. J., Chahar, B. S. & Wanke, S. E. The sintering of supported metal catalysts. *J. Catal.* **51**, 193–202 (1978).

4. Bruix, A. et al. A new type of strong metal–support interaction and the production of $H_2$ through the transformation of water on $Pt/CeO_2(111)$ and $Pt/CeO_x/TiO_2(110)$ catalysts. *J. Am. Chem. Soc.* **134**, 8968–8974 (2012).

5. Branda, M. M., Hernandez, N. C., Sanz, J. F. & Illas, F. Density functional theory study of the interaction of Cu, Ag, and Au atoms with the regular $CeO_2(111)$ surface. *J. Phys. Chem. C* **114**, 1934–1941 (2010).

6. Flytzani-Stephanopoulos, M. & Gates, B. C. Atomically dispersed supported metal catalysts. *Annu. Rev. Chem. Biomol. Eng.* **3**, 545–574 (2012).

7. Campbell, C. T. Catalyst–support interactions: electronic perturbations. *Nat. Chem.* **4**, 597–598 (2012).

8. Campbell, C. T. & Sellers, J. R. V. Anchored metal nanoparticles: effects of support and size on their energy, sintering resistance and reactivity. *Faraday Discuss.* **162**, 9–30 (2013).

9. Campbell, C. T. The energetics of supported metal nanoparticles: relationships to sintering rates and catalytic activity. *Acc. Chem. Res.* **46**, 1712–1719 (2013).

10. Hemmingson, S. L. & Campbell, C. T. Trends in adhesion energies of metal nanoparticles on oxide surfaces: understanding support effects in catalysis and nanotechnology. *ACS Nano.* **11**, 1196–1203 (2016).

11. Zhdanov, V. P. Ostwald ripening of charged supported metal nanoparticles: Schottky model. *Phys. E* **71**, 130–133 (2015).

12. Yati, I. Effects of sintering-resistance and large metal–support interface of alumina nanorod-stabilized Pt nanoparticle catalysts on the improved high temperature water gas shift reaction activity. *Catal. Commun.* **56**, 11–16 (2014).

13. Gholami, R., Alyani, M. & Smith, K. Deactivation of Pd catalysts by water during low temperature methane oxidation relevant to natural gas vehicle converters. *Catalysts* **5**, 561–594 (2015).

14. Liu, W. et al. Single-atom dispersed Co–Nl–C catalyst: structure identification and performance for hydrogenative coupling of nitroarenes. *Chem. Sci.* **7**, 5758–5764 (2016).

15. Tauster, S. J., Fung, S. C. & Garten, R. L. Strong metal–support interactions. Group 8 noble metals supported on titanium dioxide. *J. Am. Chem. Soc.* **100**, 170–175 (1978).

16. Maat, H. J. & Moscou, L. A study of the influence of platinum crystallite size on the selectivity of platinum reforming catalysts. In *Proc. 3rd Int. Cong. Catal.* 1277 (North Holland Publishing Company, Amsterdam, 1965).

17. Chandler, B. D. Strong metal–support interactions: an extra layer of complexity. *Nat. Chem.* **9**, 108–109 (2017).

18. Matsubu, J. C. et al. Adsorbate-mediated strong metal–support interactions in oxide-supported Rh catalysts. *Nat. Chem.* **9**, 120–127 (2017).

19. Addou, R. et al. Influence of hydroxyls on Pd atom mobility and clustering on rutile $TiO_2(011)$-2 x 1. *ACS Nano* **8**, 6321–6333 (2014).

20. Strayer, M. E. et al. Charge transfer stabilization of late transition metal oxide nanoparticles on a layered niobate support. *J. Am. Chem. Soc.* **137**, 16216–16224 (2015).

21. Lu, Z. S. & Yang, Z. X. Interfacial properties of $NM/CeO_2(111)$ (NM = noble metal atoms or clusters of Pd, Pt and Rh): a first principles study. *J. Phys. Condens. Matter* **22**, 10 (2010).

22. Rim, K. T. et al. Charging and chemical reactivity of gold nanoparticles and adatoms on the (111) surface of single-crystal magnetite: a scanning tunneling microscopy/spectroscopy study. *J. Phys. Chem. C* **113**, 10198–10205 (2009).

23. Abbet, S. et al. Acetylene cyclotrimerization on supported size-selected $Pd_n$ clusters ($1 \leq n \leq 30$): one atom is enough! *J. Am. Chem. Soc.* **122**, 3453–3457 (2000).

24. Chen, Y. Improved performance of supported single-atom catalysts via increased surface active sites. *Catal. Commun.* **75**, 74–77 (2016).

25. Campbell, C. T. & Mao, Z. Chemical potential of metal atoms in supported nanoparticles: dependence upon particle size and support. *ACS Catal.* **7**, 8460–8466 (2017).

26. Hastie, T., Tibshirani, R. & Friedman, J. *Linear Methods for Regression* (Springer, New York, 2009).

27. Giordano, L., Baistrocchi, M. & Pacchioni, G. Bonding of Pd, Ag, and Au atoms on MgO(100) surfaces and MgO/Mo(100) ultra-thin films: a comparative DFT study. *Phys. Rev. B* **72**, 11 (2005).

28. Hinnemann, B. & Carter, E. A. Adsorption of Al, O, Hf, Y, Pt, and S atoms on $\alpha$-$Al_2O_3$ (0001). *J. Phys. Chem. C* **111**, 7105–7126 (2007).

29. Melnikov, V. V., Yeremeev, S. V. & Kulkova, S. E. Theoretical investigations of 3d-metal adsorption on the $\alpha$-$AL_2O_3$ (0001) surface. *Russ. Phys. J.* **54**, 704–712 (2011).

30. Hernández, N. C., Graciani, J., Márquez, A. & Sanz, J. F. Cu, Ag and Au atoms deposited on the $\alpha$-$Al_2O_3(0001)$ surface: a comparative density functional study. *Surf. Sci.* **575**, 189–196 (2005).

31. Hosokawa, S., Taniguchi, M., Utani, K., Kanai, H. & Imamura, S. Affinity order among noble metals and $CeO_2$. *Appl. Catal. A* **289**, 115–120 (2005).

32. Si, R. & Flytzani-Stephanopoulos, M. Shape and crystal-plane effects of nanoscale ceria on the activity of Au-$CeO_2$ catalysts for the water-gas shift reaction. *Angew. Chem. Int. Ed.* **47**, 2884–2887 (2008).

33. Noronha, F. B., Fendley, E. C., Soares, R. R., Alvarez, W. E. & Resasco, D. E. Correlation between catalytic activity and support reducibility in the $CO_2$ reforming of methane over $Pt/Ce_xZr_{1-x}O_2$ catalysts. *Chem. Eng. J.* **82**, 21–31 (2001).

34. Szabova, L., Camellone, M. F., Huang, M., Matolin, V. & Fabris, S. Thermodynamic, electronic and structural properties of $Cu/CeO_2$ surfaces and interfaces from first-principles DFT + U calculations. *J. Chem. Phys.* **133**, 234705 (2010).

35. Murgida, G. E. & Ganduglia-Pirovano, M. V. Evidence for subsurface ordering of oxygen vacancies on the reduced $CeO_2(111)$ surface using density-functional and statistical calculations. *Phys. Rev. Lett.* **110**, 246101 (2013).

36. Artiglia, L. et al. Introducing time resolution to detect $Ce^{3+}$ catalytically active sites at the $Pt/CeO_2$ interface through ambient pressure X-ray photoelectron spectroscopy. *J. Phys. Chem. Lett.* **8**, 102–108 (2017).

37. Eldar, Y. C. & Kutyniok, G. *Compressed Sensing: Theory and Applications* (Cambridge Univ. Press, Cambridge, 2012).

38. 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.* **114**, 105503 (2015).

39. Ghiringhelli, L. M. et al. Learning physical descriptors for materials science by compressed sensing. *New J. Phys.* **19**, 023017 (2017).

40. Goldsmith, B. et al. Uncovering structure–property relationships of materials by subgroup discovery. *New J. Phys.* **19**, 013031 (2017).

41. Pilania, G. et al. Machine learning bandgaps of double perovskites. *Sci. Rep.* **6**, 19375 (2016).

42. Medasani, B. et al. Predicting defect behavior in B2 intermetallics by merging ab initio modeling and machine learning. *npj Comput. Mater.* **2**, 1 (2016).

43. Oliynyk, A. O., Adutwum, L. A., Harynuk, J. J. & Mar, A. Classifying crystal structures of binary compounds AB through cluster resolution feature selection and support vector machine analysis. *Chem. Mater.* **28**, 6672–6681 (2016).

44. Hong, W. T., Welsch, R. E. & Shao-Horn, Y. Descriptors of oxygen-evolution activity for oxides: a statistical evaluation. *J. Phys. Chem. C* **120**, 78–86 (2016).

45. John, J. & Bloch, A. N. Quantum-defect electronegativity scale for nontransition elements. *Phys. Rev. Lett.* **33**, 1095–1098 (1974).

46. Zunger, A. Systematization of the stable crystal structure of all AB-type binary compounds: a pseudopotential orbital-radii approach. *Phys. Rev. B* **22**, 5839 (1980).

47. Waber, J. T. & Cromer, D. T. Orbital radii of atoms and ions. *J. Chem. Phys.* **42**, 4116–4123 (1965).

48. Miedema, A. R., de Châtel, P. F. & de Boer, F. R. Cohesion in alloys— fundamentals of a semi-empirical model. *Phys. B + C* **100**, 1–28 (1980).

49. Brown, I. D. *The Chemical Bond in Inorganic Chemistry: The Bond Valence Model* (Oxford Univ. Press, Oxford, 2002).

50. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).

51. Kresse, G. & Furthmuller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).

52. Perdew, J. P. et al. Atoms, molecules, solids, and surfaces: applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B* **46**, 6671–6687 (1992).

53. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).

54. Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **13**, 5188 (1976).

55. Dudarev, S. L., Botton, G. A., Savrasov, S. Y., Humphreys, C. J. & Sutton, A. P. Electron-energy-loss spectra and the structural stability of nickel oxide: an LSDA + U study. *Phys. Rev. B* **57**, 1505–1509 (1998).

56. Kanoun, M. B., Reshak, A. H., Kanoun-Bouayed, N. & Goumri-Said, S. Evidence of Coulomb correction and spin–orbit coupling in rare-earth dioxides $CeO_2$, $PrO_2$ and $TbO_2$: an ab initio study. *J. Magn. Magn. Mater.* **324**, 1397–1405 (2012).

57. Mayernick, A. D. & Janik, M. J. Methane activation and oxygen vacancy formation over $CeO_2$ and Zr, Pd substituted $CeO_2$ surfaces. *J. Phys. Chem. C* **112**, 14955–14964 (2008).

58. Krcha, M. D. & Janik, M. J. Examination of oxygen vacancy formation in Mn-doped $CeO_2$ (111) using DFT + U and the hybrid functional HSE06. *Langmuir* **29**, 10120–10131 (2013).

59. Neugebauer, J. & Scheffler, M. Adsorbate-substrate and adsorbate-adsorbate interactions of Na and K adlayers on Al(111). *Phys. Rev. B* **46**, 16067–16080 (1992).

## Author contributions
N.J.O. completed the DFT calculations and associated data analysis. A.S.M.J. completed the statistical learning analysis. The project idea was conceived by M.J.J. and T.P.S. All authors contributed to writing the manuscript and approved the final version.

## Competing interests
The authors declare no competing interests.

## Additional information
**Supplementary information** is available for this paper at https://doi.org/10.1038/s41929-018-0094-5.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Correspondence and requests for materials** should be addressed to M.J.J. or T.P.S.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.