ARTICLE   OPEN

Check for updates

Data-driven models for ground and excited states for Single Atoms on Ceria

Julian Geiger[1,2], Albert Sabadell-Rendón[1,2], Nathan Daelman [1] and Núria López [1] ✉

Ceria-based single-atom catalysts present complex electronic structures due to the dynamic electron transfer between the metal atoms and the semiconductor oxide support. Understanding these materials implies retrieving all states in these electronic ensembles, which can be limiting if done via density functional theory. Here, we propose a data-driven approach to obtain a parsimonious model identifying the appearance of dynamic charge transfer for the single atoms (SAs). We first constructed a database of (701) electronic configurations for the group 9–11 metals on $CeO_2(100)$. Feature Selection based on predictive Elastic Net and Random Forest models highlights eight fundamental variables: atomic number, ionization potential, size, and metal coordination, metal–oxygen bond strengths, surface strain, and Coulomb interactions. With these variables a Bayesian algorithm yields an expression for the adsorption energies of SAs in ground and low-lying excited states. Our work paves the way towards understanding electronic structure complexity in metal/oxide interfaces.

INTRODUCTION

Single-atom catalysts (SACs) are a promising class of materials that provide the ultimate miniaturization of the active site[1–5]. They present the best utilization of expensive metal atoms, given they can be prepared in high concentrations and defined speciation[6,7]. Various supports have been used to fabricate stable SACs, including other metals, (doped) carbons, metal-organic frameworks, and most commonly oxides, for which the definition was coined[1,3,8]. Stabilization of single atoms (SAs) on oxides is often attributed to a strong metal-support interaction (SMSI), in which electron density is donated from the metal to the support, thus making reducible oxides particularly suitable host materials[3,9]. However, the reverse electron transfer (from the oxide to metal nanoparticles) is also possible[10]. Dispersing metals up to single atoms is very efficient in oxygen-rich environments of Fe_3O_4 and ceria, and stable SAs of rhodium, palladium, platinum and gold[8,11–14] have been characterized.

Owed to their isolated nature, SAs exhibit unusual electronic properties. For instance, in an AgCu single-atom alloy catalyst, free-atom like states of copper were reported[15], while strong electronic coupling in a Pd_1/Fe_2O_3 SAC resulted in improved alkyne semi-hydrogenation performance[16]. Moreover, the oxidation state of the isolated platinum atom in Pt_1/CeO_2 dynamically changes due to the electronic coupling of the metal and ceria support[17]. For ceria, this process results in localized Ce^{3+} centers (polarons). The electron transfer process is assisted by lattice phonons and was therefore coined as phonon-assisted metal-support interaction (PAMSI)[17]. A Born–Haber model was devised to understand the origin of the dynamic behavior of the Pt_1/CeO_2 system that retrieved the energy difference between the adsorbed Pt in its neutral and charged state with concomitant reduction of the oxide. To compensate for the ionization potential of the metal atom, as well as the reduction and distortion of the oxide, covalent Pt–O interactions contribute, and large favorable changes in Coulomb interactions tip the balance.

The dynamic nature of the metal oxidation state (mOS) in Pt_1/CeO_2 challenges the classical assignment of a static charge to SACs[9], thus limiting the applicability of traditional structure-activity relationships. Consequently, Pt_1/CeO_2 needs to be represented as an electronic ensemble with similar geometric structures, but significantly different electron density distributions. Highly active metastable states, which are easily populated under operando conditions, can considerably influence material properties and catalytic reactivities, as has been shown for geometric ensembles[18]. The dynamic effect significantly increases structural and electronic complexity, thus requiring sophisticated multiscale modeling approaches[19]. Similarly, assessing the properties of multiple low-lying electronic states is a non-trivial task. As seen for Pt, these states are closely spaced within the energy spectrum, and the number of possible distributions grows rapidly with the number of electrons exchanged by the isolated metal and the oxide.

Thus, the prediction of dynamic behavior is challenging and we lack a model that provides a physical expression for the interaction between the SAs and the reducible oxide support. Machine-learning (ML) techniques applied to materials modeling hold promise to overcome these hurdles[20–22]. Data-driven models based on readily available physical properties have proven to simplify the analysis of complex configurational spaces. For instance, a Gaussian Process Regression model was used to augment first-principles estimates of the reduction temperatures of 38 metal oxides[23]. ML methods further predicted adsorption energies of SACs on various oxides[24], shed light on the influence of surface modifications on MgO(100) supports[25], and identified aggregation trends compromising SA stability[26]. A Least Absolute Shrinkage and Selection Operator (LASSO) model was used to single out the 75 most representative descriptors out of 300k candidates governing the adsorption of transition metals (TMs) on oxide supports[27]. Many of these exploratory efforts have resulted in robust models, but require large feature spaces and offer limited explainability.

[1]Institute of Chemical Research of Catalonia (ICIQ), The Barcelona Institute of Science and Technology (BIST), Avinguda Països Catalans 16, 43007 Tarragona, Spain. [2]Universitat Rovira i Virgili, Avinguda Catalunya, 35, 43002 Tarragona, Spain. ✉email: nlopez@iciq.es
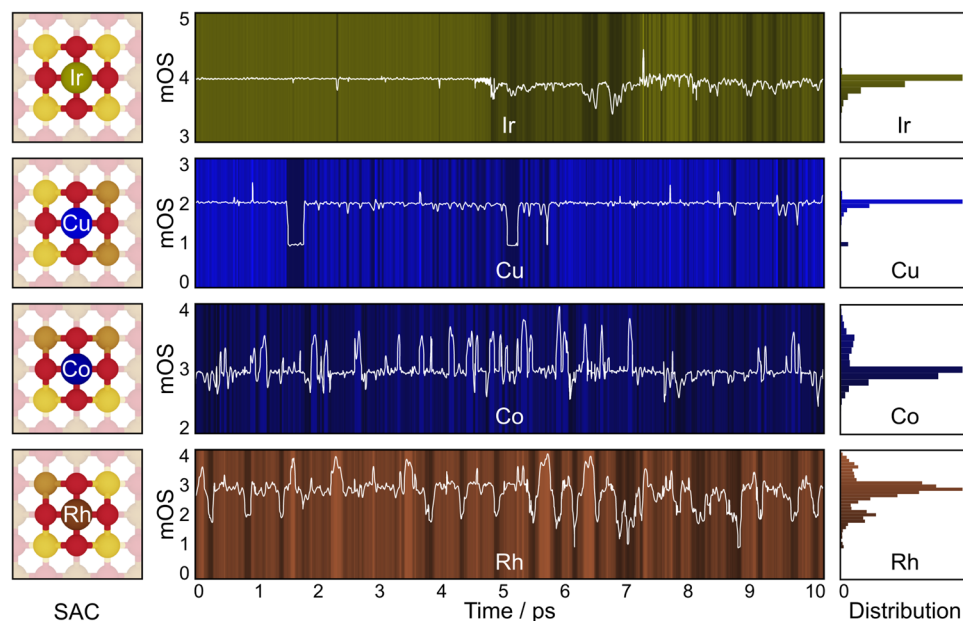
Fig. 1 Molecular Dynamics of selected CeO$_2$(100)-4O-based SACs investigated for metal oxidation state (mOS) dynamics at typical working temperature (600 K). The left panels show top views of the initial, unoptimized SAC structures, with ground state Ce^{3+} distributions indicated. colour code: O^{2-}, red; Ce^{4+}, orange; Ce^{3+}, yellow (metal species annotated; one subsurface polaron for Co-4O). The center panels depict the continuous mOS progressions during the 10 ps trajectories, indicated via the white lines and background colour gradients, where brighter (darker) colour corresponds to higher (lower) mOS values. Distributions of states at mOS increments of 0.1 are shown in the right panels. The full trajectories of the BOMD simulations can be seen in the Supplementary Videos 1–4.

In this work, we employed a data-driven approach to augment the simple representation of a physics-based two-state Born–Haber model, constructed for Pt$_1$/CeO$_2$[17]. To this end, we have created an extensive database of (701) ground and low-lying excited states for CeO$_2$(100)-based SACs of the group 9–11 metals, accounting for their rich and varied redox chemistry, adapting up to five distinct metal charge states, as well as alternative coordination environments of the ceria support. We applied a toolbox of ML techniques to reproduce the energies of ground and excited states for all SACs. Predominant features were then used to construct a parsimonious, interpretable model that predicts the appearance of dynamic electron transfer and rationalizes main contributions.

RESULTS AND DISCUSSION

Metal adsorption on CeO$_2$ surfaces

To analyze the stability of single metal atoms on the low-energy facets of ceria (see Supplementary Fig. 1 and Supplementary Note 1), we evaluated their adsorption energies by means of DFT (PBE+U) via: $E_{Ads} = E_{total} - E_{surface} - E_{bulk\ metal}$, where E_{total} denotes the energy of the SAC, $E_{surface}$ that of the unreconstructed oxide surface and $E_{bulk\ metal}$ the metal cohesive energy per atom (Supplementary Tables 1 and 2). Exothermic (endothermic) adsorption indicates favorable dispersion in the form of single atoms (nanoparticles). On the lowest energy facet (111), as well as on (110), adsorption is always endothermic with respect to the bulk metal, the only exceptions being cobalt and nickel on (110).

In contrast, adsorption on unreconstructed (2O) CeO$_2$(100) is exothermic for cobalt, nickel, copper, silver and gold. Moreover, lattice oxygen atoms in CeO$_2$(100) easily diffuse and rearrange to form oxygen-rich domains[28]. These modified coordination environments involve three or four surface oxygen atoms, denoted 3O and 4O, respectively. 3O-sites stabilize the first-row TMs cobalt, nickel and copper, while adsorption on square-planar (100)-4O is exothermic for all metals (except for Au and Ag), rendering them as stable SACs. These coordination motifs are very common on

ceria as they appear at CeO$_2$(111) step edges[29] and the corners of cuboctahedral ceria nanoparticles[30], agreeing with experimental observations[29,31]. Upon adsorption, electrons from the deposited metal atoms can be transferred to the ceria support and localize at Ce^{4+} centers, forming Ce^{3+} polarons. Thus, counting reduced cerium centers can be used as a proxy for the metal oxidation state, mOS[30], avoiding difficulties in charge assignment[32]. Particularly, for Pt$_1$/CeO$_2$(100) dynamic electron exchange between the metal atom and the support was found at around 500 K according to our previous ab initio Born-Oppenheimer Molecular Dynamics (BOMD)[17].

Dynamic behavior of M$_1$/CeO$_2$(100) SACs

To assess if the dynamic charge transfer is common to other metals, we performed 10 ps BOMD simulations of the stable CeO$_2$(100)-4O SACs at 600 K for cobalt, rhodium, iridium and copper (Supplementary Note 2). Figure 1 shows the continuous mOS progressions and frequency distributions during the BOMD trajectories for each system. Interestingly, the very redox active metal iridium remains in a charge state of 4+ throughout, with 3+ states persisting for less than 30 fs and charge transfer events being limited to the oxide (i.e., polaron hopping; see Supplementary Fig. 2). On the other hand, copper presents a 2+ ground state and forms two long-lived (about 200 fs) Cu$^+$ species, while cobalt fluctuates between mOS values of 2+ and 4+, transitioning through the 3+ ground state. Co^{2+} states are short-lived (< 10 fs) and precede transitions to 4+ states, which are stable for up to 70 fs. Rhodium shows the most dynamic behavior and adapts mOS values ranging from 1+ to 4+. The longest lifetime of this system in its 3+ ground state is below 600 fs, while 2+ (4+) states last up to 200 fs (150 fs), and also short-lived (<50 fs) 1+ states are accessible. Overall, for copper, cobalt and rhodium, we identify ~7, 27, and 65 states at varying mOS that persist for more than 30 fs, respectively.

Thus, the BOMD trajectories show that all SACs except Ir$_1$/CeO$_2$ present a dynamic mOS, albeit to varying extents. It is therefore challenging to predict mOS dynamics ad-hoc, since a poor redox
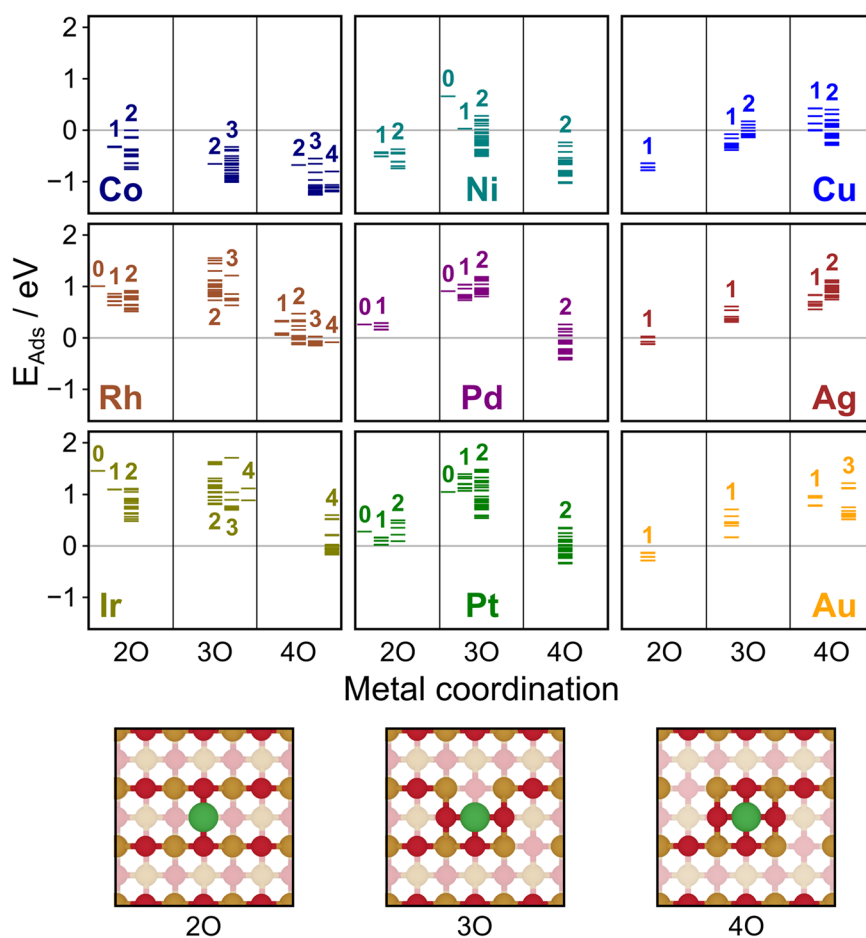
Fig. 2 Adsorption energies for ground and accessible excited states of the investigated CeO$_2$(100)-based transition metal (TM) SACs. The individual values were obtained via: $E_{Ads} = E_{total} - E_{2O\ surface} - E_{bulk\ metal}$. Each TM is represented by one panel. The different coordination environments (illustrated in the bottom row) are separated horizontally within each panel. Metal oxidation states, mOS, are annotated above/below the respective energy distributions.

active metal such as copper does present the effect, while a highly multivalent metal like iridium does not. However, an extended investigation with BOMD for other metals, coordination environments and longer timescales necessary to ascertain ground states and dynamicity is computationally not possible.

Sampling of SAC electronic ensembles

Alternatively, the dynamic properties of CeO$_2$(100)-based SACs at typical working temperatures can be inferred from a static investigation of ground and excited states[17]. Thus, we performed a systematic exploration of electronic ensembles for the group 9–11 transition metals (M = cobalt, rhodium, iridium, nickel, palladium, platinum, copper, silver, gold), surface oxygen coordinations (N_O), metal oxidation states (mOS), and Ce^{3+} distributions resulting from electron transfer to the different cerium centers in the surface. By considering all possible Ce^{3+} configurations up to an mOS of 2+, we have created 1242 initial structures, see Supplementary Note 3, that converged to a dataset of 701 distinct optimized geometries, as presented in Fig. 2.

We find that each metal can be stabilized by at least one oxygen coordination environment on the CeO$_2$(100) surface. However, there is great variability throughout the set: while for the group 9 and 10 metals, square-planar 4O coordination presents the ground, 2O an intermediate and 3O a metastable state, for group 11, adsorption energies become more endothermic with increasing valency of the metal atom. Moreover, the ground state mOS largely depends on the coordination number,

where, higher valency generally leads to more highly charged metal species. In numerous cases, SACs at a given coordination and charge state show the presence of low-lying excited states that cover energy ranges up to 1 eV in magnitude, see Supplementary Table 3. Thus, the electronic ensembles for different mOS frequently overlap, indicative of dynamic behavior at finite temperatures. The ground states deduced from the static sampling reproduce the majority species observed in the BOMD simulations (Fig. 1), and the frequencies and lifetimes of dynamic mOS deviations closely follow the energy distributions of the ensembles. For instance, in addition to its 3+ ground state, Co$_1$/CeO$_2$(100)-4O mainly occupies 4+ states, as expected from the small energy difference between the two ensembles, see Fig. 2. Similarly, the static mOS behavior of Ir$_1$/CeO$_2$(100)-4O relates to the lack of accessible states with a charge other than 4+.

Feature space construction

To investigate the main contributions ruling the dynamic charge transfer, we built data-driven ML models that reproduce the adsorption energies of ground and accessible low-lying excited states of the SACs. Our initial pool of atomic metal descriptors is based on the set proposed by O'Connor et al.[27], containing among others: the atomic number (Z), the cumulative ionization potential (IP), electronegativity, (χ), and orbital levels/radii, amounting to a total of nine features. Following previous ML models for the prediction of ground state adsorption energies of SAs on oxide supports[24,27,33], we added thermodynamic data for metal–metal
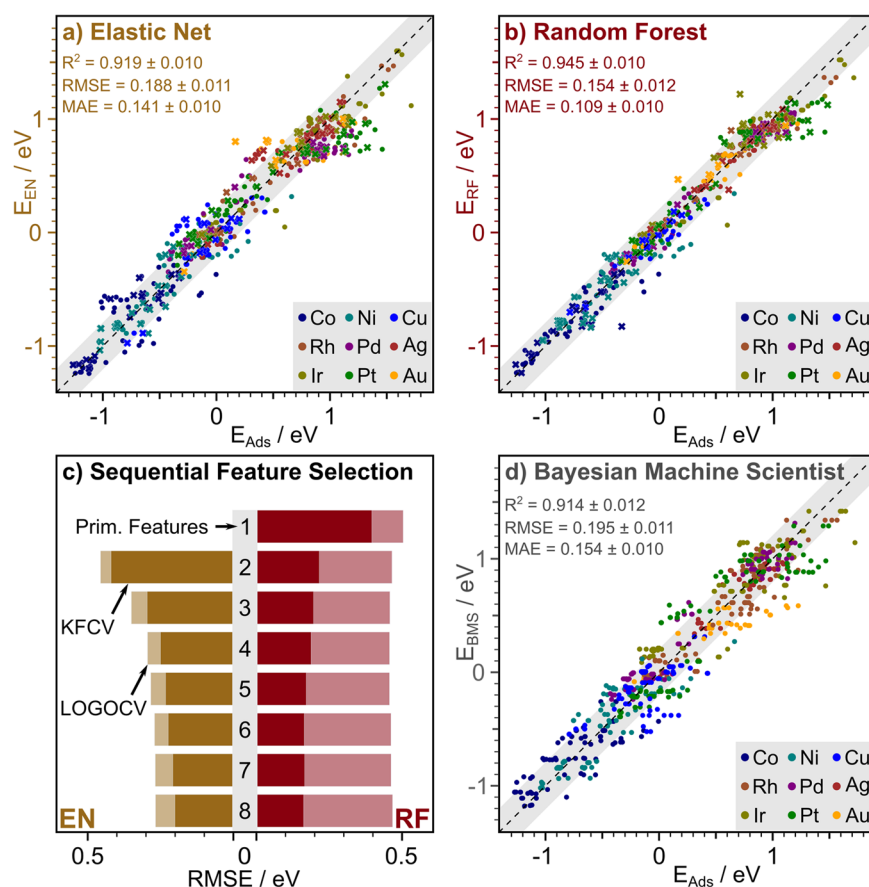
Fig. 3 Predictive ML models constructed from the DFT dataset of adsorption energies (E_{Ads}) of ground and excited states of ceria-based group 9–11 transition metal (TM) SACs. Predictions by the best-performing Elastic Net (EN), Random Forest (RF), and Bayesian Machine Scientist (BMS) models are shown in the panels **a**, **b**, and **d**, with training (testing) data points indicated as circles (crosses). The gray areas mark a region with a deviation of up to 0.2 eV. Panel **c** shows the courses of RMSE values for the EN and RF models during sequential feature selection (SFS), evaluated via K-fold cross-validation (KFCV; opaque bars) and leave-one-group-out cross-validation (LOGOCV; transparent bars). The use of secondary features for Elastic Net (see main text) requires a minimum of two descriptors in the reduced primary space. The actual error values and selected primary features obtained during SFS, panel **c**, are listed in Supplementary Table 6.

and metal–oxygen interactions[34]. Specifically, we evaluated the bond enthalpies of the metal–metal, $\Delta G^{M–M}$, and metal–oxygen bonds, $\Delta G^{M–O}$, both for the diatomic molecules and for the bulk systems, see Supplementary Table 4.

Among the oxide descriptors, studies on oxygen defects on ceria demonstrate that the geometric properties of local Ce^{3+} distributions crucially affect the relative stability of vacancies[35]. Thus, we augmented the feature space with topological descriptors of the oxide: (i) the average number of surface oxygen bound to the Ce^{3+} centers, denoted ϵ, indicative of surface strain[35]; and (ii) the $M–Ce^{3+}$ and $Ce^{3+}–Ce^{3+}$ distances, accounting for Coulomb interactions between these centers. To remove the need for explicit DFT optimizations, the distances are expressed in units of primitive cell vectors. Owing to the inhomogeneity of the dataset, we used statistical measures (i.e., the sum, minimum, mean, maximum and standard deviation) of the respective distance sets as ML features, instead of the individual values, see Supplementary Note 4 and Supplementary Fig. 4.

Predictive ML models

In Fig. 3, we present a summary of the predictive ML models. In all cases, we employed K-fold cross-validation (KFCV) using a 5-fold data split and show predictions obtained by the best-performing model, with mean errors and standard deviations indicated. Linear regression (LR) with metal-only descriptors led to a convoluted description, proving that host descriptors are crucial to untangle the

metal clusters, see Supplementary Fig. 5. However, as fully linear models cannot capture the physics of the metal–oxide interaction, predictions stay poor (Supplementary Table 5).

In a next step, nonlinearity was introduced via a secondary feature space of second-order products and ratios of all combinations of primary descriptors[27] (871 in total), making regularization essential to combat overfitting. Hence, we use the Elastic Net (EN) method[36], as it combines l1-and l2-regularization and therefore achieves feature subset selection and robust predictions, even in the presence of correlated variables, see Supplementary Note 5 and Supplementary Figs. 4 and 6. With the secondary feature space, the EN model achieves an excellent predictive performance of 0.19 eV (RMSE) and 0.14 eV (MAE), see Fig. 3a. Complementary, a nonlinear Random Forest (RF) model, see Supplementary Note 6, provides similar predictive performance of 0.15 eV (RMSE) and 0.11 eV (MAE) with the primary feature space, see Fig. 3b. Thus, our models confirm that nonlinearity is crucial to reproduce the energy distributions of the electronic ensembles for each SAC, and validate the suitability of our descriptor pool. However, the large number (86) of convoluted features retained by EN and the poor interpretability of RF make it difficult to readily derive physical insight.

Feature space reduction

To identify the most representative descriptors, we then applied sequential feature selection (SFS), while evaluating model performances with KFCV, as well as leave-one-group-out cross-validation
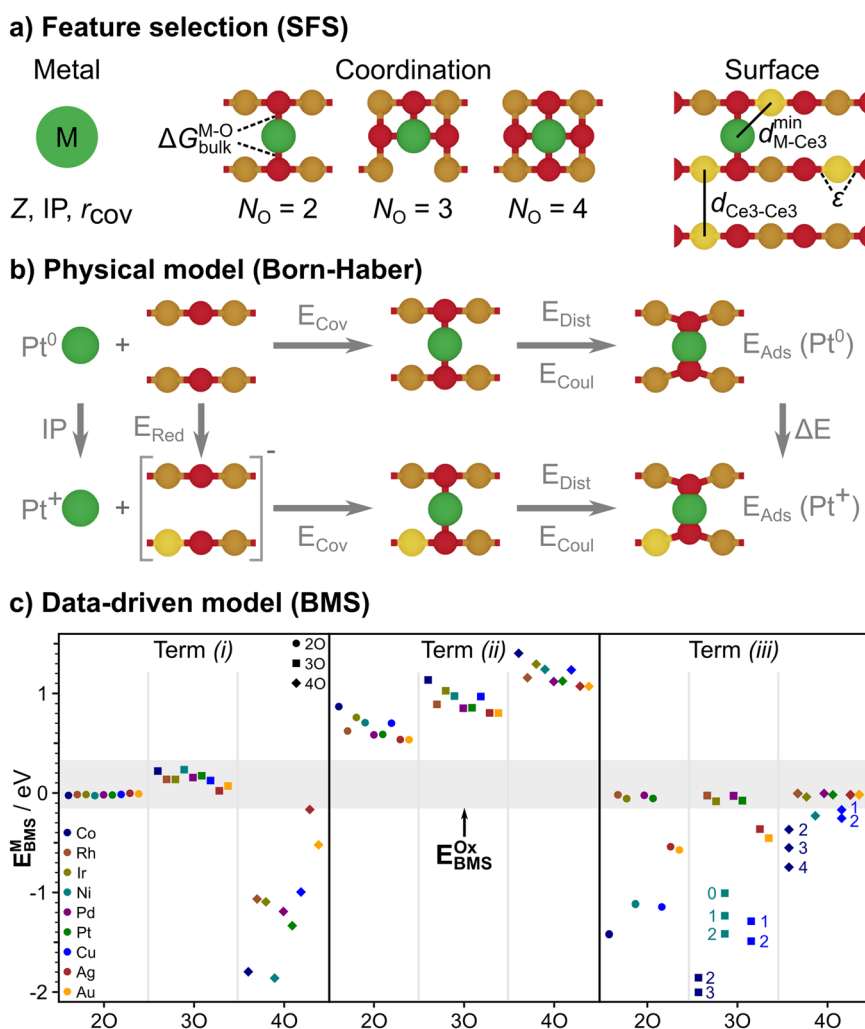
a) Feature selection (SFS)

Metal

ΔG_{bulk}^{M-O}

Z, IP, r_{cov}

Coordination

$N_O = 2$ $N_O = 3$ $N_O = 4$

Surface

d_{M-Ce3}^{min}

$d_{Ce3-Ce3}$

ϵ

b) Physical model (Born-Haber)

Pt^0 + ... E_{Cov} ... E_{Dist} ... E_{Ads} (Pt^0)

IP E_{Red} E_{Coul} ΔE

Pt^+ + [...]⁻ E_{Cov} ... E_{Dist} ... E_{Ads} (Pt^+)

E_{Coul}

c) Data-driven model (BMS)

Term (i)   Term (ii)   Term (iii)

● 2O ■ 3O ◆ 4O

E_{BMS}^M / eV

E_{BMS}^{Ox}

Co, Rh, Ir, Ni, Pd, Pt, Cu, Ag, Au

Fig. 4 Schematic of the data-driven construction of a closed-form model for the interaction of single metal atoms with the CeO₂(100) support, obtained by the Bayesian Machine Scientist (BMS). **a** Representative features selected via sequential feature selection (SFS) from predictive ML models. **b** Physical Born–Haber model previously used to describe the metal-support interaction of a Pt₁/CeO₂(100) SAC at fixed coordination and two charge states (energy terms: Red = Reduction; Cov = Covalent; Dist = Distortion; Coul = Coulomb; Ads = Adsorption). **c** Energy contributions given by each individual term of E_{BMS}^M. The range of energy contribution due to the support, E_{BMS}^{Ox}, is indicated as gray background. Relevant mOS values for term (iii) are annotated.

(LOGOCV), see Supplementary Note 7). Figure 3c) shows the RMSE values of the Elastic Net and Random Forest models during SFS up to the inclusion of eight primary features, while the full data is presented in Supplementary Table 6. The use of Random Forest with LOGOCV leads to considerable errors throughout, which we attribute to the poor extrapolation capabilities of the method. On the other hand, the EN model converges to low errors of around 0.20 eV, irrespective of the applied data partitioning. Predictions by the Elastic Net model during leave-one-group-out cross-validation are presented in Supplementary Fig. 8.

Importantly, as feature subsets are expanded during SFS, the errors of both models quickly converge. Thus, a reduced set of representative physical descriptors is sufficient to capture the complex interactions governing metal adsorption energies of ceria-based SACs. As the actual feature sets obtained from SFS vary with model choice and validation procedure, see Supplementary Table 6, we instead identified frequently occurring feature classes that directly map to distinct physical properties of the systems, see Fig. 4a. These properties, and corresponding representative features, are (i) the metal species, Z; (ii) its charge state, IP (cumulative up to mOS); (iii) its size, r_{cov}; (iv) the coordination environment, N_O; (v) the covalent contribution of

the metal–oxygen bonds, ΔG_{bulk}^{M-O}; (vi) surface strain and distortion, ϵ; (vii) Coulomb interactions between the metal and the support, $min(d_{M-Ce3})$; and lastly (viii) Coulomb interactions between Ce^{3+} centers, $d_{Ce3-Ce3}$. Notably, this reduced space of eight representative features contains all contributions earlier suggested in the Born–Haber model[17], as well as relevant geometric properties of the ceria support identified by Murgida et al.[35].

Search for a closed-form model

Based on the reduced set of representative features, we then searched for a closed-form model for the metal–oxide interaction that generalizes the two-state Born–Haber cycle (Fig. 4b), to all considered metals, coordination environments, oxidation states, and surface electron distributions. Such a model has the advantage that mathematical analysis of the obtained functional form can provide additional physical insight. We employed the Bayesian Machine Scientist (BMS)[37], which samples the space of possible functional forms that fit the training data (further details are provided in the Methods Section). During preliminary runs with all eight features (Supplementary Table 7), BMS only retained dissimilar, reduced subsets, see Supplementary Note 8. Due to the

greater energy differences between the metal ensembles, compared to within the individual distributions, BMS mostly retained metal features, while oxide variables were discarded. Therefore, initial models (Supplementary Eqs. 7 to 9) had low predictive power and were elusive to physical interpretation.

Consequently, we decomposed the adsorption energies, E_{Ads}, into a predominant metal and a separate oxide contribution. Training the first BMS model with only the five metal features (see above) gives an approximate positioning of each metal ensemble in the appropriate region of the energy spectrum, E_{BMS}^M. To retrieve the remaining oxide contributions, approximated by E_{BMS}^{Ox}, we then subtracted E_{BMS}^M from the DFT adsorption energies and trained a second BMS model on the residuals, using the remaining three host descriptors. Following this approach, we obtained consistent and interpretable functional forms for both contributions, as presented in Eqs. (1) to (3). The full, additive model gives accurate adsorption energy predictions (RMSE = 0.20 eV and MAE = 0.15 eV), as shown in Fig. 3d. Relevant fitting constants are reported in Supplementary Table 8.

$$E_{Ads} = E_{Ads}^M + E_{Ads}^{Ox} \approx E_{BMS}^M + E_{BMS}^{Ox} = E_{BMS}. \tag{1}$$

$$E_{BMS}^M = \overbrace{N_O c_1 \left(\Delta G_{bulk}^{M-O} + c_2\right)^{N_O} \cos(r_{cov})}^{(i)} + \overbrace{N_O c_3 + \Delta G_{bulk}^{M-O\,2}}^{(ii)} \\ - \overbrace{\frac{N_O r_{cov}^2 (IP + Z)^{-c_2 - N_O}}{c_4\left(\Delta G_{bulk}^{M-O} + c_5\right)^2 \left(\Delta G_{bulk}^{M-O} \cdot (IP+Z)^{-c_2-N_O} + c_4\right)\left(r_{cov} + \cos\left(\frac{c_6 r_{cov}}{c_4}\right)\right)^2}}^{(iii)} \tag{2}$$

$$E_{BMS}^{Ox} = \frac{1}{c_1}\left(d_{Ce^3 - Ce^3} + \frac{\min(d_{M-Ce^3})}{c_2^2 \min(d_{M-Ce^3}) - c_2 d_{Ce^3-Ce^3}} \cos\left(c_1\left(c_1 \epsilon(c_1 + c_2) + \epsilon^2\right)\right)\right) \tag{3}$$

Deriving physical insight from the closed-form model

For better interpretability, we evaluated each of the three constituent terms of E_{BMS}^M (Eq. (2), see above), denoted (i), (ii), and (iii), separately, while their individual energy contributions are presented in Fig. 4c. Term (i) spans from –1.86 to 0.24 eV and attributes the greatest stability to highly coordinated (small) metals that form strong bonds with oxygen. Thus, it accounts for metal–oxygen bonding in the different coordination environments, and predicts the following stability trends: $N_O = 4$ (exothermic) > $N_O = 2$ (thermoneutral) > $N_O = 3$ (endothermic), in agreement with expectations derived from coordination chemistry. Thus, term (i) correctly reproduces the dependence of the adsorption energies on the coordination number for the group 10 metals (Ni, Pd, Pt, Rh, and to some extent, Ir (see Fig. 2). However, it provides a poor approximation for the group 11 metals (Cu, Ag, Au), as they do not follow the same energy trends.

Term (ii) is endothermic, spanning from 0.54 to 1.40 eV, and only depends on ΔG_{bulk}^{M-O} and N_O. Contrary to term (i), it particularly destabilizes high coordination and strong metal–oxygen bonds, thus we mainly attribute it to distortion. Term (ii) provides a first correction to the estimates of term (i), and further distinguishes the different metals and coordination environments in the energy spectrum.

Lastly, term (iii) is given by a fraction that involves all five metal features and spans from –2.00 to 0.00 eV. It therefore introduces a distinction by the mOS. The contribution of term (iii) is exothermic for the first-row metals cobalt, nickel, and copper, particularly at lower coordination $N_O = 2$ and 3. In the limit of large atomic numbers and covalent radii, that is, for the second and third row TMs, term (iii) approaches zero, therefore leaving these metals almost unaffected. Owing to their weak metal–oxygen bond strength, the effect is somewhat mitigated for silver and gold, for

which term (iii) slightly stabilizes low coordinations. Overall, the influence of term (iii) is highly dependent on the specific metal atom and mOS state, thus providing a final refinement of the estimates given by terms (i) and (ii).

The remaining oxide contributions, approximated by E_{BMS}^{Ox} and illustrated as gray background in Fig. 4c, are mainly attributed to Coulomb interactions (within the support, and between the metal and Ce^{3+} centers), as well as lattice strain induced by the oxygen-rearrangements and surface polarons. E_{BMS}^{Ox} particularly penalizes configurations with increased lattice strain, which span from –0.15 to 0.33 eV, and are therefore of the same magnitude as polaron hopping barriers in ceria (0.40 eV)[38], or the reconstruction energy from 2O to 3O (0.30 eV, see Supplementary Note 1).

The investigated ML models (EN, RF, and BMS) all provide similar predictive accuracy. However, as outlined above, the closed functional expression obtained with BMS is physically interpretable. Supplementary Fig. 9 presents the BMS predictions in direct comparison to the original data of Fig. 2. Overall, deviations for ground states are within the range of the PBE+U inherent error for the ceria reduction energy (up to 0.4 eV)[39]. In certain cases, ground state mOS for a given metal and coordination are not correctly identified by the BMS. We attribute these discrepancies to the inherently small energy differences for these systems. For instance, DFT adsorption energies of Co-4O in its 3+ and 4+ charge states differ by only 0.06 eV, see Fig. 2. On the other hand, highly excited states are generally less well-reproduced, causing a contraction of the energy spans of the individual distributions. Nonetheless, the BMS model correctly reproduces ensemble overlaps, and thus mOS dynamicity.

Owing to its compact nature, we can compare the generalized model obtained from BMS to the phenomenological Born–Haber (BH) cycle, which only provided the energy difference for two states (Pt and Pt^+), in a fixed coordination environment (2O). While the BH model is derived from physical insights, BMS functional forms are data-driven. The BH is based on seven local variables, including the IP, surface distortion/reduction, a covalent contribution, and Coulomb interactions, which were evaluated using explicit DFT distances and Bader charges. Instead, the BMS accounts for nine metals, three coordination environments, and up to five different oxidation states, while using only eight variables, none of which require explicit DFT. Owing to its significant contribution, the IP appears in both models. The covalent term, E_{Cov}, in BH was approximated by the $Pt(H_2O)_2$-binding energy. In the BMS, it is replaced by the more general metal–oxygen bond-formation enthalpy, ΔG_{bulk}^{M-O}. As a bulk property, ΔG_{bulk}^{M-O}, also contains non-local ionic interactions that were approximated via the Coulomb term in the BH model (see Supplementary Note 4). The remaining contributions in the BH cycle are the reduction and surface distortion of ceria, both of which are rather small, as no additional oxygen-rearrangements were considered. In BMS, these effects are condensed in the surface term, E_{BMS}^{Ox}, which further accounts for long-range repulsion and elastic properties of the surface due to the localized polarons.

Finally, to provide an outlook on the expected generalization capabilities of our BMS model, we consider the impact of different metals and reducible oxide supports. Retaining the same functional form and atomic descriptors, the adsorption energies of other SA metals can be predicted by fine tuning the coefficients, as evaluated for our set of metals through leave-one-group-out cross-validation, see Supplementary Table 9. As for the oxides, the surface contributions account for local strain and polaron structure, and can thus be adapted to other reducible oxide supports using equivalent descriptors. Accounting for defects, such as oxygen vacancies, will require future extensions of the method[40].

In summary, we have shown that the dynamic charge transfer between isolated single-metal atoms and ceria supports is ubiquitous and can be predicted via an interpretable, parsimonious

mathematical expression for the metal-support interactions. Our data-driven model employs a set of eight variables, including metal properties like the atomic number, covalent radius, ionization potentials, metal–oxygen bond strengths, and the coordination number, as well as oxide contributions, accounting for surface strain and Coulomb interaction. It generalizes a previous, physical Born–Haber model and reproduces the adsorption energies for ground and low-lying excited states for a variety of CeO_2(100)-based SACs. The proposed methodology of augmenting physics-based models with data-driven machine-learning techniques allows for a generalization of dynamic electron transfer in metal-support interactions and paves the way towards the introduction of complex electronic structure contributions in the modeling of single atoms for heterogeneous catalysis.

METHODS

Density functional theory

DFT simulations were performed with the Vienna Ab Initio Simulation Package (VASP; version 5.4.4)[41] and the Perdew–Burke–Ernzerhof (PBE) functional[42]. For the cerium atoms, an additional Hubbard U correction[43], with an effective U_{eff} value (Dudarev's approach[44]) of 4.5 eV was applied[45]. Core electrons were treated with the projector augmented wave (PAW) method[46,47] using the appropriate PAW-PBE pseudopotentials, while valence electrons were expanded with plane waves up to a basis set cutoff of 500 eV. Total energies were evaluated at the Gamma point, and validated by comparison to (3 x 3 x 1) k-point sampling, see Supplementary Fig. 3. Electronic convergence was set to 10^{-6} eV and atomic positions were converged until residual forces fell below 0.01 eV Å^{-1}.

Slab models for the (111) and (100) surfaces were constructed as (3 x 3), and for the (110) surface as (2 x 2) supercells, based on optimized ceria bulk (theoretical lattice parameter: 5.491 Å). They extend nine, nine, and six atomic layers along the vertical direction, with the bottom three, four, and three layers fixed at the optimized bulk positions, respectively. At least 10 Å of vacuum were added on top of the surfaces to avoid non-physical interactions between periodic images. Ce^{3+} centers, and consequently discrete mOS states, were assigned through the localized magnetic moments of cerium atoms, where we applied a threshold of 0.8 μ_B, based on previous work[17].

For the Born-Oppenheimer Molecular Dynamics (BOMD) simulations, we used reduced (2 x 2) supercells of CeO_2(100). They were conducted within the canonical NVT ensemble (constant number of particles, volume and temperature), using the Nosé-Hoover thermostat[48] at an average temperature of 600 K. Data was collected for 10 ps trajectories, with a time-step of 1 fs. For the MD simulations, we lowered the electronic convergence threshold to 10^{-5} eV. Continuous mOS progressions along the MD trajectories (center panels of Fig. 1) were obtained by summing up the absolute magnetic moments of all surface and subsurface cerium atoms, without the application of a magnetization threshold. The distributions shown in the right panels of Fig. 1 represent states at discrete mOS values at increments of 0.1 μ_B.

Statistical modeling

All the data follows FAIR principles (findable, accessible, interoperable, and reusable) according to the guidelines outlined by Artrith et al.[22]. Predictive models (LR, RF, EN)[36] were evaluated using KFCV with a 5-fold data split, and LOGOCV procedures, see Supplementary Table 6. Data centering and standardization were applied when necessary. Reported error measures (R^2, RMSE, MAE) were averaged over the different train-test splits, with standard deviations indicated. Presented energy predictions correspond to the best-performing models, as identified during the respective cross-validation procedure. Hyperparameter tuning for the EN and RF models is provided in Supplementary Figs. 6 and 7. For the final EN model, we used values of $\alpha = 1 \times 10^{-3}$ and l1-ratio $= 0.999$ throughout (low total amount of regularization, high fraction of l1). The RF was constructed as an ensemble of 128 trees, which were grown to a maximum depth of eight. At each split, randomly selected features amounting to 60% of the given pool were evaluated. Training was performed on bootstrapped data sets. For data visualization, we made extensive use of the Plotly Python library[49].

The symbolic equation search was performed via the Bayesian Machine Scientist (BMS)[37], which samples the space of possible functional forms that fit the training data using using Markov Chain Monte-Carlo. As the space of mathematical functions to be explored by BMS grows exponentially with the number of variables, the method requires a simplified descriptor space[50]. We limited the number of steps to 10,000, of which the first 1000 were discarded. The maximum depth was set to 100 iterations per step. At each step, we examined the following BMS outputs: (i) the current function and its complexity, (ii) the Bayesian Information Criterion (BIC), (iii) the sum of squared errors (SSE), and (iv) the values of all the fitting constants. Functions were evaluated by mathematical analysis, focusing on accuracy and simplicity. For the separate metal (oxide) models we used six (three) fitting constants, and suitable priors accordingly. BMS allows continuous, as well as discrete variables as inputs. In our setup, the latter are given by the metal features and the coordination number, both of which separate distinct classes of data. In practice, BMS frequently achieves separation of data classes and approximation of polynomial terms by the use of trigonometric functions, as their compact nature lowers the applied penalty on model complexity.

DATA AVAILABILITY

DFT optimized structures and Molecular Dynamics trajectories are available in the ioChem-BD database[51] (https://doi.org/10.19061/iochem-bd-1-230).

CODE AVAILABILITY

Code and raw data can be retrieved from our GitLab repository (https://doi.org/10.19061/ceria-sac).

REFERENCES

1. Qiao, B. et al. Single-atom catalysis of CO oxidation using Pt_1/FeO_x. Nat. Chem. **3**, 634–641 (2011).
2. Wang, A., Li, J. & Zhang, T. Heterogeneous single-atom catalysis. Nat. Rev. Chem. **2**, 65–81 (2018).
3. Lang, R. et al. Single-Atom Catalysts Based on the Metal−Oxide Interaction. Chem. Rev. **120**, 11986–12043 (2020).
4. Hai, X. et al. Scalable two-step annealing method for preparing ultra-high-density single-atom catalyst libraries. Nat. Nanotechnol. **17**, 174–181 (2022).
5. Datye, A. K. & Guo, H. Single atom catalysis poised to transition from an academic curiosity to an industrially relevant technology. Nat. Commun. **12**, 895 (2021).
6. Resasco, J. et al. Uniformity is Key in Defining Structure−Function Relationships for Atomically Dispersed Metal Catalysts: The Case of Pt/CeO_2. J. Am. Chem. Soc. **142**, 169–184 (2020).
7. Xu, Y. et al. Revealing the Correlation between Catalytic Selectivity and the Local Coordination Environment of Pt Single Atom. Nano Lett. **20**, 6865–6872 (2020).
8. Hulva, J. et al. Unraveling CO adsorption on model single-atom catalysts. Science **371**, 375–379 (2021).
9. Lykhach, Y. et al. Counting electrons on supported nanoparticles. Nat. Mater. **15**, 284–288 (2016).
10. Cho, Y. et al. Disordered-Layer-Mediated Reverse Metal−Oxide Interactions for Enhanced Photocatalytic Water Splitting. Nano Lett. **21**, 5247–5253 (2021).
11. Jones, J. et al. Thermally stable single-atom platinum-on-ceria catalysts via atom trapping. Science **353**, 150–154 (2016).
12. Qiao, B. et al. Highly Efficient Catalysis of Preferential Oxidation of CO in H_2-Rich Stream by Gold Single-Atom Catalysts. ACS Catal. **5**, 6249–6254 (2015).
13. Khivantsev, K. et al. Economizing on Precious Metals in Three-Way Catalysts: Thermally Stable and Highly Active Single-Atom Rhodium on Ceria for NO Abatement under Dry and Industrially Relevant Conditions. Angew. Chem. Int. Ed. **60**, 391–398 (2021).
14. Muravev, V. et al. Interface dynamics of Pd-CeO_2 single-atom catalysts during CO oxidation. Nat. Catal. **4**, 469–478 (2021).
15. Greiner, M. T. et al. Free-atom-like d states in single-atom alloy catalysts. Nat. Chem. **10**, 1008–1015 (2018).
16. Gao, R. et al. Pd/Fe_2O_3 with Electronic Coupling Single-Site Pd−Fe Pair Sites for Low-Temperature Semihydrogenation of Alkynes. J. Am. Chem. Soc. **144**, 573–581 (2022).
17. Daelman, N., Capdevila-Cortada, M. & López, N. Dynamic charge and oxidation state of Pt/CeO_2 single-atom catalysts. Nat. Mater. **18**, 1215–1221 (2019).
18. Zhang, Z., Zandkarimi, B. & Alexandrova, A. N. Ensembles of Metastable States Govern Heterogeneous Catalysis on Dynamic Interfaces. Acc. Chem. Res. **53**, 447–458 (2020).

19. Wang, Y., Kalscheur, J., Su, Y. Q., Hensen, E. J. M. & Vlachos, D. G. Real-time dynamics and structures of supported subnanometer catalysts via multiscale simulations. *Nat. Commun.* **12**, 5430 (2021).

20. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).

21. Keith, J. A. et al. Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems. *Chem. Rev.* **121**, 9816–9872 (2021).

22. Artrith, N. et al. Best practices in machine learning for chemistry. *Nat. Chem.* **13**, 505–508 (2021).

23. Garrido Torres, J. A. et al. Augmenting zero-Kelvin quantum mechanics with machine learning for the prediction of chemical reactions at high temperatures. *Nat. Commun.* **12**, 7012 (2021).

24. Tan, K., Dixit, M., Dean, J. & Mpourmpakis, G. Predicting Metal−Support Interactions in Oxide-Supported Single-Atom Catalysts. *Ind. Eng. Chem. Res.* **58**, 20236–20246 (2019).

25. Liu, C. Y., Zhang, S., Martinez, D., Li, M. & Senftle, T. P. Using statistical learning to predict interactions between single metal atoms and modified MgO(100) supports. *NPJ Comput. Mater.* **6**, 102 (2020).

26. Su, Y. Q. et al. Stability of heterogeneous single-atom catalysts: a scaling law mapping thermodynamics to kinetics. *NPJ Comput. Mater.* **6**, 144 (2020).

27. O'Connor, N. J., Jonayat, A. S., 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.* **1**, 531–539 (2018).

28. Capdevila-Cortada, M. & López, N. Entropic contributions enhance polarity compensation for $CeO_2(100)$ surfaces. *Nat. Mater.* **16**, 328–334 (2017).

29. Dvořák, F. et al. Creating single-atom Pt-ceria catalysts by surface step decoration. *Nat. Commun.* **7**, 10801 (2016).

30. Figueroba, A., Kovács, G., Bruix, A. & Neyman, K. M. Towards stable single-atom catalysts: Strong binding of atomically dispersed transition metals on the surface of nanostructured ceria. *Catal. Sci. Technol.* **6**, 6806–6813 (2016).

31. Kunwar, D. et al. Stabilizing High Metal Loadings of Thermally Stable Platinum Single Atoms on an Industrial Catalyst Support. *ACS Catal.* **9**, 3978–3990 (2019).

32. Walsh, A., Sokol, A. A., Buckeridge, J., Scanlon, D. O. & Catlow, C. R. A. Oxidation states and ionicity. *Nat. Mater.* **17**, 958–964 (2018).

33. Iyemperumal, S. K., Pham, T. D., Bauer, J. & Deskins, N. A. Quantifying Support Interactions and Reactivity Trends of Single Metal Atom Catalysts over TiO_2. *J. Phys. Chem. C.* **122**, 25274–25289 (2018).

34. Haynes, W. M. *CRC Handbook of Chemistry and Physics 97th Edition* (CRC Press LLC Florence, Taylor & Francis Group, 2016).

35. Murgida, G. E., Ferrari, V., Ganduglia-Pirovano, M. V. & Llois, A. M. Ordering of oxygen vacancies and excess charge localization in bulk ceria: A DFT+U study. *Phys. Rev. B* **90**, 115120 (2014).

36. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York, 2009).

37. Guimerà, R. et al. A Bayesian machine scientist to aid in the solution of challenging scientific problems. *Sci. Adv.* **6**, eaav6971 (2020).

38. Plata, J. J., Márquez, A. M. & Sanz, J. F. Electron Mobility via Polaron Hopping in Bulk Ceria: A First-Principles Study. *J. Phys. Chem. C.* **117**, 14502–14509 (2013).

39. Capdevila-Cortada, M., Łodziana, Z. & López, N. Performance of DFT+U Approaches in the Study of Catalytic Materials. *ACS Catal.* **6**, 8370–8379 (2016).

40. Geiger, J. & López, N. Coupling Metal and Support Redox Terms in Single-Atom Catalysts. *J. Phys. Chem. C* https://doi.org/10.1021/acs.jpcc.2c03710 (2022).

41. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).

42. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).

43. Hubbard, J. Electron correlations in narrow energy bands. *Proc. Math. Phys. Eng. Sci.* **276**, 238–257 (1963).

44. Dudarev, S. & Botton, G. Electron-energy-loss spectra and the structural stability of nickel oxide: An LSDA+U study. *Phys. Rev. B* **57**, 1505–1509 (1998).

45. Fabris, S., Gironcoli, S. D., Baroni, S., Vicario, G. & Balducci, G. Taming multiple valency with density functionals: A case study of defective ceria. *Phys. Rev. B* **71**, 041102 (2005).

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

47. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).

48. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695 (1985).

49. Plotly Technologies Inc. *Collaborative Data Science* (Plotly Technologies Inc, 2015). https://plot.ly.

50. Pablo-García, S., García-Muelas, R., Sabadell-Rendón, A. & López, N. Dimensionality reduction of complex reaction networks in heterogeneous catalysis: From linear-scaling relationships to statistical learning techniques. *WIREs Comput. Mol. Sci.* **11**, e1540 (2021).

51. Álvarez-Moreno, M. et al. Managing the Computational Chemistry Big Data Problem: The ioChem-BD Platform. *J. Chem. Inf. Model.* **55**, 95–103 (2015).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

J.G. performed the calculations. J.G., A.S.-R., and N.D. carried out the data analysis and constructed the machine-learning models. J.G., A.S.-R., N.D., and N.L. prepared the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-022-00852-1.

Correspondence and requests for materials should be addressed to Núria López.

Reprints and permission information is available at http://www.nature.com/reprints

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