Going Beyond the d-Band Center to Describe CO₂ Activation on Single-Atom Alloys

Zhongti Sun, Zhilong Song, and Wan-Jian Yin*

The computational screening of single-atom alloys (SAAs) is challenging owing to their complicated electronic structure compared with traditional metal or alloy catalysts and the consequent lack of an accurate and low-cost activity descriptor to replace expensive adsorption energy calculations. Herein, a data-driven approach involving consecutive classification and regression is explored to identify the descriptor in the form of algebraic operators representing atomic information for predicting the adsorption energies of CO₂ molecules on SAAs. The best descriptor significantly outperforms the d-band center model in terms of accuracy and computational overhead. This study provides a fundamental understanding of the bonding strength on SAA surfaces and also an effective approach for the high-throughput screening of promising catalysts.

1. Introduction

Computational screening for novel and improved catalyst materials requires the precise and low-cost prediction of vital parameters, such as the adsorption energies of reactive species, which are crucial for understanding and simulating surface-related reactions.[1] However, adsorption energy calculations are not only time-consuming, but also rely on computational methods, and hence, may not meet the requirements of high-throughput material design. This has necessitated the development of effective approaches to predict adsorption energies with an accuracy analogical to that of previous approaches but which use simpler parameters to compute the target properties. For traditional transition metals (TMs) or alloy catalyst materials, the d-band center model,[2] standard thermodynamics scaling relation,[3] and Brønsted–Evans–Polanyi (BEP) relationships[4] have revealed a linear correlation of the d-band center, reaction energies, and activation energies with the bonding strength of reaction intermediates; these findings have significantly contributed toward improving the efficiency of modern catalyst design.[5] These approaches can draw its catalytic performance in terms of the simple electronic structure of the TM surface, allowing us to rapidly predict the adsorption energy using just one base parameter: the energy level of the d-band. Even though the d-band center of active site has proved to be extensively useful in the rational design of catalyst in chemical and electrochemical reactions,[6] it is still not an accurate reactivity descriptor for metal surfaces[7] and other complex material systems that deviate from traditional scaling relations.

Single-atom alloys (SAAs) are specific alloy classes and also a type of single-site catalyst, where few guest atoms disperse across the surface of the host matrix, such that no bonds are formed between neighboring guest atoms; this allows one to regulate the selectivity of the guest atom and activity of the host atom.[6] In principle, the effect of the isolation site on the SAAs may also break down the known limitations of traditional scaling relations, rendering the aforementioned d-band center not an exact descriptor. In addition, this unusual electronic structure with free-atom-like d-states of the guest atom may also result in SAAs with unprecedented catalytic properties.[8] To efficiently achieve rational design of SAAs, Monasterial et al. revealed that the guest atom’s p-band center is superior to d-band center on the hydrogen adsorption strength.[9] Fung et al., proposed an analytical expression of methane adsorption energy with the number of occupied and unoccupied s-, p-, and d-states using statistical learning algorithm.[10] Nevertheless, these descriptor or features all rely on the theoretical simulations, are very time consuming, and method dependent. Recently, a data-driven screening technique, together with outlier detection methods, was also raised to identify the potentially optimal performance of SAAs, deviating from thermodynamic scaling relations by Dasgupta et al.[11] Even though this method may also generate some unexpected and promising SAA catalyst, the complexity and interpretability of model may also hinder its high-throughput screening, particularly for experimentalists. Hence, it is essential to develop a data-driven approach to identify the best-performing descriptor in the form of algebraic functions with just base parameters.[12]

Z. Sun
School of Materials Science and Engineering
Jiangsu University
Zhenjiang 212013, P. R. China

Z. Song, W.-J. Yin
College of Energy
Soochow Institute for Energy and Materials InnovationS (SIEMIS) and Jiangsu Provincial Key Laboratory for Advanced Carbon Materials and Wearable Energy Technologies
Soochow University
Suzhou 215006, P. R. China

E-mail: yinwanjian@gmail.com, wjyin@suda.edu.cn

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/aesr.202100152.

© 2021 The Authors. Advanced Energy and Sustainability Research published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/aesr.202100152
As a representative example, we focus on the activation of CO₂ molecules from a linear free state to a bent chemisorption configuration in the first step of CO₂ reduction. CO₂ uptake is of critical importance to the subsequent catalytic conversion process. Thus far, no accurate and simple parameter has been proposed to describe CO₂ activation on SAAs.

In this work, a descriptor for CO₂ activation on SAAs was developed via the sure independence screening and sparsifying operator (SISSO) algorithm, a compressed sensing approach with first-principles-based inputs. By identifying the CO₂ adsorption energy descriptor with only atomic information as the primary feature, a two-step classification and regression methodology was constructed. From the classification, a 2D descriptor is adequate to separate the physical and chemical adsorption configurations. Finally, on the basis of the chemical adsorption energy database, we propose an explicit, analytical expression to quantitatively describe the correlation of the CO₂ adsorption energy with the primary features. The validation root-mean-squared error (RMSE) and maximum absolute error (MaxAE) of the 1D descriptors are 0.12 and 0.30 eV, respectively, as well as the best descriptor with 0.05 eV for RMSE and 0.21 eV for MaxAE, are all superior to those obtained by the traditional d-band center model (0.21 and 0.57 eV, respectively) in terms of accuracy and computational overhead. Our study demonstrates the advantage of data analysis and also offers a novel approach that is accurate and low cost to describe the binding energies of reactive species, which will aid in the computational screening of promising catalysts.

2. Results and Discussion

The adsorption energies of CO₂ molecule (E_ads) for Co-, Ni-, and Cu-based SAAs are simulated within the framework of the density functional theory (DFT) theory. Among these, TMs are usually served as reactive catalyst materials in the CO₂ hydrogenation reactions. The corresponding close-packed surfaces containing Co (001), Ni (111), and Cu (111) are selected as the host matrixes of SAAs. 28 kinds of TM elements, except Hg (which is toxic) from the periodic table, are served as the guest atoms of SAAs. Considering all nonequivalent high-symmetry sites close to the guest atom containing the top, bridge, face-centered cubic (fcc), and hexagonal-close-packed (hcp) hollow sites, the C=O bond of the CO₂ molecule is placed at these possible locations (Figure S1, Supporting Information). Verification with all possible adsorption configurations demonstrates that there are six kinds of most stable adsorption models among the 84 datasets, as marked in Figure 1; details are shown in Figure S2–S4, Supporting Information. The five kinds of bent CO₂ models (0–5) are denoted as the chemisorption model, that is, “chem,” and the remaining quasilinear CO₂ model (6) is denoted as the physisorption model, that is, “phys.” Figure 1b shows the most favorable adsorption configuration for each SAA, indicating that the adsorption models with the “chem” form transform into the free state-like “phys” form, with increasing (the occupancy number of d-orbitals) for the guest atom from left to right. This phenomenon can be qualitatively understood by the classical chemical bond theory, whereby a variation in the bonding strength mainly arises from the antibonding state, and a lower d-band energy level (i.e., more occupied d-orbital electrons) of active sites makes the antibonding state possess a lower energy level with more occupancy. This lowers the adsorption strength, increasing the likelihood of a “phys” form, and vice versa. In addition, the dissimilar “chem” configurations of CO₂ (O₅⁰–C₅³⁺–O₄⁻) can be partially attributed by the difference of electronegativity between the guest (“g”) and host (“h”) atom. If the electronegativity of “g” is smaller than “h,” it may induce obvious charge transfer from “g” to “h,” resulting in the positively charged guest atom and negatively charged host.
atom. Therefore, O atom of CO$_2$ may bond to “g” atom, and C linking to “h” atom, similar to “@”, =@, and =@ models of Figure 1b. Interestingly, the guest atom at the “@” model obviously deviates from the basal surface after the capture of the CO$_2$ molecule, such as Sc@Co, Y@Co, Zr@Co, and Y@Ni system. These may be contributed by the larger covalent radius of the molecule, such as Sc@Co, Y@Co, Zr@Co, and Y@Ni system.

Figure 2 shows a heat map of $E_{\text{ads}}$ for the most favorable site, which also encompasses the pristine surface (where the guest atom is the same as the host matrix on the SAAs), as also marked in the histogram form of Figure S6, Supporting Information. For Cu-based SAAs, our results are further supported by those of Lu et al.[13c] On SAAs with an identical host matrix, the absolute values of $E_{\text{ads}}$ decrease from ≈1.30–0.30 eV with increasing $\varphi$ ($\varphi \leq 6$) of the guest atom. Furthermore, with an increase in the $P$ (periodic number) of the guest atom, the bonding strength of CO$_2$ is also enhanced. Although the disparate guest atoms embedding into the same host surface may generate various adsorption characters, the uniform guest atom on the different host matrixes with diverse $\varphi$ would also lead to a variation in $E_{\text{ads}}$, for example, Y@Cu with an $E_{\text{ads}}$ of =−0.98 eV and Y@Co with an $E_{\text{ads}}$ of =−1.28 eV. This implies that the CO$_2$ adsorption properties are not only related to the $P$ and $\varphi$ values of the guest atom, but also to the $\varphi$ of the host atom. For a guest atom from the late TM series ($\varphi > 6$), the adsorption strength of CO$_2$ molecule is very low, with an $E_{\text{ads}}$ of ≈−0.20 eV. Consequently, the SAAs are more likely to retain a linear form of the free state, indicating that the involvement of a late TM has little effect on CO$_2$ activation.

Figure 2. Heat map of CO$_2$ adsorption energy ($E_{\text{ads}}$) on a) Co-, b) Ni-, c) and Cu-based SAAs. The horizontal and vertical axes represent the respective subgroup and $d$-orbital shell of the guest atom. The specific adsorption energy is also shown in units of eV. The significantly red and blue regions represent very strong and very weak CO$_2$ bonding, respectively. “/” indicates nonexistence.

To better elucidate the variation in $E_{\text{ads}}$ for different SAAs, we first investigate the correlation between $E_{\text{ads}}$ and the d-band center of the guest atom ($E_d$) in the entire database (Figure 3a). Compared with previous work, our method of calculating the d-band center that projects to the guest atom provides better correlations with other properties than calculating the d-band center bias from the basal surface, further enhancing the coordination unsaturation of $d$-orbitals. In all, besides the occupied number of $d$-orbitals $\varphi$, the electronegativity and covalent radius of the guest and host atoms may also synergistically regulate the adsorption configurations.

Figure 3. a) Relationship between the adsorption energy of CO$_2$ ($E_{\text{ads}}$) and d-band center ($E_d$) on SAAs. The black solid line represents the fitting line obtained using the chemical adsorption database. The fitting equation is $y = 0.26x + 0.76$ ($R^2 = 0.62$). The SAAs inside the green-highlighted region include “phys” and “chem” models. The inset shows the distribution of the prediction error with the chemical adsorption data using the d-band center model. b) Classification with “chem” (red) and “phys” (blue) using CVS. The black line represents the separation of convex domains using the support-vector machine with a linear kernel. The hollow and solid circles represent the training and test data, respectively.
projected on 1) the guest atom and its first adjacent host atoms or 2) the entire slabs.\cite{16} When $E_d$ is less than $\approx -2.0$ eV, the “phys” model is prevalent in all SAAs, with the quasi-linear form by an $E_{\text{ads}}$ of $\approx -0.2$ eV, while in the region above an $E_d$ of $\approx -0.5$ eV, the SAAs are likely to capture CO$_2$ molecules with the bent configuration. However, the SAAs inside the green-highlighted region of Figure 3a with $E_d$ values in the range of $-2.0$–$0.5$ eV exhibit various “phys” models, as well as a few “chem” models. Therefore, it is difficult to accurately distinguish the “phys” and “chem” configurations using the same $E_d$. If we adopt the entire chemical adsorption database, the changing trends for $E_{\text{ads}}$ and $E_d$ just match those of the d-band center theory qualitatively. The fitting correlation between $E_{\text{ads}}$ and $E_d$ yields $E_{\text{ads}} = -0.26E_d - 0.76$, with an $R^2$ of 0.62. The RMSE and MaxAE are 0.21 and 0.59 eV, respectively. The distribution of the prediction error is also shown in the inset of Figure 3a; it is slightly on the higher side and could hinder the precise design of the material. According to the previous d-band center model, the higher the energy level of $E_d$, the higher the binding energy of the CO$_2$ molecule. Nevertheless, it is evident that the provided linear correlation predicted by the d-band center model fails for CO$_2$ molecule uptake on SAAs. The model relations also do not hold for the adsorption of other species catalyzed by SAAs.\cite{16}

Thus, this standard and simple d-band center correlation does not hold for the adsorption of CO$_2$ on SAAs. Moreover, the calculation of $E_d$ for each SAA is highly computationally demanding, considering the numerous candidates. This necessitates the development of a novel accurate and low-cost descriptor for the high-throughput screening of material design in SAAs. The SISSO is a state-of-the-art method to identify descriptors and can acquire explicit, analytical functions between the material parameters and target properties. Table 1 and Table S1, Supporting Information elucidate the elementary properties of the host (h) and guest (g) atoms for SAAs to serve as primary features of the SISSO method, including the occupancy number of d-orbitals ($\psi$), periodic number (P), Pauling electronegativity ($\chi$), first ionization energy (IE), electron affinity (EA), covalent radius (R), and dipole polarizability (DP); in all, there are 14 base parameters. We then determine the Pearson correlation for some simple descriptors with $E_{\text{ads}}$, as the target parameter, based on the entire database, and subsequently discuss the best-performing, more-complicated descriptor and predictive power for data points not contained in the training dataset. Table S2, Supporting Information, shows the correlation of primary features ($\Phi_0$) with the $E_{\text{ads}}$ of our database. The top seven base parameters are all elementary properties of guest atoms; the most relevant is $\Phi_0$. This implies that the inclusion of the guest atom plays a crucial role in determining $E_{\text{ads}}$, especially $\Phi_0$, which has also proved to be a significant factor for the bonding strength.\cite{17} Table S3 and S4, Supporting Information, show the correlations for the top ten 1D descriptors identified by the first and second iterations of feature constructions ($\Phi_1$ and $\Phi_2$), respectively. Notably, these descriptors exhibit better correlation with $E_{\text{ads}}$ than the best primary feature ($\Phi_0$) and can be easily obtained via simple algebraic operations on the base parameters. For example, the best descriptor in the $\Phi_2$ feature construction is $\ln \frac{R}{\Phi} + \frac{\Phi}{\Phi}$ with a correlation coefficient of 0.9295, higher than 0.8875 for $\Phi_0$. The increase in correlation leads to a lower RMSE and MaxAE, from 0.18 eV and 0.41 eV to 0.14 eV and 0.35 eV, respectively. This demonstrates that a descriptor with just one base parameter is difficult to entirely capture the material property of interest, and a combination of primary features is required.

To accurately describe $E_{\text{ads}}$, using the provided elementary properties, the SISSO method is employed, with the aim of proposing a precise yet low-cost descriptor via a two-step classification and regression procedure. Initially, the classification method is adopted to distinguish the “phys” and “chem” results using CV5 (fivefold crossvalidation), as marked in Figure 3b. This figure also shows a black line, representing data calculated by the support-vector machine with a linear kernel, to help visualize the separation of convex domains. SISSO can successfully distinguish “phys” (blue) and “chem” (red) configurations using the 2D convex map with $|E_{A_h} - E_{A_g} - E_{A_0}|$ and $\frac{\Phi_i}{\psi_i}$ descriptors on the basis of the second-feature construction ($\Phi_2$). The fifth subsample was also verified to be distributed into both sides of the black line, which can be entirely sorted. As seen, the second descriptor $\frac{\Phi_2}{\psi_2}$ plays a crucial role in the classification, especially for the primary feature $\Phi_0$, due to the wide range from 1 to 10. Based on the aforementioned analysis, the electronegativity $\chi$ and covalent radius $R$ may also synergistically influence the “chem” and “phys” form. If we use the entire database without CV5, we will also obtain the same descriptors to classify the “phys” and “chem” results perfectly in Figure S8, Supporting Information. This signifies that the clear and analytical expression provided by the SISSO approach has better generalization ability. Thus, the SISSO can accurately achieve the classification of the CO$_2$ adsorption model for SAAs.

The SISSO method can not only be used to identify explicit, analytical descriptors for target property classification, but also for quantitatively evaluating the $E_{\text{ads}}$ for SAAs via regression analysis based on the “chem” database. This dataset is divided into a training set (80% of the database) and test set (20% of the database), where descriptor identification is performed with CV5 using the training set. The validation error is defined as the average of the test errors obtained for each of the five subsamples in the training set, and the optimal descriptor is the formula with the least validation RMSE. Figure 4a and Table S5, Supporting Information, display the resulting training and validation RMSE for each combination of hyperparameters. Both, the training and validation errors, are generally reduced; then, when implementing higher dimensions and larger rungs, overfitting may also

Table 1. Elementary properties of host and guest atoms as primary features of the SISSO method. Note that the SAAs include host and guest atoms, with a total of 14 features.

| Abbreviation | Name                           | Unit         |
|--------------|--------------------------------|--------------|
| $\psi$       | Occupation number of d orbital | dimensionless|
| R            | Covalent radius                | Å            |
| DP           | Dipole polarizability          | C m$^2$/V    |
| IE           | Ionization energy              | eV           |
| $\chi$       | Pauling electronegativity      | Dimensionless|
| EA           | Electron affinity              | eV           |
| P            | Periodic number                | Dimensionless|
occur as evidenced by a flat or even slightly increased RMSE. The optimal dimensionality of the descriptor is 6D, with a validation RMSE of 0.05 eV and MaxAE of 0.21 eV in the $\Phi_2$ feature space. The corresponding descriptor and its correlation with $E_{\text{ads}}$ are shown in Table S5, Supporting Information. Although the 6D $\Phi_2$ descriptor possesses the best accuracy, it is extremely complex and not convenient for high-throughput screening, especially for experimentalists working on material design. In the $\Phi_2$ feature space, the 1D descriptor is $P_g/C_3(\phi_g/C_0\phi_h)$ with a validation RMSE of 0.12 eV and MaxAE of 0.30 eV. This descriptor is relevant to the periodic number of the guest atom ($P_g$) and occupation number of $d$-orbitals for the guest ($\phi_g$) and host atoms ($\phi_h$). This 1D $\Phi_2$ descriptor offers specific interpretability and is convenient to derive the physical significance. Interestingly, the explicit, analytical function can also be obtained using all chemical adsorption energy data as training data in Figure S9, Supporting Information, indicating the generalization ability of this descriptor. This descriptor can also support our fundamental understanding of the bonding strength on SAA surfaces.

For the 6D and 1D $\Phi_2$ descriptors, we also display a violin plot of the absolute error distribution using the remaining test set and compare it with that obtained using the traditional $d$-band center descriptor with the same dataset in Figure 4b. Most errors (75%) are below 0.14 eV for the 6D and 0.26 eV for the 1D descriptors, and the MaxAEs are only 0.20 eV and 0.27 eV, respectively. In comparison with $E_d$, we also perform the corresponding CV5 using 80% of the database as a training set to obtain the linear correlation of $E_{\text{ads}}$ with $E_d$, leading to an average validation RMSE of 0.21 eV and MaxAE of 0.57 eV, which are higher than those with the 6D and 1D $\Phi_2$ descriptors. In addition, most errors for the test set are below 0.30 eV (upper blue density plot of Figure 4b), which is thus poorer than the 6D $\Phi_2$ as well as 1D $\Phi_2$ descriptors. In addition, most errors for the test set are below 0.30 eV (upper blue density plot of Figure 4b), which is thus poorer than the 6D $\Phi_2$ as well as 1D $\Phi_2$ descriptors. In Figure 4c–e and Figure S10, Supporting Information, we compare the DFT-calculated versus SISSO-predicted $E_{\text{ads}}$ using $\Phi_2$ 6D and $\Phi_2$ 1D, and the $d$-band center model-predicted $E_{\text{ads}}$. The hollow and solid circles represent the training and test data, respectively.
3. Conclusion

In summary, we propose a novel model that is accurate and reliable to describe the CO$_2$ adsorption energy on SAAs using the two-step consecutive classification and regression SISSO approach, combining first-principles calculations. SISSO descriptors can not only successfully distinguish the “phys” and “chem” configurations, but also provide an explicit, analytical expression for the correlation of $E_{\text{ads}}$ and atomic information as base parameters, based on the “chem” database. The validations RMSE and MaxAE of the 1D descriptor are 0.12 and 0.30 eV, respectively, as well as the best descriptor with 0.05 eV for RMSE and 0.21 eV for MaxAE, all lower than those of the traditional $d$-band center model (0.21 and 0.57 eV, respectively). Therefore, the SISSO-generated descriptor distinctly outperforms that obtained by the $d$-band center model in terms of accuracy and computational overhead. Our methodology may also offer a rational strategy to evaluate the adsorption energies for the computational screening of catalyst materials.

4. Experimental Section

DFT Calculations: All calculations were performed by the spin-polarized DFT method, implemented using the Vienna Ab-initio Simulation Package (VASP)$^{[18]}$ with the projector-augmented wave (PAW) pseudopotential.$^{[19]}$ The electronic exchange-correlation interactions were treated by the generalized gradient approximation (GGA)$^{[20]}$ Perdew-Burke-Ernzerhof (PBE) functional.$^{[20]}$ To accurately evaluate the weak interaction between the adsorbed molecule and surface, DFT + D3 corrections$^{[21]}$ were performed during the adsorption of the CO$_2$ molecule on the SAAs surfaces. The kinetic energy cutoff with a planewave basis was set to 400 eV. Three metals with dissimilar occupation numbers of the $d$-orbital, that is, Co (d7) with the hcp structure, Ni (d8) with the fcc structure, and Cu (d10) with the fcc structure, were selected as host matrixes of the SAAs. These metals usually served as reactive catalyst materials in the CO$_2$ hydrogenation process.$^{[13]}

The corresponding close-packed surfaces exposing Co (001), Ni (111), and Cu (111) were modeled using a periodically repeating (3 × 2√3) four-layer supercell, in which the bottom layer was fixed to the bulk position, whereas the other layers and adsorbed molecule were relaxed until the total energy and residual force per atom were less than 10$^{-5}$ eV and 0.02 eV Å$^{-1}$, respectively. Subsequently, 15 Å vacuum layers were added to the metal slabs, and 28 kinds of TM elements, except Hg (which is toxic) from the periodic table, were chosen as the guest atoms of the SAAs. The first-Brillouin-zone integration used (4 × 5 × 1) Monkhorst–Pack k-point sampling for geometrical optimization and the (6 × 7 × 1) k-point mesh for precise electronic structure simulations. The $E_{\text{ads}}$ values for the SAAs were calculated using the equation $E_{\text{ads}} = E_{\text{total}} - E_{\text{surf}} - E_{\text{CO}_2}$, where $E_{\text{total}}$ represents the total energy of SAAs with the CO$_2$ molecule, $E_{\text{surf}}$ indicates the total energy of the SAAs, and $E_{\text{CO}_2}$ is the energy of the CO$_2$ molecule. A more negative value indicated a stronger bonding interaction for CO$_2$ molecule uptake. The $d$-band center projected to the single guest atom ($E_d$) in the SAAs was calculated by the following formula.$^{[22]}$

$$E_d = \int_{-\infty}^{\infty} \epsilon \cdot \text{dos} \cdot d\epsilon$$

where dos denotes the density of states at the specific energy level and $\epsilon$ represents the energy level.

Descriptor Identification: All descriptor identifications were performed using version 3.0.2 of the SISSO algorithm, which was recently developed by Ouyang et al.$^{[14]}$ The basis of this approach was the preselection of primary features ($\Phi_1$) in Table 1, that is, the occupation number of the $d$-orbital ($\phi$), periodic number ($P$), covalent radius ($R$), or Pauling electronegativity ($\chi$), which were believed to play key roles in describing the property of interest. A set of algebraic/functional operators ($H^{(m)}$) was then combined for constructing the feature spaces of $\Phi_1$ and $\Phi_2$.

$$H^{(m)} = \{+, -, \times, \div, \exp, \log, \sqrt{\cdot}, \sqrt[3]{\cdot}, \sqrt{\cdot}, \sqrt[3]{\cdot}, -\}$$

where the superscript $m$ denotes the application of $H^{(m)}$ to primary features $\Phi_1$ and $\Phi_2$ (a dimensional analysis was performed in this case), ensuring that only physically meaningful combinations were maintained (only primary features with the same unit were added or subtracted). Finally, the SO was employed for descriptor identification in the smaller feature subspace selected by a screening process (SIS), where the size of the feature subspace was equal to the SIS value multiplied by the dimension of the descriptor. The descriptors were represented as linear combinations of features, where the features were nonlinear functions of primary features when employing algebraic/functional operators. To mitigate overfitting concerns during training, CV5 was performed in classification and regression using the SISSO approach. The entire database was shuffled and partitioned into five equally sized subsets; descriptor identification used one of the four subsets as the training set, and the remaining database served as a validation set to measure the validity of the descriptor.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors acknowledge funding support from National Natural Science Foundation of China (grant nos. 11674237 and 51602211), National Key Research and Development Program of China (grant no. 2016YF0700700), National Natural Science Foundation of Jiangsu Province of China (grant no. BK20160299), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). The theoretical work was carried out at National Supercomputer Center in Tianjin and the calculations were performed on TianHe-1(A).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Keywords

adsorption energies, CO$_2$ molecules, data-driven analytic approaches, descriptor identifications, single-atom alloys

Received: September 8, 2021
Revised: October 28, 2021
Published online: November 23, 2021

$^{[1]}$ a) A. Bruix, J. T. Margraf, M. Andersen, K. Reuter, Nat. Catal. 2019, 2, 659; b) W. T. Hong, M. Risch, K. A. Storerzinger, A. Grimaud, J. Suntivich, Y. Shao-Horn, Energy Environ. Sci. 2015, 8, 1404;
