



OPEN Interpretable, physics-informed learning reveals sulfur adsorption and poisoning mechanisms in 13-atom icosahedra nanoclusters

Raiane Ferreira Monteiro¹, João Marcos T. Palheta¹, Tulio Gnoatto Grison¹, Octávio Rodrigues Filho¹, Renato Luis Tame Parreira², Diego Guedes-Sobrinho³, Celso R. C. Rêgo⁴✉, Alexandre C. Dias⁵, Krysz Elly de Araújo Batista⁶ & Maurício J. Piotrowski¹

Transition-metal nanoclusters exhibit structural and electronic properties that depend on their size, often making them superior to bulk materials for heterogeneous catalysis. However, their performance can be limited by sulfur poisoning. Here, we use dispersion-corrected density functional theory (DFT) and physics-informed machine learning to map how atomic sulfur adsorbs and causes poisoning on 13-atom icosahedral clusters from 30 different transition metals (3d to 5d). We measure which sites sulfur prefers to adsorb to, the thermodynamics and energy breakdown, changes in structure, such as bond lengths and coordination, and electronic properties, such as ϵ_d , the HOMO-LUMO gap, and charge transfer. Vibrational analysis reveals true energy minima and provides ZPE-based descriptors that reflect the lattice stiffening upon sulfur adsorption. For most metals, the metal-sulfur interaction mainly determines adsorption energy. At the same time, distortion contributions are generally moderate, but become large in magnitude for a few metals, suggesting a stronger tendency toward adsorption-induced restructuring. Using unsupervised *k*-means clustering, we identify periodic trends and group metals based on their adsorption responses. Supervised regression models with leave-one-feature-out analysis identify the descriptors that best predict adsorption for new samples. Our results highlight the isoelectronic triad Ti, Zr, and Hf as a balanced group that stands out by combining moderate-to-strong sulfur binding with excellent structural stability. This combination suggests an optimal trade-off between the chemical activation of sulfur-containing species and resistance to poisoning-induced structural degradation. Additional DFT calculations for SO₂ adsorption reveal strong binding and a clear tendency toward dissociation on these clusters, linking electronic states, lattice response, and poisoning strength. These findings offer data-driven guidelines for designing sulfur-tolerant nanocatalysts at the subnanometer scale.

Keywords Nanoclusters, Transition-Metals, Density Functional Theory, Physics-Informed Learning, Sulfur Poisoning

The challenges associated with developing, manipulating, and understanding nanoscale atomistic systems have drawn significant attention from the research community due to fast advances in nanoscience^{1–3}. In the subnanometer regime, where quantum size effects become dominant, the nanostructures made of a few tens to several hundred atoms often show properties that differ from those of their bulk crystalline counterparts^{4,5}. This emergent behavior enables the materials design with tunable structural and electronic features, offering many opportunities in the energy conversion, optoelectronic, and catalytic fields^{6,7}. Transition metal (TM) nanoclusters (NCs) stand out in this context because of their distinct chemical reactivity, discrete electronic states, and high surface/volume ratios^{8,9}. The intrinsic fluxionality, finite size, and lack of translational symmetry characteristics result in a broad configurational landscape with many non-equivalent adsorption sites^{10–12}. As

¹Department of Physics, Federal University of Pelotas, Pelotas, RS 96010-900354, Brazil. ²Núcleo de Pesquisas em Ciências Exatas e Tecnológicas, Universidade de Franca, Franca, SP, Brazil. ³Quantum Chemistry and Thermodynamic Materials Group, Department of Chemistry, Federal University of Paraná, 81531-980 Curitiba, Brazil. ⁴Institute of Nanotechnology Hermann-von-Helmholtz-Platz, Karlsruhe Institute of Technology (KIT), 76344, Eggenstein-Leopoldshafen, Germany. ⁵Institute of Physics and International Center of Physics, University of Brasília, Brasília 70919-970, DF, Brazil. ⁶Federal Institute of Amazonas, Manaus 69086-475, AM, Brazil. ✉email: celso.rego@kit.edu

a result, TM-NCs exhibit a complex, yet highly adaptable catalytic behavior, which is particularly important in real-world scenarios (realistic environmental conditions) where toxic poisoning species and/or competing reaction pathways may be present¹³.

The catalytic deactivation caused by S-containing species is a long-standing dilemma in heterogeneous catalysis¹⁴. This challenge arises because sulfur atoms and sulfur-bearing molecules, such as H₂S, SO₂, and SO_x, show strong affinity for metal active sites, frequently leading to partial or complete catalytic activity loss^{15,16}. Although adsorption mechanisms on extended surfaces are reasonably well disseminated/understood, the behavior of S adsorption on subnanometer clusters remains far less explored. This knowledge gap is mainly due to pronounced finite-size effects, enhanced fluxionality, and the low coordination of surface atoms in TM-NCs^{17,18}. Consequently, developing robust nanocatalysts requires a fundamental understanding of S–NC interactions at the atomistic scale¹⁹.

In TM-NCs, reactivity is governed by the electronic structure, which deviates from that of bulk metals due to quantum confinement and discrete energy levels²⁰. In this line, systematic studies have shown that adsorption energies, preferred adsorption sites, and even reaction mechanisms can vary significantly across isomers separated by only a few tenths of an electronvolt²¹. For specific geometrical motifs, such as the icosahedron (ICO), for example, the interplay between geometric shell closure and *d*-band occupation creates interesting periodic trends in adsorption behavior across the TM series^{22,23}. Thus, understanding how S interacts with these electronically diverse motifs provides valuable insights into the design of more resilient catalytic nanostructures.

Previous experimental and theoretical studies have pointed out the critical impact of S poisoning on metal nanoparticles and NCs¹⁴. For instance, studies on Au, Pd, and Pt NCs have demonstrated that S can induce substantial structural rearrangements, change the *d*-band center, and even lead to NC fragmentation at elevated coverages^{20,24}. On the other hand, another work on subnanometer clusters shows that S adsorption may stabilize unexpected geometries²⁵. In addition to recent machine learning (ML) advances, which have accelerated materials discovery, enabling large-scale exploration of NC potential-energy landscapes^{26–28}, the ML models trained on first-principles (density functional theory-DFT) data have been effective in identifying periodic trends, predicting adsorption energies, and capturing subtle structure, and property relationships in metal NCs²⁹. By incorporating these approaches (DFT+ML) into our calculations and analysis, we not only accelerate the screening of TM–S interactions but also uncover systematic behaviors that would be difficult to detect through DFT alone, further strengthening the predictive power of nanocatalyst design.

Here, we investigate sulfur adsorption and incipient poisoning on 13-atom icosahedral transition-metal nanoclusters (TM₁₃) spanning the 3*d*, 4*d*, and 5*d* series. The icosahedral (ICO) motif is among the most stable geometries in the subnanometer regime^{30–33}, consisting of a central atom surrounded by 12 surface atoms at the vertices of a regular ICO, and providing a high-symmetry, close-packed platform with well-defined adsorption sites (top, bridge, hollow). Using dispersion-corrected, spin-polarized DFT, we characterize pristine and S-adsorbed ICO clusters through a set of physically motivated descriptors that span energetics (e.g., E_{ads} and its interaction/distortion decomposition), electronic structure (e.g., ε_d , E_{gap} , charge transfer), and lattice dynamics (vibrational fingerprints and ZPE-based descriptors). We then introduce a physics-informed interpretation of sulfur adsorption and poisoning trends by combining these descriptors with interpretable machine learning (ML): unsupervised clustering maps systematic, periodic-table-level similarities in adsorption response, while supervised, model-agnostic LOFO analyses quantify which descriptors contribute robustly to out-of-sample prediction and thus to mechanistic interpretation. This integrated DFT+ML framework yields transferable rules linking adsorption strength to electronic-state alignment and adsorption-induced lattice response. It motivates the selection of the isoelectronic triad Ti₁₃, Zr₁₃, and Hf₁₃ as representative, chemically resilient case studies for explicit SO₂ adsorption, where follow-up DFT provides direct validation of the ML-guided trends.

Methodology and computational details

Density Functional Theory: Our calculations were performed using spin-polarized DFT^{34,35} as implemented in the Vienna *Ab Initio* Simulation Package (VASP)^{36,37}, employing the projector augmented-wave (PAW) method^{38,39} to describe the electron–ion interactions. The exchange–correlation energy was described by the generalized gradient approximation (GGA) with the Perdew–Burke–Ernzerhof (PBE) functional^{40,41}, including dispersion corrections via the empirical D3 corrections, as proposed by Grimme^{42–44}, to include the attractive non-local long-range van der Waals (vdW) interactions. In addition, spin–orbit coupling (SOC) was included and tested, which could be particularly important for 4*d* and 5*d* elements, where relativistic effects may be significant. Otherwise, in our approach, the valence electrons were treated using a scalar-relativistic approximation, while the core electrons were treated fully relativistically^{45,46}.

The TM-NC systems (for pristine and atomically/molecularly adsorbed cases) were modeled from an isolated 13-atom ICO, i.e., as non-periodic systems through a cubic box (supercell) with a side length of 20 Å, providing a minimum separation distance of approximately 12 Å, which is enough to avoid the spurious self-interactions among the system and their periodic images. All systems were fully optimized until residual forces were below 0.01 eVÅ^{−1}, using a plane-wave cutoff energy of 500 eV and an energy convergence threshold of 10^{−6} eV in the self-consistent cycle. For density of states (DOS), we used a cutoff energy value that was 1.5 times higher. For Brillouin zone integration, a single **k**-point, specifically, the Γ -point, was used for all calculations. A small Gaussian smearing parameter of 1 meV was applied to prevent fractional occupation of electronic states.

The reference energies of the isolated atoms entering Eqs. (1) and (2) were obtained from single-atom calculations performed in large cubic supercells using spin-polarized DFT and Γ -point sampling. For open-shell atoms, finite initial magnetic moments were assigned, and the electronic states were allowed to relax without symmetry constraints. The resulting reference energies were taken from the lowest-energy converged spin-polarized solutions, consistent with the expected atomic high-spin occupations and Hund-type filling rules.

This procedure ensures that the free-atom references are physically appropriate for the binding-energy and adsorption-energy definitions adopted here.

It is important to mention that the energetic numerical precision near the equilibrium geometry calculated with the VASP package is on the order of 0.001 eV per atom⁴⁷. Furthermore, the PBE functional has no fitted parameters, which represents an advantage in studying NCs, since for these systems there is reduced experimental data. The PBE accuracy is 0.28 eV when measured relative to a known isomerization-energy dataset, considered difficult for theory⁴⁸. In our case, we expect better accuracy, since we are comparing relative energies within families of related NC species. Finally, the vibrational frequency (ν) calculations were performed to determine the $3N - 6$ vibrational modes for all systems, for which the Hessian matrix was calculated using finite differences, as implemented in VASP, considering the displacing of each atom in each direction by ± 0.01 Å. Further computational details, including convergence tests, are provided in the Supporting Information (Tables S1–S4).

In the last part of this work, after the selection process, to obtain the most stable SO₂ adsorption site on Ti₁₃, Zr₁₃, and Hf₁₃ NCs, we explored the analyte-substrate potential energy surface employing a strategy based on *Ab Initio* Molecular Dynamics (AIMD) simulations and additional initial configurations built by design principles (to complement the AIMD calculations). The AIMD simulations started with SO₂ and NC atomic configurations in arbitrary orientations, followed by a thermalization process at 300 K, using a Nosé-Hoover thermostat within an NVT ensemble^{49,50}. Our AIMD simulations were run for 5 ps, with a time step of 1 fs, to generate the final adsorbed snapshot, which was subsequently structurally optimized through DFT-PBE+D3 calculations.

Our simulation workflow follows the FAIR (Findable, Accessible, Interoperable, Reusable) and TRUE (Transparent, Reproducible, Usable by Others, and Extensible) data principles^{51–53}. Here, we ensure that the complete set of data and methods produced in this study is open, reproducible, and straightforward for others to reuse or build upon. In line with these guidelines, we provide full provenance of data sources, document uncertainties where applicable, and include clear reuse instructions^{54–58}. All DFT inputs and outputs, processed descriptor tables, machine-learning workflows, trained models, and analysis scripts generated in this work are openly available in the public repository <https://github.com/KIT-Workflows/Physics-Informed-Nanoclusters>, enabling direct reuse, validation, and extension by the community.

Atomic Configurations: We have considered the following systems: 13-atom ICO NCs of all TM elements in (i) pristine case (bare clusters, TM₁₃), in a first-order approximation, and (ii) atomic-sulfur-adsorbed NC case (S/TM₁₃), in a second-order approximation; after that, TM selected 13-atom ICO NCs in (iii) molecular-sulfur-based-adsorbed NC case (SO₂/TM₁₃), as illustrated schematically in Fig. 1. As well as the respective free atom calculations of all these systems and the isolated SO₂ molecule calculation. Therefore, it is evident that

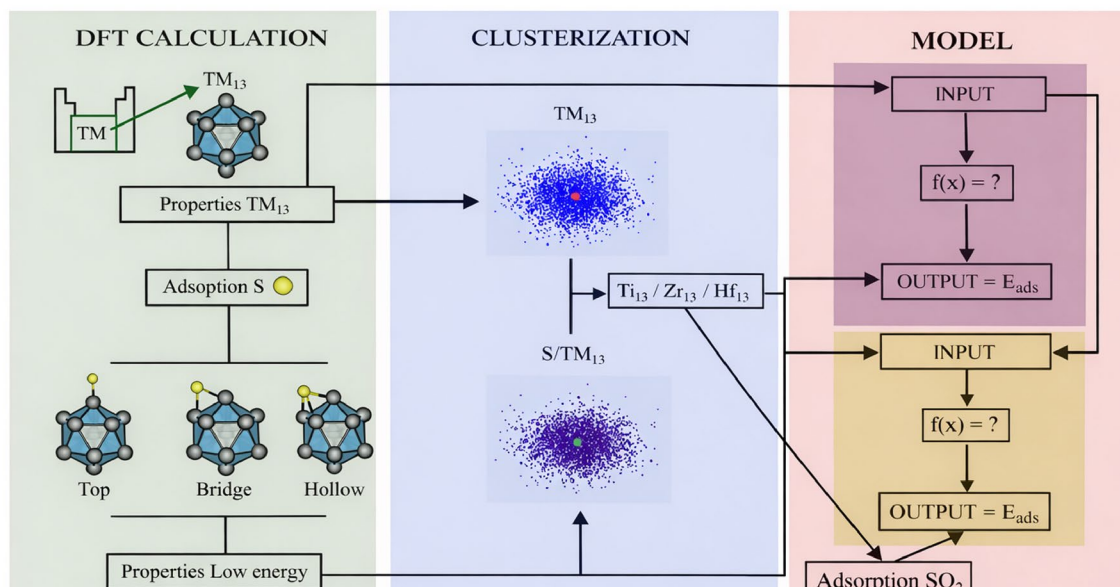


Fig. 1. Schematic workflow illustrating the combined first-principles and machine-learning strategy adopted in this work. Starting from DFT calculations, 13-atom ICO TM-NCs in pristine form (TM₁₃) and after atomic S adsorption (S/TM₁₃), considering top, bridge, and hollow binding sites. Structural, energetic, electronic, and vibrational descriptors extracted from DFT calculations were standardized and used as inputs to unsupervised clustering analyses with *k*-means and principal component analysis (PCA), enabling the identification of chemically similar groups across the TM series. Based on the clustering outcomes, selected nanoclusters (Ti₁₃, Zr₁₃, and Hf₁₃) were further employed in explicit SO₂ adsorption calculations. The descriptor set was used in interpretable supervised regression models (ElasticNetCV, RidgeCV, and Explainable Boosting Machine, EBM), allowing the prediction and rationalization of S adsorption energetics and supporting the selection of chemically resilient NCs.

our main focus is to test the composition (among all TM elements) rather than geometric motif; consequently, the structural configuration considered for our 13-atom NCs of all TM elements are initially set to the ICO, which is the lowest energy configuration for several TM_{13}NCs ^{33,59}. The ICO motif is a quasi-spherical particle, with close-packed high-symmetry (I_h), composed by a central atom and 12 equidistant atoms. All ICO-based calculations performed here had their geometries relaxed, without any symmetry or spin constraints, to allow structural and magnetic symmetry breaking and a more complete exploration of the potential energy surface.

As mentioned, spin polarization was included in all calculations. No constraints on the total magnetic moment were imposed, and the electronic structure was allowed to relax self-consistently starting from the default initial magnetic moments provided by VASP. We note that although this approach captures stable, self-consistent spin solutions, it does not guarantee the identification of the global spin ground state, particularly for open-shell transition-metal nanoclusters, where multiple spin configurations may be close in energy.

The two main sets of configurations: TM_{13} and S/TM_{13} are showed in Supporting Information (Figure S1). The first set (TM_{13}) is used as a substrate for the adsorption of S atoms. We explored all non-equivalent positions, considering an approach similar to that used for low-Miller-index surfaces, i.e., onefold (top), twofold (bridge), and threefold (hollow) sites on NCs. After a relative total-energy analysis of the adsorption sites, we obtained the S/TM_{13} set. Consequently, we systematically considered all non-equivalent sites with distinct coordination (with respect to the ICO motif) and compositions (with respect to all TM elements from the periodic table). Then, after our selection procedure, we have chosen three TM-NCs for the last calculation step, i.e., $\text{SO}_2/\text{TM}_{13}$. For this case, exploring the surface potential energy through AIMD simulations and placing the SO_2 molecule in all non-equivalent positions on NCs, considering trivial sites and specific cases with side-on and end-on orientations. It is important to note that AIMD simulations provide a more complete set of adsorption configurations and allow testing of their thermal stability.

Property Analysis: To characterize the energetic stability of the ICO NCs, we calculated the binding energy per atom, $E_b^{\text{TM}_{13}}$, as

$$E_b^{\text{TM}_{13}} = \frac{E_{\text{tot}}^{\text{TM}_{13}} - 13E_{\text{tot}}^{\text{TM free-atom}}}{13}, \quad (1)$$

where $E_{\text{tot}}^{\text{TM}_{13}}$ is the total energy of the NC and $E_{\text{tot}}^{\text{TM free-atom}}$ the energy of each isolated atomic species. This metric enables comparison of relative stabilities across different TM-NCs. In analogy, to characterize the energetic stability of the S/TM_{13} systems, we calculated the binding energy per atom, $E_b^{\text{S}/\text{TM}_{13}}$, as

$$E_b^{\text{S}/\text{TM}_{13}} = \frac{E_{\text{tot}}^{\text{S}/\text{TM}_{13}} - E_{\text{tot}}^{\text{S free-atom}} - 13E_{\text{tot}}^{\text{TM free-atom}}}{14}, \quad (2)$$

where $E_{\text{tot}}^{\text{S}/\text{TM}_{13}}$ is the total energy of the S/TM_{13} and $E_{\text{tot}}^{\text{S free-atom}}$ the energy of sulphur atomic species.

The adsorption energy (E_{ads}) of an atomic or molecular species ($X = \text{S}$ or SO_2) was defined as

$$E_{\text{ads}} = E_{\text{tot}}^{X/\text{TM}_{13}} - E_{\text{tot}}^{\text{TM}_{13}} - E_{\text{tot}}^X, \quad (3)$$

where $E_{\text{tot}}^{X/\text{TM}_{13}}$, $E_{\text{tot}}^{\text{TM}_{13}}$, and E_{tot}^X are the total energies of the adsorbed system, pristine NC, and isolated adsorbate, respectively. To gain further insight into the adsorption mechanism, E_{ads} was decomposed into interaction and distortion contributions^{60–62}:

$$E_{\text{ads}} = \Delta E_{\text{int}} + \Delta E_{\text{dis}}^{\text{TM}_{13}} + \Delta E_{\text{dis}}^X, \quad (4)$$

where ΔE_{int} is the interaction energy between frozen fragments, disregarding the energetic contribution from structural distortions, which is given by

$$\Delta E_{\text{int}} = E_{\text{tot}}^{X/\text{TM}_{13}} - E_{\text{tot}}^{\text{TM}_{13} \text{ frozen}} - E_{\text{tot}}^{X \text{ frozen}}, \quad (5)$$

while $\Delta E_{\text{dis}}^{\text{TM}_{13}}$ and ΔE_{dis}^X denote the distortion penalties of the NC and adsorbate, respectively, given by

$$\Delta E_{\text{dis}}^{\text{TM}_{13}} = E_{\text{tot}}^{\text{TM}_{13} \text{ frozen}} - E_{\text{tot}}^{\text{TM}_{13}}, \quad (6)$$

and

$$\Delta E_{\text{dis}}^X = E_{\text{tot}}^{X \text{ frozen}} - E_{\text{tot}}^X, \quad (7)$$

where $E_{\text{tot}}^{\text{TM}_{13} \text{ frozen}}$ and $E_{\text{tot}}^{X \text{ frozen}}$ are the total energies of the frozen NC and X at their original positions within the X/TM_{13} system without X and TM_{13} parts, respectively.

These values quantify the energetic effect associated with adopting the adsorption-induced fragment geometries; consequently, the ΔE_{dis}^X term is null for $X = \text{S}$. In the present work, $\Delta E_{\text{dis}}^{\text{TM}_{13}}$ is defined as the energy difference between the pristine nanocluster and the nanocluster geometry obtained after removing the adsorbate from the relaxed adsorbed configuration while preserving the adsorption-induced structure as a frozen fragment. Positive values, therefore, correspond to an energetic penalty associated with deforming the pristine nanocluster into the adsorption-induced geometry. Negative values indicate that the adsorption-induced bare-cluster geometry is lower in energy than the original pristine reference adopted in this study. Because our

analysis focuses on composition trends within a common ICO motif, with all TM₁₃ systems initialized from the icosahedral structure and subsequently relaxed, such negative values are interpreted as adsorption-assisted reconstruction toward a geometry more favorable than the original ICO-derived reference.

We have obtained the $(3N - 6)$ vibrational frequencies (ν) by calculating the dynamic Hessian matrix elements within the finite difference method implemented in VASP. This approach involved two atomic displacements, with each atom moved in each direction (positive and negative) by 0.01 Å, allowing the numerical construction of second derivatives of the total energy with respect to atomic displacements. The resulting Hessian matrix was subsequently diagonalized to yield the vibrational eigenmodes and their corresponding frequencies. The vibrational analysis has multiple purposes. (i) The absence of imaginary frequencies confirms that all optimized pristine and S or SO₂-adsorbed NCs correspond to true local minima on the potential energy surface. (ii) The complete vibrational spectra provide a characteristic fingerprint for each system, enabling detailed comparisons of structural motifs and adsorption-induced changes. (iii) In the specific case of the isolated SO₂ molecule, the computed vibrational frequencies were validated against available experimental and theoretical data, ensuring the reliability of the adopted computational protocol. These reference values also allow direct comparison between gas-phase and adsorbed SO₂, helping elucidate frequency shifts associated with charge transfer, bond weakening or strengthening, and/or symmetry breaking upon adsorption on the TM-NCs. (iv) Finally, the minimum and maximum vibrational frequencies, together with the zero-point energy (ZPE), were systematically extracted for all systems. The ZPE was computed as

$$E_{\text{ZPE}} = \frac{1}{2} \sum_i h\nu_i, \quad (8)$$

where the sum runs over all vibrational modes. These vibrational descriptors were subsequently incorporated into the ML model as physically motivated features, complementing structural, electronic, and energetic inputs.

Structural descriptors included the average bond length, d_{av} , and effective coordination number, ECN, obtained from the effective coordination concept^{63,64}. In addition to calculating d_{av} and ECN for pristine systems, to quantify structural changes induced by adsorption, we also computed relative deviations as

$$\Delta d_{\text{av}} = \frac{(d_{\text{av,ads}} - d_{\text{av}}) \times 100}{d_{\text{av}}}, \quad (9)$$

and

$$\Delta \text{ECN} = \frac{(\text{ECN}_{\text{ads}} - \text{ECN}) \times 100}{\text{ECN}}, \quad (10)$$

where the values of $d_{\text{av,ads}}$ and ECN_{ads} are obtained after adsorption, with molecules subsequently removed from the analyses. Thus, Δd_{av} and ΔECN parameters provide a direct measure of NC expansion/contraction, distortions, and coordination changes upon adsorption. We also considered the minimum NC-adsorbate distance, i.e., $d_{\text{TM-S}}$, using VESTA software⁶⁵.

Electronic structure analyses included projected DOS (PDOS) calculations and charge redistribution analysis. From PDOS, we have obtained the center of gravity of the occupied d -states, ε_d , which is derived from the d -band model⁶⁶. It provides the energetic position of the metal d -states relative to the Fermi level, and is related to the adsorption energy magnitude of an adsorbate on the TM systems. Consequently, a ε_d closer to the Fermi energy (0.0 eV), implies stronger hybridization between metal d -states and adsorbate orbitals, leading to enhanced adsorption strength. Additionally, we evaluated the HOMO-LUMO energy gap (E_{gap}) of the NCs, defined as the energy difference between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO). It is a direct measure of the electronic hardness and chemical reactivity of finite systems, with smaller gaps typically associated with higher polarizability, enhanced charge-transfer capability, and increased chemical activity.

The charge findings, including charge redistribution effects, were evaluated by Bader charge analysis^{67,68}, in which the total electron density is partitioned into Bader volumes (atomic basins) defined by zero-flux surfaces of the charge-density gradient, $\nabla n(\mathbf{r})$. The net charge on atom α is given by

$$Q_{\alpha}^{\text{Bader}} = Z_{\alpha} - \int_{V_{\alpha}} n(\mathbf{r}) d^3 r, \quad (11)$$

where Z_{α} is the valence charge of atom α and V_{α} its corresponding Bader volume. Together, these energetic, vibrational, structural, and electronic descriptors capture the interplay between electronic stability, reactivity, and adsorption strength, enabling a comprehensive analysis of the interaction mechanisms between TM-NCs and S-containing adsorbates.

For transparency, reproducibility, and ease of interpretation^{51,69,70}, all descriptors used in the ML analyses are explicitly listed in Table S6a and Table S6b of the Supporting Information, together with their definitions, units, and the subset of analyses in which they were employed (unsupervised, supervised, or both). In brief, the descriptor space combines energetic quantities (E_b , E_{ads} , ΔE_{int} , ΔE_{dis}), structural quantities (d_{av} , ECN, Δd_{av} , ΔECN , $d_{\text{TM-S}}$), vibrational quantities (ν_{min} , ν_{max} , E_{ZPE}), and electronic quantities (ε_d , E_{gap} , and Bader-charge-based measures of charge redistribution).

Machine Learning Models: To gain more insight about our system, we are applying a set of Machine-learning (ML) techniques to (i) rationalize periodic trends in atomic S adsorption across the set of 30 TM₁₃

ICO nanoclusters and (ii) guide the selection of representative systems for subsequent explicit SO₂ adsorption studies. Given the limited dataset size, our ML strategy prioritizes physical interpretability, robustness under cross-validation, and consistency with first-principles descriptors over maximizing raw predictive performance.

We derived a set of descriptors from DFT calculations to describe both pristine TM₁₃ nanoclusters and their S-adsorbed forms (S/TM₁₃). These features include energetic, structural, vibrational, and electronic properties related to adsorption and nanocluster stability. Examples are per-atom binding energy E_b , structural measures like d_{av} and ECN, vibrational values such as ν_{min} , ν_{max} , and E_{ZPE} , and electronic descriptors like ε_d , Bader charge, and E_{gap} . We also included adsorption-related energies (E_{ads} , ΔE_{int} , ΔE_{dis}) as descriptors to decompose the adsorption process into interaction and structural distortion contributions, enabling a more transparent interpretation of adsorption trends across the TM₁₃ nanoclusters. Before applying machine learning, we standardized all numerical descriptors using the `StandardScaler` from `SCIKIT-LEARN`⁷¹ so that features with different units would contribute equally in both distance-based and regularized regression analyses.

The unsupervised and supervised analyses were carried out on related, but not identical, data representations. For the unsupervised analysis, we considered standardized descriptor matrices for pristine TM₁₃ and for optimized S/TM₁₃ systems separately, to compare how sulfur adsorption reorganizes the descriptor space. For the supervised regression task, the target variable was the atomic-sulfur adsorption energy E_{ads} . In contrast, the predictive inputs were restricted to pristine-cluster descriptors, so that the models probe how much of the adsorption response can be anticipated from the properties of the clean nanoclusters alone.

Because the dataset contains only 30 transition-metal systems and strong chemical correlations exist within each d -series, we did not rely on a single random train/test split. Instead, model performance was evaluated using grouped cross-validation across chemically distinct subsets, so that each validation fold contains samples not used during fitting and belonging to a different transition-metal subgroup than the training data. This protocol provides a more stringent test of transferability than a random split and better reflects the models' intended use as a rationalization tool for periodic trends.

First, we used an unsupervised learning classification approach to find similarities and groupings among the TM₁₃ nanoclusters in descriptor space. Clustering was performed separately for pristine TM₁₃ and S/TM₁₃ systems using the k -means algorithm in `SCIKIT-LEARN`⁷¹. We set the number of clusters to $k = 10$, based on the idea that small groups of chemically similar elements would form compact regions in descriptor space across the 30 transition metals. We then compared the clustering results for pristine and S-adsorbed systems (Figures 1 and 4) to identify elements that remained close in both cases, suggesting they exhibit similar energetic, structural, vibrational, and electronic responses to S adsorption. This analysis identified Ti₁₃, Zr₁₃, and Hf₁₃ as a clear, physically meaningful group that shows strong S binding and moderate changes upon adsorption, so we chose them for detailed SO₂ adsorption calculations.

Second, to improve model interpretability, we trained a set of supervised regression models to link DFT-derived descriptors with adsorption energetics. Here, we focus on predicting atomic-sulfur adsorption energies on TM₁₃ and identify which pristine descriptors provide the most robust signal under cross-validation. To balance robustness and interpretability, we used three complementary regressors: (i) `ElasticNetCV`^{71,72}, which combines L1 and L2 regularization and selects among correlated descriptors; (ii) `RidgeCV`, which uses L2 regularization to spread weight across collinear features; and (iii) the `Explainable Boosting Machine (EBM)`^{73,74}, an interpretable boosted generalized additive model that captures nonlinear main effects while staying transparent. We evaluated models and quantified feature utility using group cross-validation to reduce optimistic bias and to test generalization across chemically distinct subsets rather than within a single correlated series.

To better understand how collinearity affects our models, we used a leave-one-feature-out (LOFO) protocol (Fig. 5). For each feature f , we recalculated cross-validated predictions after removing f and measured how much generalization changed. We report $\Delta R^2(f) = R^2_{full} - R^2_{-f}$, $\Delta MAE(f) = MAE_{-f} - MAE_{full}$, and $\Delta RMSE(f) = RMSE_{-f} - RMSE_{full}$. Here, “full” means the baseline model with all features, and “- f ” means the model with feature f removed. Positive Δ values indicate that the feature improves generalization, since removing it worsens performance. Negative Δ values suggest the feature is redundant or adds noise when descriptors are correlated. When LOFO results are similar across `ElasticNetCV`, `RidgeCV`, and `EBM`, it gives a reliable, model-independent view of which pristine descriptors matter most for S adsorption trends. In practice, the LOFO analysis distinguishes descriptors with consistent positive utility across model classes from those whose contribution is model-dependent or partially redundant due to collinearity. We therefore interpret the LOFO ranking as a robustness diagnostic for mechanistic relevance rather than as a strict one-to-one measure of causal importance. This also helps us interpret sulfur adsorption and poisoning as effects of both electronic stability/reactivity and lattice-dynamical changes.

The physic-informed model analysis is used as a rationalization and selection layer rather than a replacement for first-principles calculations: it identifies descriptors with robust explanatory value, supports the clustering-guided choice of representative systems (Ti₁₃, Zr₁₃, Hf₁₃), and provides an interpretable bridge between periodic DFT trends and the subsequent explicit SO₂ adsorption results.

Results and discussion

All findings presented here originated from PBE-D3 protocol optimizations, which resulted from protocol tests considering only PBE plain calculations and hybrid protocols with vdW D3 and/or SOC corrections for all 30 TM-NCs and S/TM₁₃ systems. For illustrative purposes, some of these protocol tests are presented in Supporting Information (Figures S2 and S3).

Pristine nanoclusters

After PBE+D3 optimizations, the NCs consistently preserved the expected geometry, highlighting the robustness of the ICO motif at the subnanometer scale (see Figure S1 in Supporting Information). Vibrational frequency analysis confirmed the dynamical stability of all structures, with no imaginary modes detected, as shown in Fig. 2(a). In addition, the vibrational frequencies decrease systematically from 3*d* to 5*d* elements. This trend reflects the interplay between atomic mass and *d*-orbital characteristics: lighter 3*d* elements with compact *d*-orbitals form shorter, stiffer bonds, while heavier 5*d* metals experience relativistic *s*-contraction and *d*-expansion, leading to weaker bonding.

The binding energies per atom, Fig. 2(b), follow an approximately parabolic dependence on atomic number within each *d*-series. Maximum stability is reached for mid-series elements, consistent with optimal occupation of bonding states and minimal occupation of antibonding states, as described by the cluster orbital model⁷⁵. Important to note a reasonably good agreement between $E_b^{TM_{13}}$ and the cohesive energy for TM bulk systems (E_{coh}^{TM}), although finite-size and surface effects produce systematic offsets (higher per-atom binding for extended solids in some species and enhanced NC stabilization in others). Additionally, deviations in stability for certain elements, e.g., Cr with its half-filled d^5 configuration, highlight the role of magnetism and local electronic effects.

The results for d_{av} and ECN (shown in Figures S2 and S3 in Supporting Information) corroborate these periodic trends. While d_{av} reflects the balance between bonding/antibonding occupation, the ECN values remain close to the ideal ICO reference (6.46), indicating that overall NC geometry is preserved even as bond lengths and binding strengths vary. Together, these descriptors establish a coherent picture: the calculated binding energies indicate that the pristine ICO NCs are energetically stable against dissociation into isolated atoms across the transition-metal series, with periodic modulations driven primarily by *d*-orbital filling.

Sulfur adsorption on nanoclusters

The S adsorption on TM_{13} NCs was systematically investigated by considering the three basic adsorption sites on ICO surface, i.e., top, bridge, and hollow. For each TM, full structural relaxations were performed, leading to the optimized S/ TM_{13} configurations, which are presented in the Supporting Information (Figure S1). For most cases, the hollow site is energetically preferred, reflecting the enhanced coordination of S and the more efficient charge redistribution enabled by multi-center bonding. Nevertheless, notable exceptions are identified: for W NCs, the bridge site is favored, whereas for Ir the top site is stabilized. These systems exhibit comparatively large distortion energies, indicating a pronounced geometrical resistance of the ICO cage to S binding. Such cases, together with selected hollow-site adsorptions accompanied by significant deformation, correlate with anomalous S–metal bond lengths, emphasizing that local geometry and electronic structure may override general coordination-based preferences.

Figure 2 illustrates that S adsorption induces systematic changes in both the energetic and vibrational responses of the ICO TM_{13} NCs, with trends that depend sensitively on the TM species. As shown in panel (a),

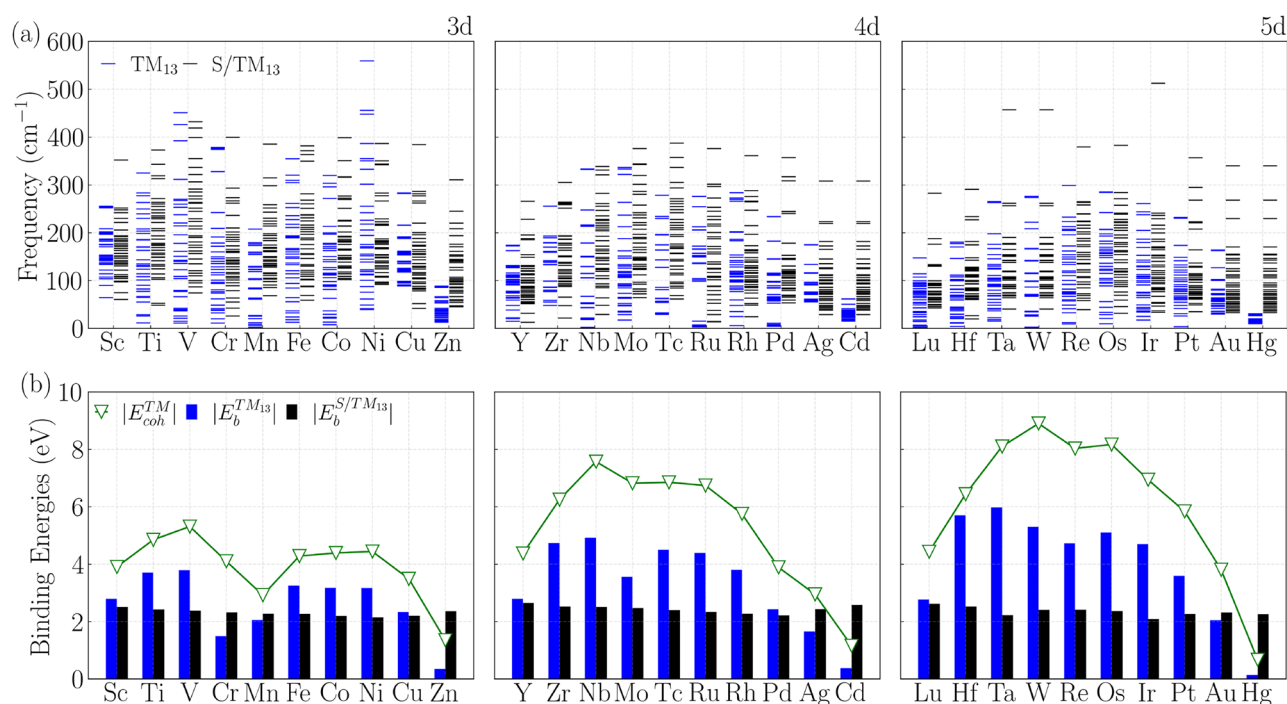


Fig. 2. Vibrational and energetic trends of TM_{13} and S/TM_{13} across the 3*d*–5*d* series. **(a)** Vibrational frequencies for pristine TM_{13} (blue) and S/TM_{13} (black) systems. **(b)** Per-atom $E_b^{TM_{13}}$ (blue) and $E_b^{S/TM_{13}}$ (black), compared with bulk cohesive energies (E_{coh}^{TM} , green).

the vibrational spectra of the S/TM₁₃ systems differ markedly from those of the pristine clusters, displaying additional high-frequency modes and a broader frequency distribution. Concurrently, panel (b) reveals a clear modification of the binding-energy behavior, i.e., the pronounced parabolic dependence of the per-atom E_b observed for pristine NCs is significantly flattened after S adsorption. The emergence of high-frequency vibrational features is consistent with the formation of localized and relatively stiff TM–S bonds, which introduce modes with larger effective force constants despite the increased S mass. Mode-projection analysis confirms that these contributions are predominantly localized at the adsorption site. The attenuated dependence of the E_b on atomic number suggests two complementary effects, the stabilizing contribution of the TM–S interaction, which is of comparable magnitude across several TMs, and a partial compensation of the intrinsic variations in intra-NC bonding that dominate the energetics of pristine TM₁₃ NCs.

A few elements, like Au, Cr, and Hg, show the opposite behavior, even though adsorption typically lowers the binding energy per atom due to structural distortions and weakening of metal–metal interactions. In these situations, the adsorbate–nanocluster complex is slightly stabilized in comparison to the pristine nanocluster due to the formation of strong metal–sulfur bonds that can compensate for the loss of metal–metal cohesion. This effect is particularly noticeable in systems with relatively weak intrinsic cohesion or modified electronic structure, such as late transition metals with filled d shells and significant relativistic contributions.

Deviations from these general trends are observed for elements with half-filled or fully filled d -shells, such as Cr, Mn, and Zn, as well as some heavier congeners (Ag, Cd, Au, and Hg). These cases highlight the role of magnetism and local electronic structure in determining adsorption energetics. In several systems, S adsorption modifies the local magnetic moments and orbital occupations, thereby altering the balance between bonding and antibonding states and ultimately affecting the binding strength. Overall, the combined energetic and vibrational descriptors indicate that S adsorption reorganizes local bonding more strongly than would be expected from bulk cohesive trends alone. Consequently, reliable predictions of S/TM₁₃ stability require explicit consideration of adsorbate–NC electronic interactions and adsorption-induced geometric relaxation.

The E_{ads} values reported in Fig. 3(a) are negative for all systems, confirming the thermodynamic favorability of S binding on TM₁₃ NCs. The adsorption strength spans from 1.39 eV for Hg to 11.70 eV for Mo, with most values lying in the 4–10 eV range. It is useful to compare these nanocluster trends with the much larger body of work on sulfur adsorption on extended transition-metal surfaces. Qualitatively, the same periodic logic is recovered: within a given d -series, early- and mid-row metals generally bind sulfur more strongly, whereas late and coinage-like metals bind more weakly. This behavior can be rationalized within the framework of the d -band model^{16,19,66,76,77}, in which the position of the metal d -band center relative to the Fermi level governs the strength of metal–adsorbate interactions through hybridization with Sp states⁷⁸. For homologous elements, the heavier $5d$ congeners can still exhibit strong adsorption because the larger radial extent of the d states and relativistic effects enhance overlap with the Sp orbitals.

The absolute adsorption energies reported here, however, are often larger than those typically reported for low-index surfaces. This is expected for subnanometer clusters: they expose low-coordinated atoms, possess narrower d bands, and can undergo local cooperative relaxations, all of which strengthen the metal–sulfur interaction relative to bulk-like terrace sites. Direct one-to-one comparison of absolute values should therefore be made with caution, because coordination environment, reference state, and site multiplicity differ between extended surfaces and 13-atom clusters. Although the d -band model provides a useful qualitative framework and is consistent with the periodic trends observed here, its quantitative applicability to nanoclusters is limited by their discrete electronic structure and pronounced quantum-size and coordination effects, which weaken simple correlations between adsorption energies and single descriptors such as the d -band center. While strong S binding may facilitate adsorbate activation, it also hampers desorption, in line with the Sabatier principle^{79,80}, and provides a rational basis for S poisoning on late-TM sites.

The E_{ads} decomposition into ΔE_{int} and $\Delta E_{\text{dis}}^{\text{TM}_{13}}$ contributions reveals that the interaction term dominates the total binding strength for all TMs, confirming that chemisorption is the primary driving force for S adsorption. In contrast, the distortion contribution is generally moderate, indicating that the ICO motif is largely preserved upon adsorption. However, several systems (Cr, Mo, Tc, W, Re, Os, and Ir) exhibit relatively large distortion-energy magnitudes (exceeding 2 eV), indicating substantial local nanocluster rearrangements. The sign of $\Delta E_{\text{dis}}^{\text{TM}_{13}}$ should be interpreted according to the definition adopted in Sec. Methodology. Positive values correspond to an energetic penalty associated with deforming the pristine ICO nanocluster into the adsorption-induced geometry. By contrast, negative values indicate that the nanocluster geometry obtained after removing the adsorbate from the relaxed adsorbed structure is lower in energy than the initially optimized pristine reference. In this sense, adsorption can drive the system toward a more favorable nanocluster geometry, even though the pristine structure was initially a local minimum. Consequently, systems with large-magnitude distortion energies combine strong chemical interactions with appreciable geometric adaptation, suggesting that S-rich environments may locally reshape pristine active sites even when adsorption remains exothermic.

Figure 3(b) presents the evolution of the ECN and d_{av} across the three TM series before and after S adsorption. These structural descriptors provide complementary insight into adsorption-induced distortions. The d_{av} and ECN variations, together with their respective changes (Δd_{av} and ΔECN ; Figures S2 and S3), indicate that most NCs retain near-ICO symmetry, with distortions localized around the adsorption site. Early $3d$ elements display larger perturbations, reflecting their lower structural rigidity and higher flexibility. In contrast, mid- and late-series NCs, particularly in the $4d$ and $5d$ rows, are more resilient, undergoing only minor geometric adjustments. Among the two descriptors, d_{av} correlates more consistently with the distortion energy, whereas ECN exhibits heightened sensitivity to adsorption-induced rearrangements. Large ECN variations are associated with substantial structural transformations, sometimes leading to motifs more stable than the pristine ICO geometry (e.g., Cr, Mo, W, and Re), as evidenced by negative $\Delta E_{\text{dis}}^{\text{TM}_{13}}$. Conversely, positive distortion energies indicate cases in which S adsorption perturbs the ICO framework without conferring additional stabilization.

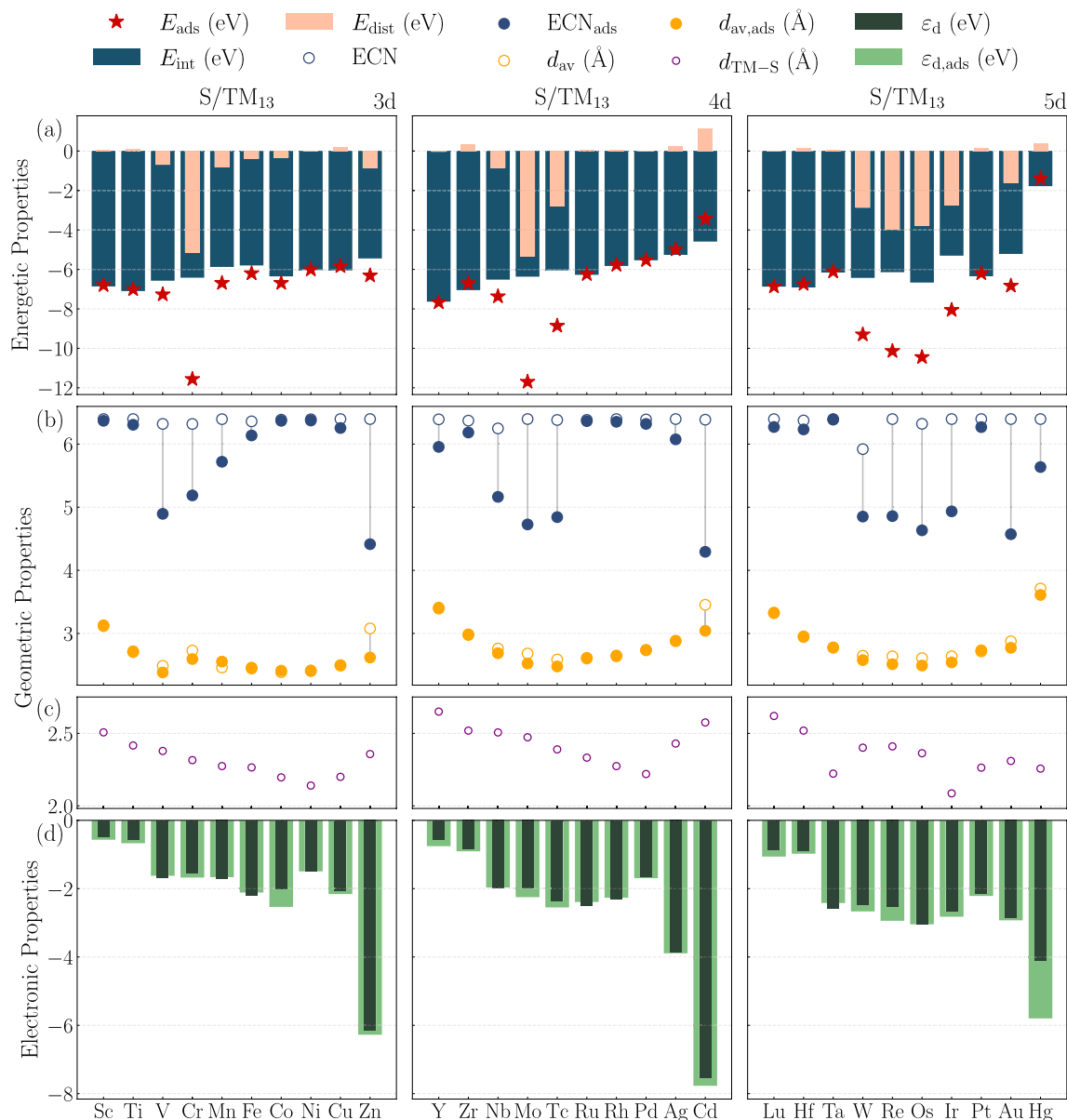


Fig. 3. Energetic, structural, and electronic descriptors for S adsorption on TM_{13} NCs across the $3d$, $4d$, and $5d$ series. (a) Adsorption energy decomposition into interaction (ΔE_{int}) and distortion ($\Delta E_{\text{dis}}^{\text{TM}_{13}}$) contributions, together with total adsorption energies (E_{ads}). (b) Effective coordination number (ECN) and average TM-TM bond length (d_{av}) before and after S adsorption. (c) TM-S bond distance ($d_{\text{TM-S}}$) for each TM. (d) Center of gravity of the occupied d -states (ϵ_d) for pristine and S-adsorbed NCs.

These observations demonstrate that correlating distortion energies with structural descriptors is nontrivial and requires consideration of electronic effects, including adsorption-induced changes in the total magnetic moment (Table S5).

The TM-S bond distances shown in Fig. 3(c) exhibit an approximately parabolic trend along each TM row, with the shortest distances occurring for mid-row elements, where the d band is neither too empty nor too filled, thereby maximizing S-TM hybridization. Across a given group, the bond length generally increases from $3d$ to $5d$ elements due to atomic-size effects, although relativistic contraction in the $5d$ series leads to shorter-than-expected bonds for Ta and Ir. While a direct correlation between shorter TM-S distances and stronger adsorption might be anticipated, such a relationship is not universally observed. Structural distortions and preferential site changes, as in the case of S/Ir₁₃, generate distinct local chemical environments that modulate bond lengths independently of adsorption energy. Indeed, an inverse correlation between $d_{\text{TM-S}}$ and $|\Delta E_{\text{int}}|$ is observed for most systems, underscoring the structural origin of interaction strength. Excessively short TM-S bonds impose strain on the ICO cage, linking geometric frustration with the energetic fingerprints of S adsorption. In the present ICO-based framework, these negative values should therefore be interpreted as adsorption-assisted

reconstruction relative to the original pristine ICO reference, rather than as a contradiction of the distortion-energy definition itself.

Figure 3(d) compares the ε_d values for pristine and S-adsorbed NCs. Across the $3d$ to $5d$ series, ε_d shifts progressively to lower energies, reflecting the increased stabilization of the metal d states. Consistent with the qualitative spirit of the d -band model⁶⁶, metals with ε_d closer to the Fermi level tend to exhibit stronger adsorbate–metal hybridization, which is broadly consistent with the enhanced reactivity of early transition metals and the weaker S interactions observed for later-series elements. However, variations in the d -band center alone do not constitute a validation of the d -band model for nanoclusters. While the position of ε_d can still provide useful qualitative insight into metal–adsorbate interactions, the discrete electronic structure, low coordination, and finite-size effects of small nanoclusters require a more cautious interpretation.

Bader charge analysis supports this picture by showing a net electron transfer from the NC to S in all cases, driven by electronegativity differences. The largest charge-transfer occurs in the $3d$ NCs, where the d states are less stabilized and more readily available for donation. Differential charge-density plots further confirm localized electron accumulation on S and depletion on neighboring metal atoms, highlighting the mixed covalent–metallic character of the S–NC bond. At the same time, the correlation between adsorption energy and the d -band center remains relatively weak for the nanoclusters investigated here. Although the d -band model has been highly successful in rationalizing adsorption trends on extended transition-metal surfaces, its predictive power is diminished in small nanoclusters due to quantum-size effects, discrete electronic states, and low atomic coordination. Therefore, in the present work, ε_d should be interpreted primarily as a qualitative descriptor rather than as a strictly predictive variable for adsorption strength.

Physics-informed interpretation of sulfur adsorption and poisoning trends

Here, “physics-informed” refers to the use of physically motivated, DFT-derived descriptors and to their interpretation in light of established physical and chemical principles. Within this framework, we combine unsupervised clustering and LOFO analysis to identify periodic trends in atomic S adsorption across the transition-metal series and to motivate the choice of representative systems (Figures 4–6). In descriptor space, Ti_{13} , Zr_{13} , and Hf_{13} (atomic numbers 22, 40, and 72) form a clear and physically meaningful group. Consistent with the DFT energetics and structural analyses discussed above, these nanoclusters exhibit strong S binding while remaining comparatively resilient within the ICO motif. Notably, Ti, Zr, and Hf constitute an isoelectronic triad across the $3d$, $4d$, and $5d$ series, providing a controlled set of systems for comparing periodic trends and relativistic effects.

We emphasize that PCA is used here only as a linear, low-dimensional visualization of the standardized descriptor space. In contrast, the k -means classification is performed in the full descriptor space after standardization. Therefore, the two-dimensional PCA maps in Fig. 4(a) should be interpreted as a projection that helps visualize chemical organization, not as the space in which all descriptor information is retained. In this projected space, the most compact and persistent group is formed by Ti_{13} , Zr_{13} , and Hf_{13} , which remain close in both the pristine and adsorbed representations. By contrast, isolated points or sparsely populated regions correspond to chemically distinctive systems whose descriptor combinations differ markedly from the majority trend, typically due to anomalous magnetic/electronic occupations, filled or half-filled d -shell effects, unusual adsorption-site preferences, or large adsorption-induced structural distortions. We list the corresponding element-wise PCA coordinates and cluster assignments in Table S7 of the Supporting Information.

Before interpreting feature utility, we quantify the absolute predictive accuracy of the full supervised models under the same grouped cross-validation protocol used throughout this work. ElasticNetCV achieves $R^2 = -0.431$, MAE = 1.936 eV, and RMSE = 2.517 eV; RidgeCV achieves $R^2 = -0.109$, MAE = 1.728 eV, and RMSE = 2.217 eV; and EBM achieves $R^2 = -0.059$, MAE = 1.762 eV, and RMSE = 2.166 eV. These grouped-cross-validation results indicate that the models retain only modest extrapolative accuracy across chemically distinct transition-metal subgroups, with RidgeCV and EBM outperforming ElasticNetCV. We therefore primarily use supervised analysis as an interpretable trend-rationalization and descriptor-screening tool rather than as a high-accuracy predictive model.

Figure 4(a) shows that the pristine and adsorbed nanoclusters occupy related but distinct regions in descriptor space, demonstrating that sulfur adsorption systematically reorganizes the energetic, structural, vibrational, and electronic fingerprints of the TM_{13} systems. The displacement of the points upon adsorption reflects the combined effect of metal–sulfur bonding, charge redistribution, and local structural adaptation. Importantly, some systems remain close to one another in both representations. The most notable example is the isoelectronic Ti/Zr/Hf triad, which stays grouped in both the PCA projections and the k -means assignments [Fig. 4(b)]. This persistence indicates that these nanoclusters respond to sulfur adsorption in a chemically analogous way: they bind sulfur strongly, yet avoid the larger adsorption-induced rearrangements observed for more distortion-prone cases.

Considering the three regression models together with the clustering results (Fig. 4) and the LOFO ranking (Fig. 5), Ti_{13} , Zr_{13} , and Hf_{13} emerge in an intermediate adsorption regime: S binds strongly enough to promote activation of S-containing molecules, yet not so strongly that it is systematically associated with large adsorption-induced destabilization of the scaffold. This regime aligns with the Sabatier principle, which predicts optimal catalytic behavior at intermediate binding strengths (neither too weak for activation nor too strong, which would lead to poisoning and slow turnover).

Although several transition-metal nanoclusters exhibit adsorption energies within a similar intermediate range, the Ti_{13} , Zr_{13} , and Hf_{13} clusters stand out due to their combination of moderate sulfur binding and remarkable structural stability upon adsorption. This behavior contrasts with the more pronounced structural distortions that many other transition-metal clusters undergo upon interaction with sulfur species. In catalysis, where resistance to poisoning-induced structural degradation is desirable, such stability is especially appealing.

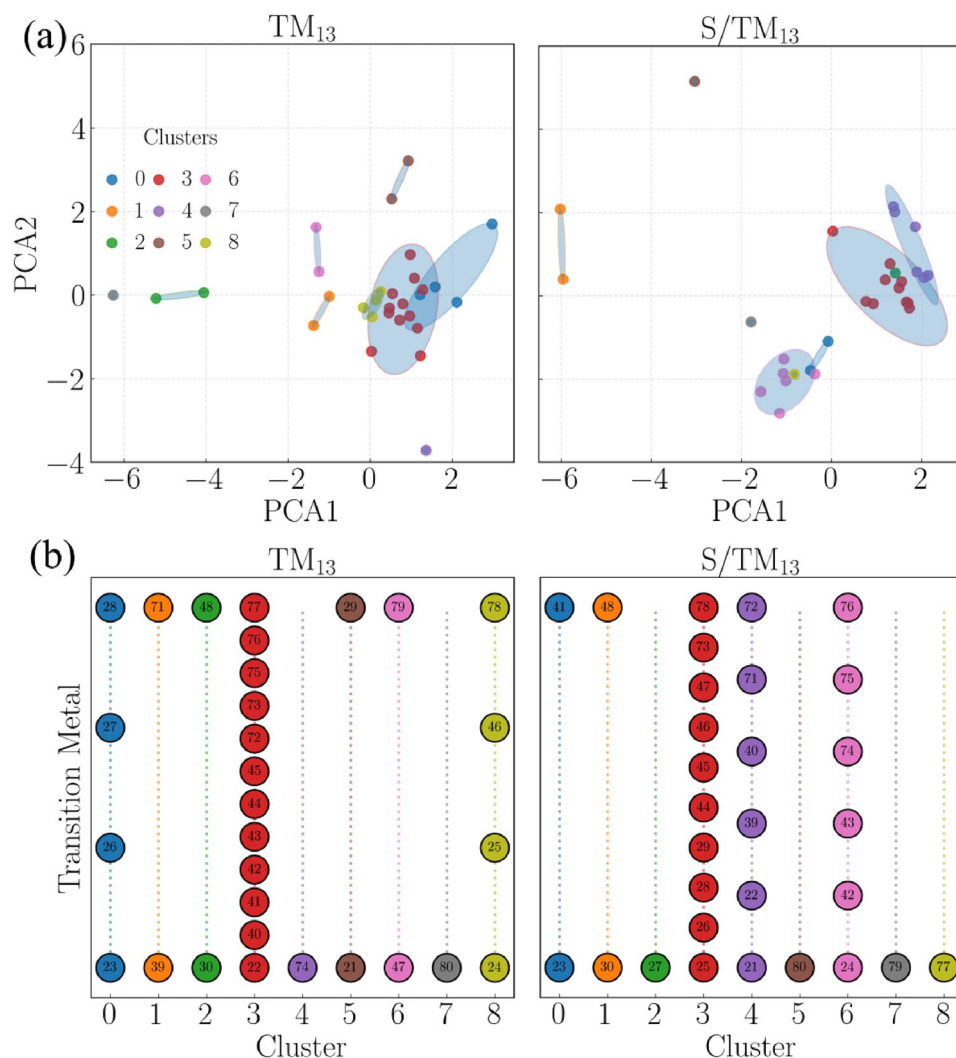


Fig. 4. (a) Two-dimensional PCA projections of standardized DFT-derived descriptors for pristine TM_{13} (left) and S/TM_{13} (right) NCs. Colors denote k -means cluster assignments, while ellipses indicate the dispersion of each cluster in descriptor space. (b) Cluster membership of TM-NCs obtained from k -means classification for TM_{13} (left) and S/TM_{13} (right) systems. Each point represents a TM element, identified by its atomic number, and elements that remain grouped in both classifications exhibit similar responses to S adsorption. The PCA maps are shown only for visualization of the descriptor organization; the k -means classification was performed in the full standardized descriptor space.

Bringing together the results from both unsupervised grouping and supervised trend-finding gives us a solid, data-driven reason to pick Ti_{13} , Zr_{13} , and Hf_{13} as reliable and chemically tough platforms for detailed SO_2 adsorption studies and related S-poisoning situations. When prioritizing accurate predictions over the identification of general trends, the next step would be to add adsorption-response descriptors (like ΔE_{int} , ΔE_{dis} , Δd_{av} , ΔECN , $d_{\text{TM-S}}$, or charge transfer). These will directly capture the structural and electronic changes that drive unusually strong or weak binding cases.

Figure 6 provides a sample-resolved comparison between the DFT reference adsorption energies and the predictions of the three supervised models across the $3d$, $4d$, and $5d$ transition-metal series. While the grouped cross-validation metrics above quantify the overall predictive accuracy, Fig. 6 shows how the models perform metal-by-metal. All three regressors reproduce parts of the broad periodic variation of $|E_{\text{ads}}|$, but the deviations remain substantial under chemically distinct group holdout. In particular, extrapolation to the $3d$ series is the most challenging, whereas transfer across the $4d$ and $5d$ series is comparatively more stable, especially for RidgeCV and EBM. In this sense, Fig. 6 complements the LOFO analysis by showing where the models follow the DFT trends and where their predictive errors concentrate.

Our combined unsupervised and supervised ML approaches aren't meant to replace first-principles calculations. Instead, they give us a practical, physics-informed way to understand periodic trends, see which pristine descriptors still matter during cross-validation, and pick the right materials. In this light, Ti_{13} , Zr_{13} , and Hf_{13} consistently stand out as a group that binds just right: S sticks well enough to activate S-containing

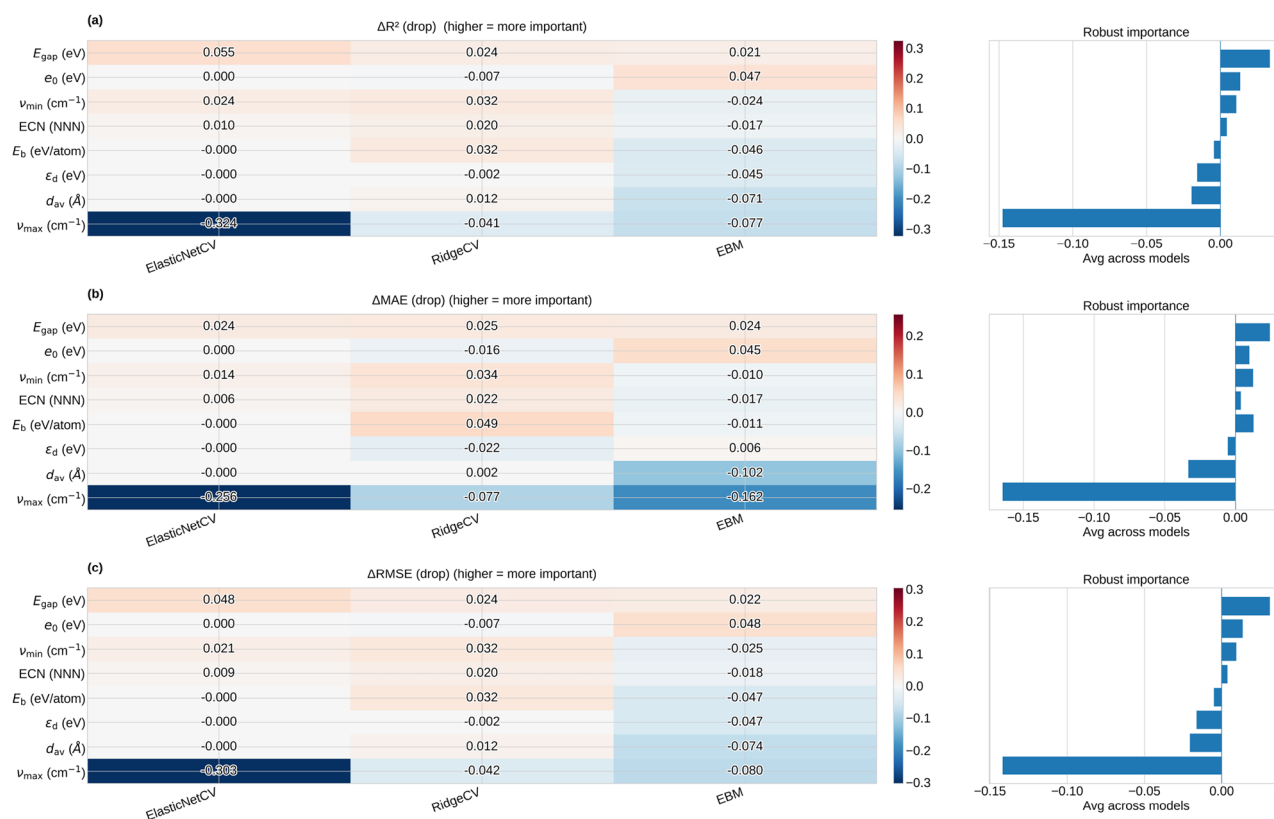


Fig. 5. Leave-one-feature-out (LOFO) analysis measures how removing each pristine nanocluster descriptor affects model generalization when predicting atomic sulfur adsorption energies on TM_{13} icosahedra. The panels show performance drops as (a) ΔR^2 , (b) ΔMAE , and (c) ΔRMSE , all evaluated with group cross-validation using three regression models: ElasticNetCV, RidgeCV, and Explainable Boosting Machine (EBM). The heatmaps on the left display the effects for each model, while the bar plots on the right show the average impact across models, ranking feature importance (the y-axis is the same as in the left panels).

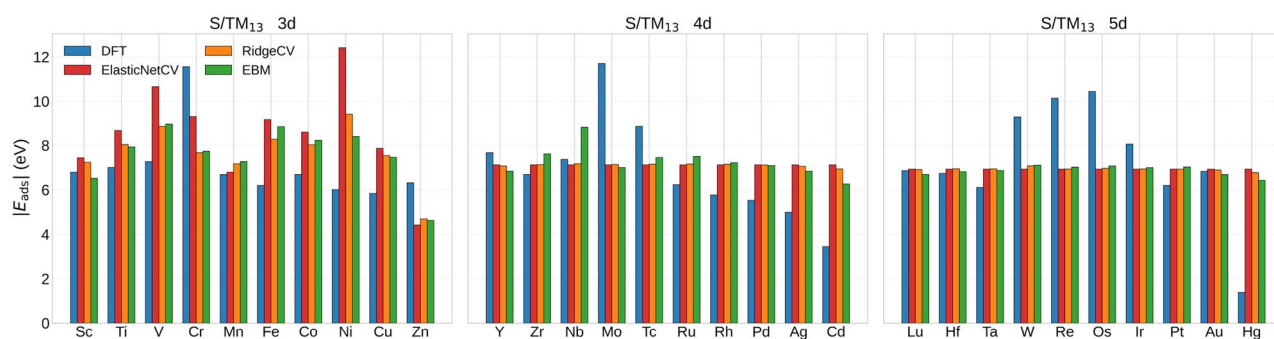


Fig. 6. Comparison between DFT adsorption energies with machine learning predictions for atomic sulfur adsorption on TM_{13} icosahedral nanoclusters across the $3d$, $4d$, and $5d$ transition metal series. For each metal, the bars show both the DFT reference value of $|E_{\text{ads}}|$ for S/TM_{13} and the predictions from three different machine learning models (ElasticNetCV, RidgeCV, and Explainable Boosting Machine, EBM) trained using pristine TM_{13} descriptors as inputs and atomic-sulfur adsorption energies as targets.

molecules, but not so much that it causes significant destabilization. That sort of balance is precisely in line with the Sabatier principle. It provides a solid, data-driven rationale for choosing these nanoclusters as robust, representative platforms for detailed SO_2 adsorption studies and other S-poisoning scenarios.

Case study validation: SO_2 adsorption

We performed explicit first-principles calculations of SO_2 adsorption on the corresponding selected 13-atom ICO NCs: $\text{SO}_2/\text{Ti}_{13}$, $\text{SO}_2/\text{Zr}_{13}$, and $\text{SO}_2/\text{Hf}_{13}$, using our DFT-PBE+D3 protocol, to validate the trends inferred from the ML analysis. These NCs were identified by the unsupervised learning stage as representative and chemically resilient candidates within the TM series. For each system, we have run AIMD simulations, as presented in Supporting Information (Figure S4), exploring multiple adsorption geometries (downward S and O orientations and a tilted orientation) at the hollow, bridge, and top sites. In Fig. 7, the lowest energy dissociative adsorption geometries in panel (a) with the respective Bader charge density isosurfaces, and the corresponding PDOS in panel (b), while the lowest energy molecular adsorption geometries are presented in panel (c) with the respective Bader charge density isosurfaces, and the corresponding PDOS in panel (d). Additionally, in Table 1 are presented the E_b , E_{ads} , ΔE_{int} , ΔE_{dis} , Δd_{av} , ΔECN , $d_{\text{TM-S}}$, $d_{\text{TM-O}}$, $d_{\text{S-O}}$, ε_d , and ν values for $\text{SO}_2/\text{Ti}_{13}$, $\text{SO}_2/\text{Zr}_{13}$, and $\text{SO}_2/\text{Hf}_{13}$.

For all three NCs (Ti_{13} , Zr_{13} , and Hf_{13}), the putative global minimum corresponds to dissociative adsorption of the SO_2 molecule. In these configurations, the S–O bonds are strongly elongated and effectively cleaved, leading to the formation of robust metal–S and metal–O bonds, as illustrated in Fig. 7(a). In contrast,

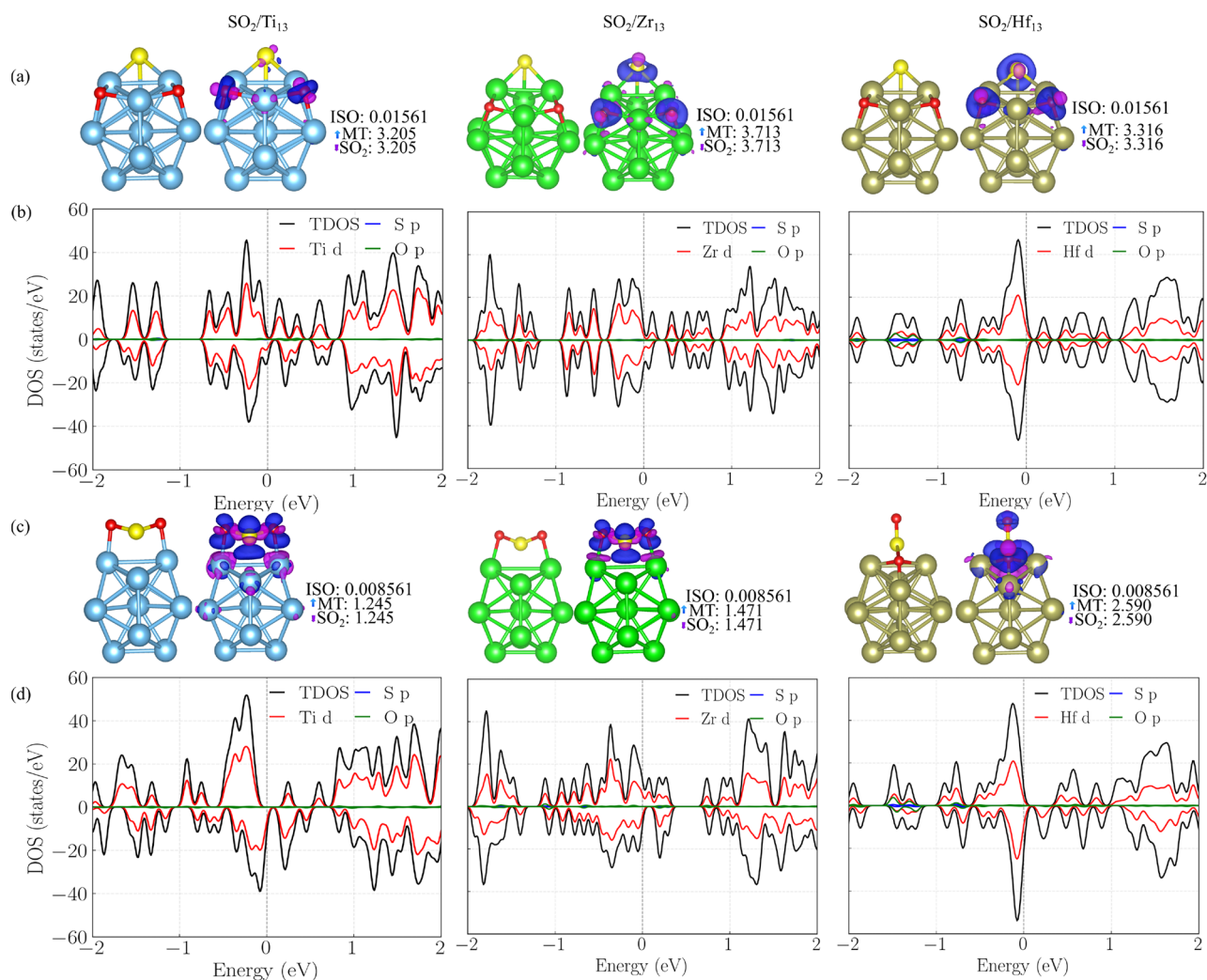


Fig. 7. The lowest-energy adsorption configurations and electronic structure of SO_2 on 13-atom ICO NCs. (a) Dissociative adsorption geometries of SO_2 on Ti_{13} , Zr_{13} , and Hf_{13} , together with the corresponding Bader charge density difference isosurfaces. (b) Projected density of states (PDOS) for the dissociated configurations, showing the total DOS and the contributions from TM d , S p , and O p states. (c) Most stable molecular (non-dissociated) adsorption geometries and associated charge density distributions. (d) PDOS for the intact molecular adsorption configurations. The Fermi level is set to zero energy in all panels.

	SO ₂ /Ti ₁₃	SO ₂ /Zr ₁₃	SO ₂ /Hf ₁₃
E_b (eV/atom)	-4.584 (-4.065)	-5.374 (-4.863)	-5.320 (-4.954)
E_{ads} (eV)	-13.063 (-4.745)	-12.356 (-4.179)	-12.681 (-6.826)
ΔE_{int} (eV)	-25.416 (-6.472)	-25.41 (-6.472)	-25.100 (-13.182)
$\Delta E_{dis}^{TM_{13}}$ (eV/atom)	0.016 (0.011)	0.360 (0.043)	0.342 (.288)
$\Delta E_{dis}^{SO_2}$ (eV/atom)	12.016 (1.923)	12.591 (1.733)	5.879 (12.6225)
Δd_{av} (%)	1.216 (0.517)	1.535 (0.497)	1.444 (1.158)
ΔECN (%)	-1.166 (-3.595)	-1.666 (-3.017)	-1.336 (-1.633)
d_{TM-S} (Å)	2.428 (2.487)	2.580 (2.670)	2.560 (2.437)
d_{TM-O} (Å)	1.956 (1.913)	2.098 (2.073)	2.069 (2.068)
d_{S-O} (Å)	4.198 (1.660)	4.489 (1.644)	4.410 (1.537)
ε_d (eV)	-0.870 (-1.125)	-1.317 (-1.345)	-1.178 (-1.181)
	664.102 (476.699)	660.040 (446.927)	942.482 (471.810)
ν (cm ⁻¹)	638.769 (438.667)	637.756 (414.553)	540.731 (433.483)
	464.116 (420.727)	482.428 (411.623)	404.675 (420.523)

Table 1. The structural, energetic, electronic, and vibrational properties of the most stable SO₂-adsorbed configurations on Ti₁₃, Zr₁₃, and Hf₁₃ ICO NCs. Values reported outside parentheses correspond to the lowest-energy dissociative adsorption configurations, whereas values in parentheses refer to the most stable molecularly adsorbed (intact) SO₂ geometries. The listed quantities include the binding energy per atom (E_b), adsorption energy (E_{ads}), interaction and distortion energy contributions (ΔE_{int} , ΔE_{dis}), relative changes in average bond length (Δd_{av}) and effective coordination number (ΔECN), characteristic metal-S and metal-O bond lengths, the S-O bond distance, the d -band center (ε_d), and representative molecule vibrational frequencies (ν).

molecularly adsorbed configurations, in which the SO₂ molecule remains intact (Fig. 7(c)), are systematically higher in energy and therefore metastable, as quantitatively reflected by the adsorption energies listed in Table 1.

The energetic decomposition analysis confirms that dissociative adsorption is driven by very large and negative interaction energies (ΔE_{int}), which overwhelmingly dominate over the distortion penalties associated with both the NC ($\Delta E_{dis}^{TM_{13}}$) and the adsorbate ($\Delta E_{dis}^{SO_2}$). This behavior is particularly pronounced for Zr₁₃ and Hf₁₃, which exhibit the most negative adsorption energies and interaction terms, reflecting their higher chemical affinity toward S- and O-containing species. Despite strong adsorption, the relatively modest Δd_{av} and ΔECN values further indicate that the underlying ICO framework of the NCs remains largely intact, consistent with their classification as structurally resilient adsorption platforms.

Additional insight into the bonding mechanism is obtained from the PDOS shown in Fig. 7(b). In the dissociated configurations, the S p states hybridize strongly with the TM d manifold near the Fermi level, providing a clear electronic signature of covalent metal-S bonding. This hybridization is accompanied by a redistribution of the metal d states, consistent with the downward shift of the d -band center reported in Table 1. Such electronic reorganization is well known to stabilize strongly bound adsorbates and directly supports the trends predicted by the ML descriptors.

The structural fingerprints of dissociative adsorption are further corroborated by the pronounced elongation of the S-O bond distances, which increase from typical gas-phase values to more than 4 Å in the most stable configurations. This bond cleavage is also reflected in the vibrational analysis, where the spectra become dominated by lower-frequency metal-S and metal-O modes, consistent with the formation of strong chemisorbed fragments. Overall, the explicit SO₂ adsorption results place Ti₁₃, Zr₁₃, and Hf₁₃ in an intermediate adsorption regime where S-containing species bind strongly enough to undergo activation and dissociation, yet without inducing catastrophic structural NC degradation. This behavior is fully consistent with the Sabatier principle^{79,80}, and directly validates the ML-guided selection of these NCs as chemically resilient and catalytically relevant platforms for S-containing molecules.

Conclusions

We performed a systematic investigation of the stability, reactivity, and sulfur tolerance of 13-atom icosahedral transition-metal nanoclusters by combining dispersion-corrected first-principles calculations with interpretable machine-learning analyses. By screening the $3d$, $4d$, and $5d$ series within a fixed ICO motif, we isolated transferable structure-property relationships that disentangle intrinsic nanocluster stability from adsorption-driven changes. The DFT results show that the pristine-cluster binding energy follows an approximately parabolic dependence on atomic number within each d series, with maximal stability near mid-series elements, consistent with optimal filling of bonding states in the cluster-orbital picture. Upon atomic S adsorption, adsorption energetics are dominated by the interaction contribution. At the same time, distortion penalties are typically moderate, and the vibrational fingerprints reveal localized stiff TM-S modes and adsorption-induced broadening of the spectra. In the associated descriptor space, unsupervised clustering organizes the

transition metals into chemically meaningful groups, and model-agnostic LOFO analyses across complementary regression models identify which descriptors contribute most consistently under stringent cross-validation, thereby supporting mechanistic interpretation even when extrapolative predictive accuracy remains limited. Guided by this combined unsupervised-supervised framework, we identified the isoelectronic triad Ti_{13} , Zr_{13} , and Hf_{13} as representative, chemically resilient platforms that balance strong S binding with limited adsorption-induced structural perturbation. Explicit DFT calculations of SO_2 adsorption on these nanoclusters validate this selection: the preferred adsorption is strong and predominantly dissociative, driven by large metal-adsorbate interaction energies that outweigh the structural distortion costs. At the same time, the ICO framework remains largely preserved. Electronic-structure analysis shows pronounced hybridization between S-*p* and metal-*d* states near the Fermi level together with a downward shift of the *d*-band center, providing a consistent electronic fingerprint for activation and poisoning propensity; concomitantly, the structural and vibrational signatures corroborate bond weakening/cleavage in SO_2 and the emergence of metal-S/metal-O modes in the dissociated states.

Our results place Ti_{13} , Zr_{13} , and Hf_{13} in an intermediate adsorption regime where S-containing molecules can be activated without catastrophic degradation of the catalytic scaffold, aligning with the Sabatier principle and highlighting these mid-series nanoclusters as promising candidates for sulfur-tolerant catalysis. More broadly, this study demonstrates that integrating first-principles energetics, electronic descriptors, and lattice-dynamical fingerprints with interpretable ML (including robustness checks across model classes) enables a rational, transferable mapping from periodic-table chemistry to adsorption and poisoning mechanisms in sub-nanometer catalysts.

Data availability

In addition to the Supporting Information, all data sets and workflows used in this study are openly available at <https://github.com/KIT-Workflows/Physics-Informed-Nanoclusters>. The repository includes representative DFT input/output files, processed descriptor tables, scripts for preprocessing and feature standardization, PCA and *k*-means analyses, supervised regression and group-cross-validation workflows, LOFO-analysis scripts, trained models, and figure-generation notebooks/scripts used in the manuscript.

Received: 23 January 2026; Accepted: 24 April 2026

Published online: 04 May 2026

References

- Alivisatos, A. P. Semiconductor clusters, nanocrystals, and quantum dots. *Science* **271**, 933–937 (1996).
- Schmid, G. *Nanoparticles: From Theory to Application* 2 edn. (Wiley-VCH, 2011).
- Maiti, S. et al. A comprehensive review of atomically precise metal nanoclusters with emergent photophysical properties towards diverse applications. *Chem. Soc. Rev.* **54**, 1585–1844. <https://doi.org/10.1039/D4CS00962B> (2025).
- Jena, P. & Sun, Q. Super atomic clusters: design rules and potential for building blocks of materials. *Chem. Rev.* **118**, 5755–5870. <https://doi.org/10.1021/acs.chemrev.7b00524> (2008).
- Castleman, A. W. & Jena, P. Clusters: A bridge across the disciplines of environment, materials science, and biology. *Proc. Natl. Acad. Sci.* **103**, 10554–10559. <https://doi.org/10.1073/pnas.0601780103> (2006).
- Bell, A. T. The impact of nanoscience on heterogeneous catalysis. *Science* **299**, 1688–1691. <https://doi.org/10.1126/science.1083671> (2003).
- Aslam, U., Rao, V. G., Chavez, S. & Linic, S. Catalytic conversion of solar to chemical energy on plasmonic metal nanostructures. *Nat. Catal.* **1**, 656–665. <https://doi.org/10.1038/s41929-018-0138-x> (2018).
- Li, J., Li, X., Zhai, H.-J. & Wang, L.-S. Au₂₀: A tetrahedral cluster. *Science* **299**, 864–867. <https://doi.org/10.1126/science.1079879> (2003).
- Cao, G. & Wang, Y. *Nanostructures and Nanomaterials: Synthesis, Properties, and Applications* 2 edn. (World Scientific, 2011).
- Yoon, B. et al. Charging effects on bonding and catalyzed oxidation of CO on Au₈ clusters on MgO. *Science* **307**, 403–407. <https://doi.org/10.1126/science.1104168> (2005).
- Sun, G. & Sautet, P. Active site fluxional restructuring as a new paradigm in triggering reaction activity for nanocluster catalysis. *Acc. Chem. Res.* **54**, 3841–3849. <https://doi.org/10.1021/acs.accounts.1c00413> (2021).
- Rêgo, C. R. C., Tereshchuk, P., Oliveira, L. N. & Da Silva, J. L. F. Graphene-supported small transition-metal clusters: A density functional theory investigation within van der Waals corrections. *Phys. Rev. B* <https://doi.org/10.1103/physrevb.95.235422> (2017).
- Campbell, C. T. The energetics of supported metal nanoparticles: Relationships to sintering rates and catalytic activity. *Acc. Chem. Res.* **46**, 1712–1719. <https://doi.org/10.1021/ar3003514> (2013).
- Bartholomew, C. H. Mechanisms of catalyst deactivation. *Appl. Catal. A Gen.* **212**, 17–60. [https://doi.org/10.1016/S0926-860X\(00\)00843-7](https://doi.org/10.1016/S0926-860X(00)00843-7) (2001).
- Mesilov, V. et al. Insights into sulfur poisoning and regeneration of Cu-SSZ-13 catalysts: In situ Cu and S K-edge XAS studies. *Catal. Sci. Technol.* **11**, 5619–5632. <https://doi.org/10.1039/D1CY00975C> (2021).
- Liu, D.-J. et al. Sulfur adsorption on coinage metal(100) surfaces: Propensity for metal-sulfur complex formation relative to (111) surfaces. *Phys. Chem. Chem. Phys.* **21**, 26483–26491. <https://doi.org/10.1039/C9CP03449H> (2019).
- Fielicke, A. et al. Structure determination of isolated metal clusters via far-infrared spectroscopy. *Phys. Rev. Lett.* **93**, 023401. <https://doi.org/10.1103/PhysRevLett.93.023401> (2004).
- Heiz, U. & Landman, U. *Nanocatalysis* Vol. 1 (2007).
- Nørskov, J. K., Abild-Pedersen, F., Studt, F. & Bligaard, T. Density functional theory in surface chemistry and catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 937–943. <https://doi.org/10.1073/pnas.1006652108> (2011).
- Häkkinen, H. The gold-sulfur interface at the nanoscale. *Nat. Chem.* **4**, 443–455. <https://doi.org/10.1038/nchem.1352> (2012).
- Ferrando, R., Jellinek, J. & Johnston, R. L. Nanoalloys: From theory to applications of alloy clusters and nanoparticles. *Chem. Rev.* **108**, 845–910. <https://doi.org/10.1021/cr040090g> (2008).
- Peraça, C. S. T., Nagurniak, G. R., Orenha, R. P., Parreira, R. L. T. & Piotrowski, M. J. A theoretical indicator of transition-metal nanoclusters applied in the carbon nanotube nucleation process: A DFT study. *Dalton Trans.* **49**, 492. <https://doi.org/10.1039/c9dt04272e> (2020).
- Piotrowski, M. J., Palheta, J. M. T. & Fournier, R. Cage doping of Ti, Zr, and Hf-based 13-atom nanoclusters: Two sides of the same coin. *Phys. Chem. Chem. Phys.* **26**, 13172–13181. <https://doi.org/10.1039/d4cp00518j> (2024).

24. Say, Z. et al. Unraveling molecular fingerprints of catalytic sulfur poisoning at the nanometer scale with near-field infrared spectroscopy. *J. Am. Chem. Soc.* **144**, 8848–8860. <https://doi.org/10.1021/jacs.2c03088> (2022).
25. Yin, P. et al. Sulfur stabilizing metal nanoclusters on carbon at high temperatures. *Nat. Commun.* **12**, 3135. <https://doi.org/10.1038/s41467-021-23426-z> (2021).
26. Deringer, V. L., Caro, M. A. & Csányi, G. Machine learning interatomic potentials as emerging tools for materials science. *Adv. Mater.* **31**, 1902765. <https://doi.org/10.1002/adma.201902765> (2019).
27. Zuo, Y. et al. Performance and cost assessment of machine learning interatomic potentials. *J. Phys. Chem. A* **124**, 731–745. <https://doi.org/10.1021/acs.jpca.9b08723> (2020).
28. Chmiela, S., Sauceda, H. E., Müller, K.-R. & Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nature Communications* **9**, 3887. <https://doi.org/10.1038/s41467-018-06169-2> (2018).
29. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301. <https://doi.org/10.1103/PhysRevLett.120.145301> (2018).
30. Baletto, F. & Ferrando, R. Structural properties of nanoclusters: Energetic, thermodynamic, and kinetic effects. *Rev. Mod. Phys.* **77**, 371–423. <https://doi.org/10.1103/RevModPhys.77.371> (2005).
31. Johnston, R. L. The development of metallic behaviour in clusters. *Proc. R. Soc. A: Math. Phys. Eng. Sci.* **356**, 211–230. <https://doi.org/10.1098/rsta.1998.0158> (1998).
32. Imaoka, T. et al. Magic number pt_{13} and misshapen pt_{12} clusters: Which one is the better catalyst?. *J. Am. Chem. Soc.* **135**, 13089–13095. <https://doi.org/10.1021/ja405922m> (2013).
33. Piotrowski, M. J., Piquini, P. & Silva, J. L. F. D. Density functional theory investigation of $3d$, $4d$, and $5d$ 13-atom metal clusters. *Phys. Rev. B* **81**, 155446. <https://doi.org/10.1103/PhysRevB.81.155446> (2010).
34. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136**, B864–B871. <https://doi.org/10.1103/PhysRev.136.B864> (1964).
35. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133–A1138. <https://doi.org/10.1103/PhysRev.140.A1133> (1965).
36. Kresse, G. & Hafner, J. Ab initio molecular dynamics for open-shell transition metals. *Phys. Rev. B* **48**, 13115–13126. <https://doi.org/10.1103/PhysRevB.48.13115> (1993).
37. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186. <https://doi.org/10.1103/PhysRevB.54.11169> (1996).
38. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979. <https://doi.org/10.1103/PhysRevB.50.17953> (1994).
39. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775. <https://doi.org/10.1103/PhysRevB.59.1758> (1999).
40. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868. <https://doi.org/10.1103/PhysRevLett.77.3865> (1996).
41. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple [Phys. Rev. Lett. 77, 3865 (1996)]. *Phys. Rev. Lett.* **78**, 1396–1396. <https://doi.org/10.1103/physrevlett.78.1396> (1997).
42. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements H–Pu. *J. Chem. Phys.* **132**154104. <https://doi.org/10.1063/1.3382344> (2010).
43. Grimme, S., Hansen, A., Brandenburg, J. G. & Bannwarth, C. Dispersion-corrected mean-field electronic structure methods. *Chem. Rev.* **116**, 5105–5154. <https://doi.org/10.1021/acs.chemrev.5b00533> (2016).
44. Rêgo, C. R. C., Oliveira, L. N., Tereshchuk, P. & Da Silva, J. L. F. Comparative study of Van der Waals corrections to the bulk properties of graphite. *J. Phys. Condens. Matter* **27**, 415502. <https://doi.org/10.1088/0953-8984/27/41/415502> (2015).
45. Koelling, D. D. & Harmon, B. N. A technique for relativistic spin-polarised calculations. *J. Phys. C: Solid State Phys.* **10**, 3107 (1977).
46. Takeda, T. The Scalar Relativistic Approximation. *Z. Phys. B: Condens. Matter Quanta* **32**, 43–48. <https://doi.org/10.1007/bf01322185> (1978).
47. Lejaeghere, K. et al. Reproducibility in density functional theory calculations of solids. *Science* **351**, 1415. <https://doi.org/10.1126/science.aad3000> (2016).
48. Mardirossian, N. & Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **115**, 2315–2372. <https://doi.org/10.1080/00268976.2017.1333644> (2017).
49. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268. <https://doi.org/10.1080/00268978400101201> (1984).
50. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697. <https://doi.org/10.1103/PhysRevA.31.1695> (1985).
51. Caldeira Rego, C. R. et al. SimStack: An intuitive workflow framework. *Front. Mater.* <https://doi.org/10.3389/fmats.2022.877597> (2022).
52. Schaarschmidt, J. et al. Workflow engineering in materials design within the Battery 2030+ project. *Adv. Energy Mater.* <https://doi.org/10.1002/aenm.202102638> (2021).
53. Bastos, C. et al. First-principles statistical investigation of thermodynamic behavior with excitonic effects in $Mo_{1-x}W_xSe_2$ alloys through a data-driven workflow approach. *J. Mater. Chem. A* <https://doi.org/10.1039/d5ta02721g> (2025).
54. Rêgo, C. R. C. et al. Digital workflow optimization of van der Waals methods for improved halide perovskite solar materials. (2025).
55. de Araujo, L. O. et al. Automated workflow for analyzing thermodynamic stability in polymorphic perovskite alloys. *npj Comput. Mater.* <https://doi.org/10.1038/s41524-024-01320-8> (2024).
56. Soleymanbrojeni, M., Caldeira Rego, C. R., Esmailpour, M. & Wenzel, W. An active learning approach to model solid–electrolyte interphase formation in Li-ion batteries. *J. Mater. Chem. A* **12**, 2249–2266. <https://doi.org/10.1039/D3TA06054C> (2024).
57. Bekemeier, S. et al. Advancing digital transformation in material science: The role of workflows within the materialdigital initiative. *Adv. Eng. Mater.* <https://doi.org/10.1002/adem.202402149> (2025).
58. Dalmedico, J. F. et al. Tuning electronic and structural properties of lead-free metal halide perovskites: A comparative study of 2d Ruddlesden–Popper and 3d compositions. *ChemPhysChem* <https://doi.org/10.1002/cphc.202400118> (2024).
59. Chaves, A. S., Piotrowski, M. J. & Silva, J. L. F. D. Evolution of the structural, energetic, and electronic properties of the $3d$, $4d$, and $5d$ transition-metal clusters (30tm_n systems for $n = 2–15$): A density functional theory investigation. *Phys. Chem. Chem. Phys.* **19**, 15484–15502. <https://doi.org/10.1039/c7cp02240a> (2017).
60. Yonezawa, A. F. et al. Stability changes in iridium nanoclusters via monoxide adsorption: A dft study within the van der waals corrections. *J. Phys. Chem. A* **125**, 4805–4818. <https://doi.org/10.1021/acs.jpca.1c02694> (2021).
61. Sousa, K. A. P. et al. Electrochemical, theoretical, and analytical investigation of the phenylurea herbicide fluometuron at a glassy carbon electrode. *Electrochim. Acta* **408**, 139945. <https://doi.org/10.1016/j.electacta.2022.139945> (2022).
62. Felix, J. P. C. S. et al. Molecular adsorption on coinage metal subnanoclusters: A DFT+D3 investigation. *J. Comput. Chem.* **44**, 1040–1051. <https://doi.org/10.1002/jcc.27063> (2023).
63. Hoppe, R. The coordination number — an “inorganic chameleon”. *Angew. Chem. Int. Ed.* **9**, 25–34. <https://doi.org/10.1002/anie.197000251> (1970).
64. Hoppe, R. Effective Coordination Numbers (ECoN) and Mean Active Fictive Ionic Radii (MEFIR). *Z. Kristallogr.* **150**, 23–52. <https://doi.org/10.1524/zkri.1979.150.1-4.23> (1979).

65. Momma, K. & Izumi, F. VESTA: A three-dimensional visualization system for electronic and structural analysis. *J. Appl. Cryst.* **41**, 653–658. <https://doi.org/10.1107/s0021889808012016> (2008).
66. Hammer, B. & Nørskov, J. K. *Advances in Catalysis* (Academic Press Inc, 2000).
67. Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*. International Series of Monographs on Chemistry (Clarendon Press, 1994).
68. Tang, W., Sanville, E. & Henkelman, G. A grid-based Bader analysis algorithm without lattice bias. *J. Phys. Condens. Matter* **21**, 084204. <https://doi.org/10.1088/0953-8984/21/8/084204> (2009).
69. Schaarschmidt, J. et al. Workflow engineering in materials design within the battery 2030+ project. *Adv. Energy Mater.* **12**, 2102638 (2021).
70. Mieller, B., Valavi, M. & Caldeira Rêgo, C. R. An automatized simulation workflow for powder pressing simulations using Simstack. *Adv. Eng. Mater.* <https://doi.org/10.1002/adem.202400872> (2024).
71. Pedregosa, F. et al. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
72. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc., B: Stat. Methodol.* **67**, 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x> (2005).
73. Nori, H., Jenkins, S., Koch, P. & Caruana, R. Interpretml: A unified framework for machine learning interpretability, <https://doi.org/10.48550/ARXIV.1909.09223> (2019).
74. Lou, Y., Caruana, R., Gehrke, J. & Hooker, G. Accurate intelligible models with pairwise interactions. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD' 13*, 623–631, <https://doi.org/10.1145/2487575.2487579> (ACM, 2013).
75. Pettifor, D. G. *Bonding and Structure of Molecules and Solids* (Oxford University Press, 1995).
76. Hammer, B. & Nørskov, J. K. Theoretical surface science and catalysis—calculations and concepts. *Adv. Catal.* **45**, 71–129. [https://doi.org/10.1016/S0360-0564\(02\)45013-4](https://doi.org/10.1016/S0360-0564(02)45013-4) (2000).
77. Nørskov, J. K. et al. Universality in heterogeneous catalysis. *J. Catal.* **209**, 275–278. <https://doi.org/10.1006/jcat.2002.3615> (2002).
78. Rodriguez, J. A. The chemical properties of bimetallic surfaces: Importance of ensemble and electronic effects in the adsorption of sulfur and so₂. *Prog. Surf. Sci.* **81**, 141–189. <https://doi.org/10.1016/j.progsurf.2006.02.001> (2006).
79. Sabatier, P. La. *Catalyse en Chimie Organique* (Encyclopédie de Science Chimique Appliquée, Librairie Polytechnique, Paris et Liège, 1913).
80. Medford, A. J. et al. From the sabatier principle to a predictive theory of transition-metal heterogeneous catalysis. *J. Catal.* **328**, 36–42. <https://doi.org/10.1016/j.jcat.2014.12.033> (2015).

Acknowledgements

We acknowledge support by the KIT Publication Fund of the Karlsruhe Institute of Technology. The authors are thankful for financial support from the National Council for Scientific and Technological Development (CNPq, grant numbers 303206/2025-0, 444431/2024-1, 408144/2022-0, 305174/2023-1, and 444069/2024-0), the Rio Grande do Sul Research Foundation (FAPERGS, grant number 24/2551-0001551-5), Federal District Research Support Foundation (FAPDF, grants 00193-00001817/2023-43 and 00193-00002073/2023-84), PDPG-FAPDF-CAPES Centro-Oeste (grant number 00193-00000867/2024-94), and the Coordination for Improvement of Higher Level Education (CAPES). In addition, the authors thank the “Centro Nacional de Processamento de Alto Desempenho em São Paulo” (CENAPAD-SP, UNICAMP/FINEP - MCTI project) for resources into the 897 and 570 projects, Lobo Carneiro HPC (NACAD) at the Federal University of Rio de Janeiro (UFRJ) for resources (133 project) CIMATEC SENAI at Salvador – BA, Brazil for the partnership, support through the Ogun Supercomputer, and “Laboratório Central de Processamento de Alto Desempenho” (LCPAD) financed by FINEP through CT-INFRA/UFPR projects and Santos Dumont/LNCC.

Author contributions

M.J.P. conceived the project. M.J.P., R.L.T.P., D.G.S., C.R.C.R., and A.C.D. provided the computational resources. R.F.M., J.M.T.P., T.G.G., and M.J.P. performed the calculations. O.R.F., M.J.P., and C.R.C.R. curated the data and developed the algorithmic codes. R.F.M., J.M.T.P., T.G.G., O.R.F., K.E.A.B., and M.J.P. prepared the figures and tables and drafted the initial version of the manuscript. R.L.T.P., D.G.S., C.R.C.R., A.C.D., and K.E.A.B. contributed to data curation and critically reviewed the manuscript. M.J.P. and C.R.C.R. supervised the project and led the manuscript preparation. All authors contributed to the writing and revision of the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. The project (HGF ZT-I-PF-5-261 GENIUS) underlying this publication is/was funded by the Initiative and Networking Fund of the Helmholtz Association in the framework of the Helmholtz AI project call.

Declarations

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/s41598-026-50998-x>.

Correspondence and requests for materials should be addressed to C.R.C.R.

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

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026