

# **ARCADE: A New Paradigm for a Fully Automated and Interpretable Design of Complex Catalysts**

Jianzhuo Wu,<sup>1,\*</sup> Xue Jia,<sup>2,\*</sup> Yongfeng Guo,<sup>1,\*</sup> Honglin Li,<sup>1</sup> Yu Li,<sup>1</sup> Yuhang Wang,<sup>2</sup> Jialu Li,<sup>1</sup> Guangsheng Liu,<sup>1</sup> Yufei Ding,<sup>3</sup> Hao Li,<sup>2,†</sup> and Wan-Lu Li<sup>1,4,‡</sup>

<sup>1</sup>*Aiiso Yufeng Li Family Department of Chemical and Nano Engineering,  
University of California, San Diego, La Jolla, CA, USA*

<sup>2</sup>*Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, Sendai, Japan*

<sup>3</sup>*Department of Computer Science and Engineering,  
University of California, San Diego, La Jolla, CA, USA*

<sup>4</sup>*Program in Materials Science and Engineering,  
University of California, San Diego, La Jolla, CA, USA*

## Abstract

The rational design of high-performance catalysts is pivotal for modern industry but is often hindered by the time-consuming nature of trial-and-error methods. While computational and data-driven approaches have accelerated materials discovery, significant challenges remain, especially for complex systems like multielement nanoparticles. These challenges include immense computational costs, difficulties in modeling stability and heterogeneous active sites, and a lack of interpretability in "black-box" AI models. To address these multifaceted challenges, we introduce ARCADE (Automated Rational CAtalyst DEsign), a generalizable and transferable framework for the automated and interpretable design of complex catalysts. Our framework integrates a high-throughput workflow that synergizes realistic structure generation, active-site screening, state-of-the-art pH-dependent microkinetic modeling, and short-range order (SRO) guided interpretability, establishing an end-to-end design pipeline that leverages public databases. Using the challenging design of multielement nanoparticle catalysts for CO<sub>2</sub> reduction reaction (CO<sub>2</sub>RR) as a representative system, we demonstrate that ARCADE can systematically navigate the design process, from evaluating material stability to identifying site-specific activity and selectivity. Our results show that this methodology substantially accelerate the development cycle for complex catalysts, confirming its feasibility and broad applicability. Crucially, this work establishes a new, fully integrated paradigm for the rational design and discovery of complex materials for challenging chemical reactions.

## I. Introduction

Electrocatalysts play a pivotal role in driving key energy conversion reactions, enabling the efficient transformation of abundant molecules (e.g., H<sub>2</sub>O, CO<sub>2</sub>, and N<sub>2</sub>) into value-added products such as hydrogen, synthetic fuels, and ammonia.[1] Over the past two decades, theory-guided catalyst discovery, such as the development of adsorption-energy scaling relations [2] and the seminal d-band theory[3], has provided fundamental insights into heterogeneous catalysis. More recently, advances in data-driven artificial intelligence, including the use of machine learning (ML) to predict electrocatalyst properties and to accelerate the understanding of dynamic catalytic mechanisms through Machine Learning

---

\* These authors contributed equally to this work.

† Email: li.hao.b8@tohoku.ac.jp

‡ Email: wal019@ucsd.edu

Potentials (MLPs), have further transformed catalyst research and optimization.[4, 5] However, despite these methodological advances, rationally designing complex catalytic systems, particularly multielement nanoparticles, remains limited due to their inherent structural and compositional complexity.[6, 7]

Multielement nanoparticles, typically composed of three to five or more principal elements, have recently emerged as promising candidates for heterogeneous catalysis owing to their abundant active sites, tunable surface chemistry, and entropy-stabilized structure.[8, 9] Nevertheless, several formidable challenges remain in guiding their rational design as electrocatalysts: (i) the structural complexity driven by random multielement mixing make it difficult to identify thermodynamically stable atomic-packing structure models;[10] (ii) the irregular morphology and complex surface structure of nanoparticles make the manual placement of adsorbates difficult;[11] (iii) the lack of interpretable structure–performance relationships that link the atomic structure of multielement nanoparticles with specific catalytic reaction schemes hinders rational catalyst design.[12] To overcome these challenges, several theoretical and ML approaches have been proposed;[4] however, their applicability remains limited due to difficulties in accurately capturing structural disorder and chemical complexity.

Traditional nanoparticle modeling relies on Wulff construction and global structure search. Wulff construction offers a simple way to rapidly predict equilibrium shapes from density functional theory (DFT)[13, 14] derived surface energies but neglect nonequilibrium growth, solvent or ligand effects, and multielement reconstruction.[15–17] Global optimization packages such as the USPEX or CALYPSO identify low-energy configurations, yet the search space grows combinatorially, and energy evaluations are costly.[18, 19] More importantly, they mainly utilize symmetry to extract structural features, which is often applied to crystals rather than nanoparticles. In multielement nanoparticles, thermodynamics drives intrinsically nonrandom mixing, which manifests as short-range order (SRO), and the advanced characterization studies have directly visualized SRO and confirmed its significant impact on structural defects and mechanical properties.[20] Explicitly accounting for SRO is therefore essential for generating realistic multielement nanoparticle structural models. On the other hand, ML interatomic potentials achieve near-DFT accuracy at molecular dynamics speed, enabling large-scale nanosecond simulations that capture realistic structural ensembles and dynamic reconstruction beyond the limits of Wulff or global optimization

methods.[21, 22] For nanoparticle modeling, developing a stable structure-search methodology that combines SRO-based structural feature representation with MLP-accelerated optimization is highly promising.

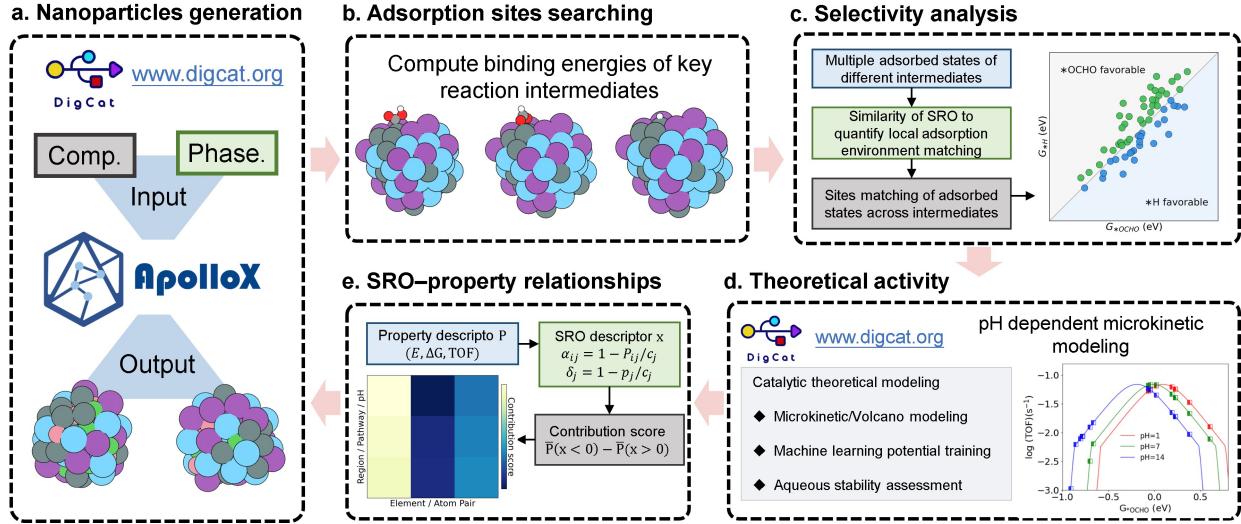
Adsorption-site determination commonly begins with symmetry-based enumeration on ideal slabs (e.g., the Pymatgen’s AdsorbateSiteFinder[23], CatKit/ASAP[11]), which is efficient but presumes crystallinity and becomes intractable as coverage increases or disorder grows. Global co-optimization frameworks that relax adsorbate and surface concurrently using surrogate MLPs or advanced potential energy surface (PES) explorers (Global Optimization using Bayesian Ensemble Estimation (GOFEE)[24], Solid-State Surface Walk (SSW)[25], San Diego Global Minimum Search (SDGMS)[26]) better escape local minima, yet they are computationally demanding and require careful uncertainty management on heterogeneous, low-symmetry nanoparticles. Recent high-throughput studies also highlight the combinatorial explosion when enumerating multi-site coverages and thus resort to stochastic sampling/filters[27, 28] underscoring the need for automated, geometry-aware active-site search beyond ideal surfaces to handle edges, kinks, and disordered nanoparticle facets. Automated and comprehensive active-site search frameworks are thus required to explore possible adsorption configurations and evaluate their adsorbate binding energies.

Regarding interpretable structure–activity relationships, descriptor-based catalytic volcano models are widely used[29, 30], but the conventional use of a fixed correction term ( $0.059 \times \text{pH}$ ) for pH-dependent energetics is inadequate to describe activity on the reversible hydrogen electrode (RHE) scale.[31] Incorporating pH-dependent thermodynamics/kinetics into descriptor-based analyses yields pH-dependent microkinetic volcano models that more faithfully capture activity evolution under realistic operating conditions and the RHE scale[32–35]. Additionally, the beneficial effects of SRO on catalytic properties, including selectivity and activity, are increasingly being recognized.[36–38] Notably, even though enhancements in catalytic performance related to SRO have been observed, establishing more systematic links between SRO features and catalytic performance is crucial for a comprehensive, mechanism-level understanding of complex multielement nanoparticle electrocatalysts. Ultimately, leveraging SRO-performance relationships to understand catalytic performance would enable the rational and interpretable design of catalysts.

To address these challenges, herein, we propose a new paradigm for automated and interpretable catalyst design, termed ARCADE (Automated Rational CAtalyst DEsign). This

generalizable framework encompasses thermodynamically stable nanoparticle generation by introducing SRO-based structural feature, automated active-site identification and binding energy evaluation, selectivity analysis through comparisons of Gibbs free energies among competing reactions, and activity analysis via advanced pH-dependent microkinetic modeling. SRO-derived descriptors are further employed to construct structure-performance relationships. Notably, the entire workflow is implemented within the Digital Catalysis Platform (*DigCat*)[39], where data generation, model construction, and theoretical analyses can be executed automatically. The resulting datasets and mechanistic insights are continuously fed back into the *DigCat* platform’s database, establishing a closed-loop ecosystem that accelerates the rational design and discovery of next-generation electrocatalysts.

In this work, the electrochemical CO<sub>2</sub> reduction reaction (CO<sub>2</sub>RR) was selected as a representative system to demonstrate the efficiency of ARCADE. Among the possible CO<sub>2</sub>RR products, formic acid/formate (HCOOH) is particularly attractive due to its high energy density, ease of storage, and wide applicability in chemical, pharmaceutical, and energy-related industries.[40, 41] Guided by initial compositional screening on the *DigCat* platform, two representative nanoparticle systems, medium-entropy La-Sn-Bi and high-entropy La-Sn-Bi-Co-Ni, were investigated. The most thermodynamically stable structures were rapidly determined via a Monte Carlo (MC)[42] based SRO-constrained conditional generative protocol, coupled with MLP-accelerated global search algorithm. After systematically searching all possible adsorption sites, the adsorption energies of \*OCHO, \*COOH, and \*H intermediates were calculated. Both compositions exhibited selectivity toward HCOOH formation and reached the peak of the theoretical volcano model, indicating high catalytic activity superior to that of the corresponding elemental nanoparticles. The constructed SRO-performance score heatmaps successfully decoupled the intertwined dependencies of stability, selectivity, and activity on local structural features, revealing clear structure-function relationships in compositionally complex catalysts. Our analysis reveals that surface Bi enrichment and bulk Co/Ni enrichment govern thermodynamic stability, while local environments enriched in La and Sn are the primary drivers for selectivity and activity, respectively, with specific atomic pairwise patterns further elucidating the origins of these metrics. Ultimately, this work establishes a transformative paradigm for decoding ’interpretable complexity,’ offering a systematic route to map intricate atomic-level disorder onto predictable catalytic function, thereby paving the way for the rational and efficient design of complex catalysts.



**FIG. 1. ARCADE workflow for data-driven nanoalloy catalysis.** **(a)** Nanoparticle generation: Given only candidate compositions and a parent phase specified via our Digital Catalysis Platform (*DigCat*), the ApolloX multicomponent amorphous-structure modeling engine samples thermodynamically stable nanoalloy structures and returns atomistic nanoparticles. **(b)** Adsorption-site searching: On each nanoparticle, possible adsorption motifs of key intermediates are automatically enumerated and relaxed to obtain binding energies. **(c)** Selectivity analysis: Computed Gibbs free energies for competing intermediates (e.g.,  $G_{*OCHO}$  vs  $G_{*COOH}$ ,  $G_{*H}$ , etc.) are cross-compared to map selectivity trends. **(d)** Theoretical activity: pH-dependent microkinetic modeling (implemented on the *DigCat*) translates adsorption energetics into predicted turnover frequencies and pH-dependent microkinetic volcano relationships. **(e)** Structure-property relationships: Aggregating results across composition/structure space enables interpretable maps that link local atomic motifs and elemental mixing patterns to stability, selectivity, and activity.

## II. Results

### A. Overall Workflow of ARCADE

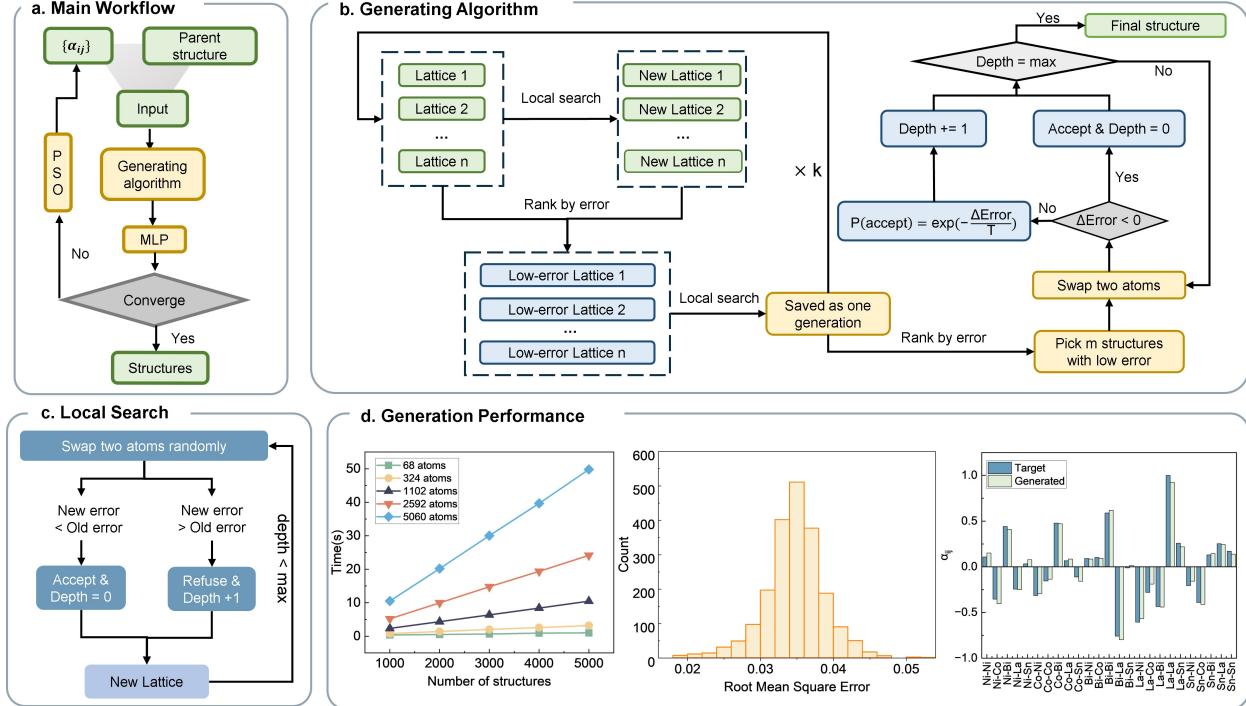
The ARCADE workflow establishes a fully data-driven framework for nanoalloy catalyst discovery, integrating structure generation, active-site identification, selectivity and activity analysis, and structure-performance relationship evaluation. All components of this workflow have been implemented within the *DigCat* platform, enabling a closed-loop discovery

framework. The workflow is shown in (**Fig. 1**). First, we deploy our upgraded **ApolloX** (Automatic Prediction by generative mOdel for Large-scaLe Optimization of X-composition materials)[43] engine to generate thermodynamically stable multicomponent nanoparticles based on user-defined compositions and parent phases, ensuring realistic structural configurations (**Fig. 1a**). By integrating MLPs with Particle Swarm Optimization (PSO) [44] and MC [42] sampling, ApolloX explicitly incorporates realistic SRO, enabling the construction of physically reasonable atomic configurations for multielement nanoparticles.

Next, the workflow automatically identifies all possible adsorption sites based on Voronoi tessellation[45, 46] and constructs absorbed structures utilizing the Atomic Simulation Environment (ASE)[47], followed by the structural optimization of key reaction intermediates to determine their binding energies (**Fig. 1b**). Then, the local adsorption environment of each intermediate can be characterized using SRO descriptors. By identifying similar local environments among different intermediates, the corresponding adsorption free energies (e.g., of  $^{*}\text{COOH}$ ,  $^{*}\text{OCHO}$ , and  $^{*}\text{H}$ ) can be compared to reveal the preferred reaction pathways and determine product selectivity (**Fig. 1c**).

Subsequently, pH-dependent microkinetic volcano modeling implemented on DigCat translates these energies into predicted turnover frequencies (TOFs) (**Fig. 1d**), explicitly capturing the variation of catalytic performance across different pH environments (detailed theoretical derivations are provided in the **Supporting Information**).

Finally, by aggregating data across diverse compositions and structures, the workflow integrates SRO descriptors, specifically the Local Density Deviation (LDD) and Warren-Cowley (W-C) parameters[48–50] (formally defined in the **Methods Section**), with property descriptors such as energies, adsorption free energy differences, and turnover frequencies (TOFs). By calculating a contribution score for specific SRO patterns, ARCADE constructs interpretable structure-performance maps that directly correlate local atomic motifs and elemental mixing patterns with catalytic performance, enabling rational and high-throughput catalyst design (**Fig. 1e**). Note that while this interpretability module is presented here as the final integration step, its specific analyses are applied iteratively throughout the study to decode stability, selectivity, and activity in turn.



**FIG. 2. The updated ApolloX workflow for structure generation and its performance.**

**(a)** Main workflow of ApolloX for structural modeling, including the model’s iterative generation process. **(b)** Details of the hybrid Generating algorithm. This algorithm consists of two stages: First, two successive steps of a greedy local search (defined in c) are applied to quickly generate a set of low-error lattices; second, these structures are further refined using a Metropolis-based search to precisely match the target SRO parameters ( $\alpha_{ij}$ ). **(c)** The greedy local search routine used in the first stage of (b). It performs random atom swaps and only accepts changes that reduce the error, enabling rapid initial optimization. **(d)** Performance benchmarking of quinary catalysts with varying sizes and quantities. (Left) Computational time as a function of the number of structures (for different numbers of atoms) on an NVIDIA 4090 GPU. (Middle) Statistical distribution of the SRO root mean square error (RMSE) for 2000 generated structures, showing an average RMSE of 0.035. (Right) A representative comparison between target  $\alpha_{ij}$  values (Target) and the generated values (Generated) for a structure with the average RMSE of 0.035.

## B. Workflow of Structure Generation and Linking Thermodynamic Stability to Short-Range Order

The first step in catalyst design is to identify thermodynamically stable structures for potential catalysts of the target reaction. To achieve this, we integrate initial compositional determination based on the *DigCat* platform and stable structure generation implemented by the ApolloX. *DigCat* hosts comprehensive literature and structural data corresponding to diverse chemical reactions and target products, facilitating the initial identification of candidate compositions and structural parent phases. After determining the catalyst components and the structural parent phase by *DigCat*, we deploy an upgraded version of ApolloX to generate stable structures for complex materials. The main workflow (**Fig. 2a**) of ApolloX operates as an iterative optimization loop. We treat SRO as a key structural feature, transforming the highly challenging structure-energy iterative optimization into a practical and feasible structure-SRO-energy iterative optimization. SRO is described using the W-C parameter  $\alpha_{ij}$ [48–50], where i and j represent different types of elements. On one hand, a MC scheme[42] is adopted to probe elemental mixing and local motif rearrangements, tightly linking SRO to structural configurations. This serves as a Generating Algorithm, enabling the generation of structures corresponding to a given SRO. On the other hand, a PSO algorithm[44] is employed to tune the SRO-energy relationship. Ultimately, this approach allows for the efficient search and generation of thermodynamically stable structures. The Generating Algorithm (**Fig. 2b**) is a hybrid, two-stage process. The first stage, which corresponds to the "global search" mentioned in the text, is a rapid greedy local search (**Fig. 2c**). It performs random atomic swaps and only accepts changes that reduce the SRO error (defined as  $\frac{\sqrt{\sum_{i,j}[\alpha_{ij}-\alpha_{ij}(\text{target})]^2}}{\sum_{i,j}}$ ), quickly producing a set of low-error candidate structures. This candidate pool is then passed to the second stage, or "detailed search," which employs a Metropolis criterion. This Metropolis search allows for a controlled exploration of the configurational space (by probabilistically accepting some higher-error swaps), enabling the algorithm to escape local minima and converge more precisely on the target  $\alpha_{ij}$  values.

Focusing on the electrochemical CO<sub>2</sub>RR to formate as a representative system, we first utilized the *DigCat* platform to perform a rigorous compositional screening to define the precise catalytic systems for our structural generation and subsequent benchmarking. From 8,439 retrieved literature records on CO<sub>2</sub>RR electrocatalysts yielding formate, 30.95% em-

ployed metal or alloy catalysts. A comprehensive summary of Faradaic efficiency data reported over the past decade is provided in Table S1, while the statistical distribution of constituent elements involved is presented in Fig. S1. Analysis of these entries revealed that Bi and Sn are the most frequently reported metal components[51–54]. It is also worth noting that recent studies indicate that La incorporation can markedly enhance catalytic performance.[55] Consequently, we targeted La-Sn-Bi as our medium-entropy alloy (MEA) system. Furthermore, guided by elements commonly incorporated into high-entropy alloys[56–58], we expanded this composition with Co and Ni to design a La-Sn-Bi-Co-Ni high-entropy alloy (HEA) system. Since most structural entries in the *DigCat* platform correspond to face-centered cubic (FCC) parent lattices, we adopted FCC lattices to construct our nanoparticle prototypes. Specifically, for both the ternary and quinary systems, we first constructed sufficiently large equimolar FCC supercells. By selecting the geometric center of the structure and applying varying cutoff radii, we generated spherical nanoparticle parent phases with five distinct sizes ranging from 2 to 8 nm, as illustrated in Fig. S2.

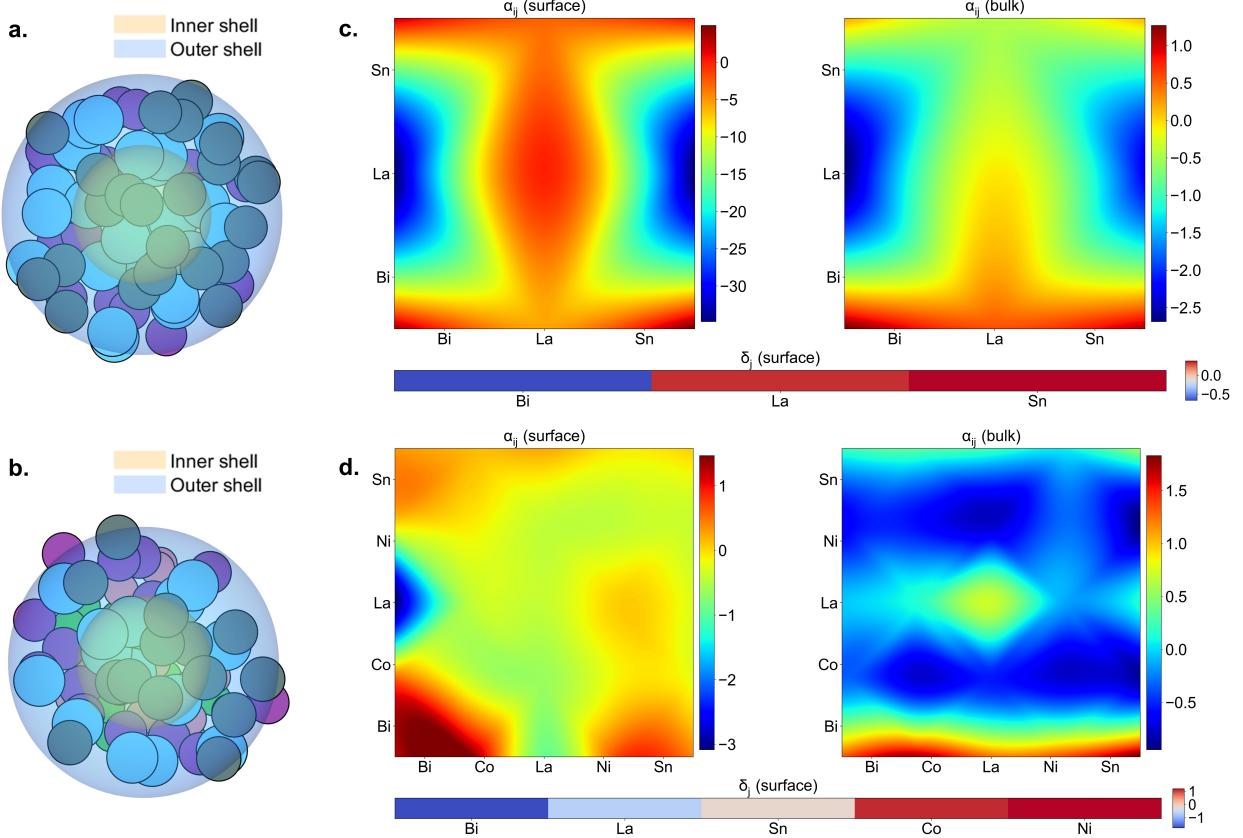
These generated nanoparticle models served as specific benchmarks to evaluate the performance of the updated generation module within the ApolloX engine. Focusing on the complex quinary system, the capability to generate structures based on target SRO is detailed in **Fig. 2d**. The algorithm demonstrates exceptional scalability, producing  $\tilde{5}000$  structures for a  $\tilde{5}000$ -atom system in under 50 seconds, while maintaining near-constant generation time per structure even as system size increases (**Fig. 2d**, left). Furthermore, it shows high fidelity: a statistical analysis of 2000 generated structures reveals a tight distribution of SRO root mean square errors (RMSE) centered around an average of 0.035 (**Fig. 2d**, middle). A representative comparison for a structure with this average error confirms the close match between the target and generated  $\alpha_{ij}$  values (**Fig. 2d**, right). Additionally, we conducted performance tests on ternary catalysts across varying sizes. As shown in Fig. S3, the module exhibits similarly outstanding performance, demonstrating the broad applicability and robustness of this generation function.

Although the MC-based generation method and MLPs are capable of handling significantly larger structures, we prioritized computational feasibility for the subsequent high-precision DFT calculations of adsorption energies. Therefore, we selected the 2 nm-sized models, specifically  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  and  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$  as the final catalyst compositions. The specific atomic ratios in these models represent the natural outcome of cutting

an equimolar FCC supercell into a nanoparticle shape. Subsequently, we applied the complete ApolloX workflow for structural generation. For each composition, we performed 15 rounds of PSO iterative cycles, with each cycle producing 100 structures. We utilized Crystal Hamiltonian Graph Neural Network (CHGNet)[59] as the surrogate MLP, and its performance benchmark (Fig. S4) confirms that it meets the required accuracy for this study. The relationship between the minimum energy of all structures and the iteration count for ternary and quinary catalysts is presented in Fig. S5 and Fig. S6, respectively. The results indicate that the total minimum energy gradually decreases and stabilizes as iterations progress, signifying a successful convergence toward the global minimum. To further validate this optimization capability, we compared the energy distribution of the 1,500 structures generated by ApolloX against a baseline of randomly shuffled structures (Fig. S7). The results demonstrate that ApolloX identifies a significantly higher proportion of structures within the low-energy regime and achieves lower absolute minimum energies than the random method, further underscoring the effectiveness of our approach.

Leveraging the extensive structural ensemble generated by the ApolloX, we next systematically investigated how SRO within the nanoparticles governs their thermodynamic stability. To disentangle surface and interior effects, each nanoparticle was partitioned into a surface shell and a bulk shell (**Fig. 3a-b**). The partition was defined using two concentric spheres centered at the geometric center of all catalyst atoms; the outer radius equals the maximum atomic distance from the center, and the inner radius is half of this maximum.

We characterized SRO using two descriptors: (i) the W-C parameter  $\alpha_{ij}$ , which captures aggregation vs dispersion between atomic pairs, and (ii) the LDD parameter  $\delta_j$ , which measures local enrichment or depletion of element  $j$ . Based on their mathematical definitions,  $\alpha_{ij}$  can be directly interpreted as describing the distribution of  $j$  atoms around  $i$  atoms, while  $\delta_j$  represents the distribution of the  $j$  element itself. Therefore, LDD parameter allows for a quick and convenient understanding of the individual contributions of different elements, while W-C parameter provides more detailed insights into the short-range ordering and structural characteristics. Notably, the two descriptors exhibit both consistency and complementarity. Their progressive relationship reflects a shared description of local chemical environments, while the enrichment-depletion trends captured by  $\alpha_{ij}$  and the aggregation-dispersion trends represented by  $\delta_j$  highlight distinct structural features. In this sense, the two descriptors offer complementary perspectives on short-range ordering.



**FIG. 3. Statistical contribution of short-range order to thermodynamic stability and schematic structures.** (a-b) Schematics of  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  (a) and  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$  (b) indicating inner-outer shells and the surface–bulk partition. (c-d) Surface and bulk Warren-Cowley parameter  $\alpha_{ij}$  score maps for  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  (c) and  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$  (d). The horizontal bar below reports surface Local Density Deviation parameter  $\delta_j$  contributions. The color scale is  $\overline{E}(x < 0) - \overline{E}(x > 0)$  (eV), where  $x \in \{\alpha_{ij}, \delta_j\}$ .

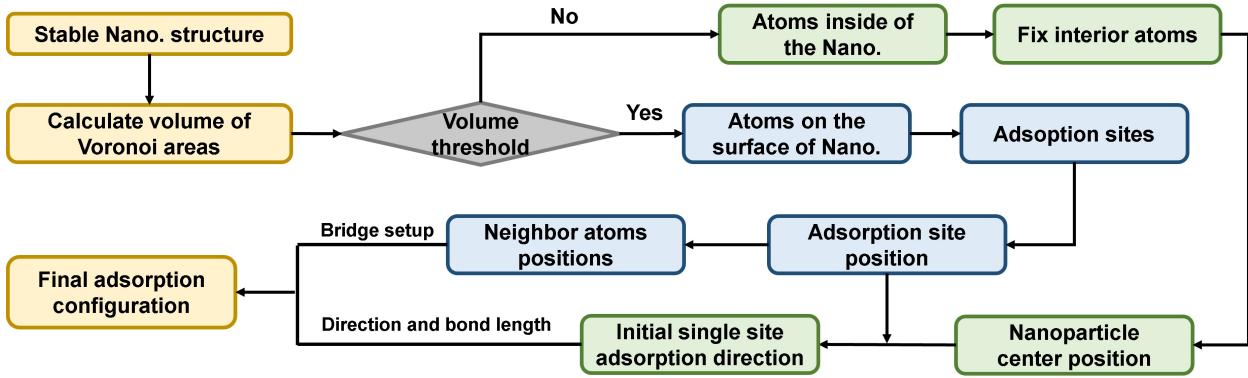
For each composition, we evaluated SRO separately in the bulk and surface shells across 1,500 structures. For each descriptor ( $\alpha_{ij}$  or  $\delta_j$ ), structures were partitioned by the sign of the descriptor. We then computed the energy contrast  $\overline{E}(x < 0) - \overline{E}(x > 0)$  (eV) as a score to quantify the stabilizing influence of that SRO motif, where  $x \in \{\alpha_{ij}, \delta_j\}$ . The results for all descriptors were combined and visualized as score maps (Fig. 3c-d). Negative values (blue) indicate that  $x < 0$  is associated with lower energy, i.e., aggregation of the pair or local enrichment of the element stabilizes the structure; positive values (red) indicate the opposite tendency, favoring dispersion of the corresponding pair or depletion of the element.

Applying this framework yields the following trends. For the  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$ , at the surface,

local composition tends to be enriched in Bi and depleted in Sn and La. Pairwise ordering in both surface and bulk tends to favor aggregation of Bi–La and Sn–La pairs, while disfavoring Sn–Bi and Bi–Bi pairs through dispersion. For the  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$ , at the surface, local composition tends to be enriched in Bi and La and depleted in Ni and Co. Surface ordering favors aggregation of Bi–La and disfavors Bi–Bi, Ni–Bi, and Sn–Bi. In the bulk area, ordering tends to favor aggregation among Co- and Ni-containing pairs (Co–Co, Ni–Co, Sn–Co, La–Ni, Sn–Ni), while disfavoring Co–Bi and Sn–Bi. Overall, Bi enrichment in the surface region and Co/Ni enrichment in the bulk both enhance the structural stability of the nanoparticle catalyst.  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  exhibits a strong consistency between bulk and surface in the aggregation/dispersion tendencies of atomic pairs, whereas  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$  shows pronounced differences between bulk and surface due to the bulk aggregation propensity of Co and Ni, highlighting the significant impact of multicomponent high-entropy effects on structure. Specifically, driven by the differences in surface energies and cohesive energies among constituent elements, high-entropy nanoparticles often exhibit non-uniform elemental distribution (e.g., surface segregation), resulting in distinct SRO between the surface and the bulk.[8, 20] These observations both substantiate the rationale for selecting Co and Ni as components of the HEA catalyst and reveal that the role of Bi extends beyond its commonly noted influence on activity[60–64] to include a substantial contribution to stability.

### C. Automated Adsorption Sites Search and Absorbate Binding Energy Calculations

After structure generation, ARCADE systematically integrates adsorption-site identification and energy evaluation for the most stable structure. We employ an autonomous, geometry-aware search protocol based on Voronoi tessellation[45, 46] to distinguish surface adsorption sites from interior atoms (**Fig. 4**). The Voronoi cell of each atom is constructed from the convex hull of its Voronoi vertices; interior atoms, being fully surrounded, have finite and relatively small cell volumes, whereas surface atoms lack neighbors in certain directions and thus exhibit much larger (sometimes unbounded) volumes (see Fig. S8 for the spatial partitioning illustration). We therefore apply a volume threshold  $V_{thr}$ : atoms with Voronoi cell volumes exceeding  $V_{thr}$  are classified as surface sites, while the remainder are fixed as interior atoms to mimic bulk constraints. This geometry-aware rule is robust to disorder and does not require predefining facets.



**FIG. 4. Schematic workflow for autonomous adsorption site identification.** Voronoi tessellation distinguishes surface sites from interior atoms based on a volume threshold. For identified surface sites, the workflow determines single-site adsorption geometries (direction and bond length) and identifies neighboring atom pairs for bridge setups, integrating these parameters to generate the final atomic configurations. Blue and green modules represent information derived from the surface and the integration of interior atoms, respectively. Note: Nano. in the flowchart is an abbreviation for nanoparticle.

For each identified surface site, the workflow derives configuration parameters through two general geometric criteria. First, for single-site adsorption, the adsorption direction is defined by the radial vector extending from the nanoparticle center through the target surface atom, with the initial bond length estimated based on the sum of the covalent radii of the adsorption site atom and the binding atom of the intermediate. Second, to establish setups for multi-site adsorption, the workflow identifies neighboring surface atom pairs and constructs the initial geometry by aligning the adsorbate's binding atoms along the radial vectors of these adjacent sites—analogous to the single-site protocol. The adsorption height is then adjusted along these directions to reconcile the intermediate's internal bond distances with the spatial separation of the surface sites, ensuring a physically plausible starting configuration prior to relaxation. For the CO<sub>2</sub>RR-to-formate pathway studied here, these protocols were applied to generate configurations for three key intermediates: \*COOH and \*H are treated as monodentate species using the single-site criteria (binding *via* the C atom and H atom, respectively), while \*OCHO is modeled as a bidentate species utilizing the bridge setup logic (binding *via* its two O atoms).

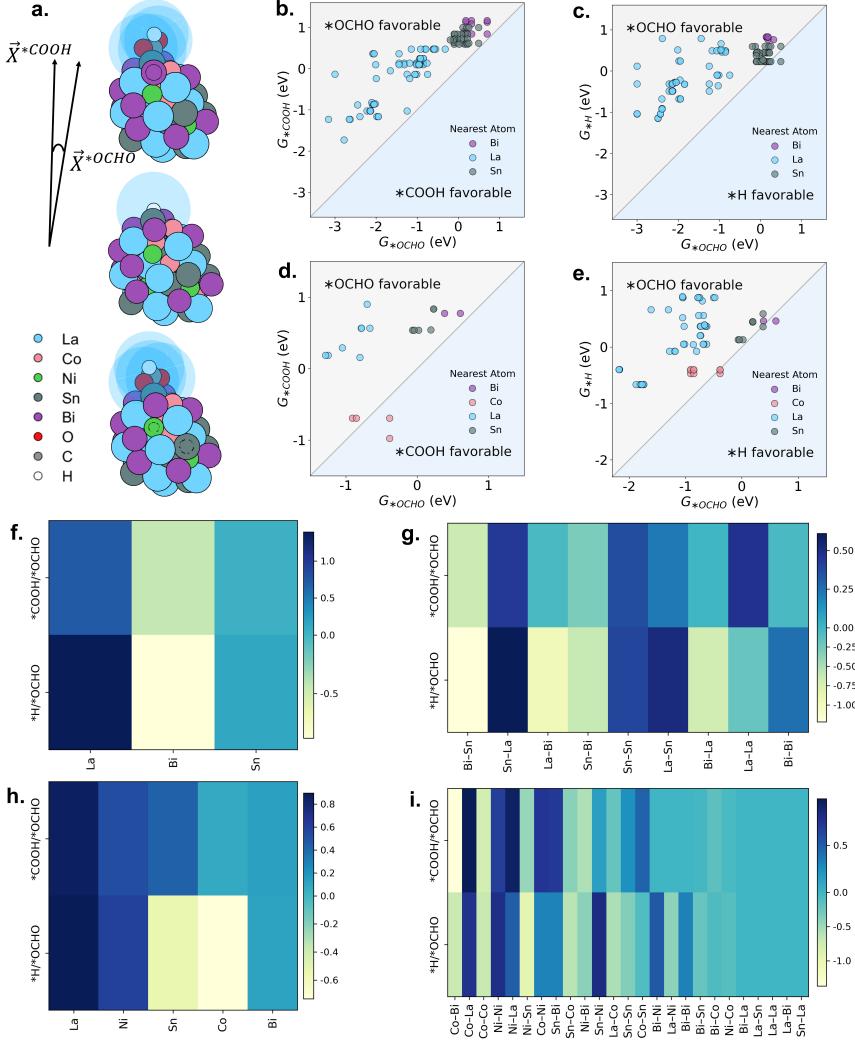
Based on the ARCADE screening results, the most stable candidates, La<sub>23</sub>Sn<sub>23</sub>Bi<sub>22</sub> and

$\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$ , were identified for further electrocatalytic analysis. Following the procedure illustrated in **Fig. 4**, the outermost atoms were determined and designated as adsorption sites, from which the adsorption structures of  $^*\text{COOH}$ ,  $^*\text{OCHO}$ , and  $^*\text{H}$  were constructed. For  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  nanoparticles, the outermost layer contains 53 atoms, yielding 53 possible adsorption structures each for  $^*\text{COOH}$  and  $^*\text{H}$ . For  $^*\text{OCHO}$ , which requires diatomic adsorption sites, 145 possible configurations were obtained. In the case of  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$  nanoparticles, the outermost layer comprises 41 atoms, corresponding to 41 adsorption structures for  $^*\text{COOH}$  and  $^*\text{H}$ , and 112 for  $^*\text{OCHO}$ . Adsorption energies for all configurations were subsequently calculated. Specifically, the calculated adsorption energies for  $^*\text{COOH}$ ,  $^*\text{H}$ , and  $^*\text{OCHO}$  on  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  are detailed in Tables S2-S4, respectively, while the corresponding data for  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$  are provided in Tables S5-S7. The resulting distributions of these adsorption energies are presented in Fig. S9. Detailed computational methods are provided in the Methods Section. Representative examples of adsorption configurations before and after optimization for different compositions and intermediates are illustrated in Fig. S10. As observed in these examples, although adsorbates were initially placed on predefined sites, they were allowed to relax freely during DFT geometry optimizations and, where energetically favorable, migrated to lower-energy positions. For  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$ , the optimized sites are:  $^*\text{COOH}$  on Sn, Bi, Ni, Co, or La;  $^*\text{OCHO}$  on diatomic sites composed of any two among Sn, Bi, Ni, Co, and La; and  $^*\text{H}$  on Sn, Bi, Ni, Co, La, or bridge sites. For  $\text{LaSnBi}$ , the optimized sites are:  $^*\text{COOH}$  on Sn, Bi, La, or bridge;  $^*\text{OCHO}$  on diatomic sites composed of any two among Sn, Bi, and La; and  $^*\text{H}$  on Sn, Bi, La, or bridge sites.

#### D. Selectivity Analysis and the Relationship with SRO

Starting with initial compositional guesses, followed by generating stable structures and performing adsorption site searches, we have established the structural modeling foundation for catalyst design. The structure-energy data obtained during structure generation have been used for catalyst stability analysis, while the adsorption site and adsorption free energy data for different intermediates from the adsorption site search will be utilized to evaluate the selectivity among competing pathways and the activity of the preferred reaction pathway.

To assess the selectivity of CO<sub>2</sub>RR, we compared the Gibbs free energies of  $^*\text{OCHO}$ ,



**FIG. 5. Catalytic selectivity analysis and the role of short-range order.** (a) Matching of intermediates configurations by local chemical environment near adsorbates, which is symbolized by the nearest adsorption site and the local SRO vector  $\mathbf{X} = \text{vec}(\alpha_{ij}) \oplus \delta_j$  (b-e) Selectivity analysis of  $G_{*OCHO}$  vs.  $G_{*COOH}$  and  $G_{*OCHO}$  vs.  $G_{*H}$  for  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  (b-c) and  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$  (d-e). (f-i) Score heat maps showing the contribution of short-range order on selectivity, quantified as  $\overline{\Delta G}(x < 0) - \overline{\Delta G}(x > 0)$ (eV), where  $x \in \{\alpha_{ij}, \delta_j\}$ . (f) and (h) report the Local Density Deviation parameter  $\delta_j$  contributions for  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  and  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$ , respectively; (g) and (i) report the Warren-Cowley parameter  $\alpha_{ij}$  contributions for  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  and  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$ , respectively.

\*COOH, and \*H adsorbed on the same site[35, 65, 66]. However, as noted above, the adsorption sites for these intermediates are not strictly identical: \*COOH and \*H typically

adsorb on a single site, whereas  $^*\text{OCHO}$  involves two adjacent atoms, and adsorbates may migrate during optimization. To address this inconsistency, we proposed using SRO to identify adsorption environments with high similarity, thereby enabling a fair comparison of their energies. We matched structures using: (i) the nearest adsorption-site atom and (ii) SRO descriptors characterizing the local environment around the adsorbate. Specifically, the “nearest adsorption-site atom” is defined as the catalyst atom participating in the pair with the shortest relative distance to any adsorbate atom, where the relative distance is the Euclidean separation divided by the sum of the two atoms’ covalent radii. To delineate a chemically meaningful local environment, we adopt an “effective interaction range” per adsorbate atom equal to the average of the sum of the van der Waals radii of the adsorbate and catalyst atoms, approximately 4.5 Å. The union of catalyst atoms lying within this range for all atoms of the adsorbate defines the adsorbate’s local chemical environment. Within this local environment, we computed: (i) the W-C parameter  $\alpha_{ij}$  and (ii) the LDD parameter  $\delta_j$  to quantify SRO. For each relaxed adsorption structure, the set of  $\alpha_{ij}$  and  $\delta_j$  values within the local environment are concatenated into a single feature vector  $\mathbf{X} = \text{vec}(\alpha_{ij}) \oplus \delta_j$  used as a label. Structures are then matched if they share the same nearest adsorption-site atom and if the cosine similarity between their local-environment SRO vectors is  $\geq 0.9$ . We chose the 0.9 threshold to accommodate minor differences in adsorbate orientation and in the range of the local environment, as well as small catalyst distortions induced during relaxation; this tolerance balances robustness with specificity. A schematic of the matching protocol for medium/high-entropy catalysts is shown in **Fig. 5a**.

For elemental catalysts, structures of a given adsorbate are matched using a geometric criterion: we computed the adsorbate position vector as the displacement from the catalyst geometric center to the adsorbate geometric center, and we required a cosine similarity  $\geq 0.9$  between these position vectors to establish a match.

For  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$ , the matching results and SRO cosine similarity data for  $^*\text{OCHO}$  vs.  $^*\text{H}$  and  $^*\text{OCHO}$  vs.  $^*\text{COOH}$  are provided in Tables S8 and S9, respectively. For  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$ , the corresponding data for  $^*\text{OCHO}$  vs.  $^*\text{H}$  and  $^*\text{OCHO}$  vs.  $^*\text{COOH}$  are detailed in Tables S10 and S11, respectively. After matching structures corresponding to each adsorbate, catalytic selectivity was analyzed as shown in **Fig. 5b-e** and Fig. S11. We quantify selectivity by comparing the binding energies of  $^*\text{OCHO}$  to those of the other two adsorbates. The difference in binding energy between adsorbates,  $\Delta G =$

$G_{*COOH/*H} - G_{*OCHO}$ , indicates both the direction and magnitude of selectivity: a positive and larger  $\Delta G$  implies that the reaction favors the \*OCHO pathway, i.e., higher selectivity.

**Fig. 5b-c** reports the \*OCHO vs. \*COOH and \*OCHO vs. \*H selectivity for  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$ , and **Fig. 5d-e** provides the corresponding results for  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$ , with the nearest adsorption-site atom annotated for each structure. All the results indicate the high  $\text{CO}_2\text{RR}$ -to-HCOOH selectivity for the nanoparticles. Fig. S11 shows the \*OCHO vs. \*COOH and \*OCHO vs. \*H selectivity for elemental catalysts. Comparing MEA/HEA catalysts with elemental catalysts, the MEA/HEA data points cluster preferentially toward the upper-left of the diagonal, indicating superior selectivity relative to the elemental references.

To explicitly assess how SRO affects catalytic selectivity, we group structures by the sign of each descriptor ( $\alpha_{ij}$  and  $\delta_j$ ), and then compare the difference between the groups' averages of  $\Delta G$ . Because the extremes of  $\Delta G$  (its maximum and minimum) are most consequential—setting the upper bound of selectivity and indicating whether any non-OCHO-favorable cases occur—we assign linearly varying weights to data points:

$$w_i = 1 - \gamma + 2\gamma \frac{|\Delta G_i - \Delta G_c|}{\Delta G_{max} - \Delta G_{min}}$$

where  $\Delta G_{max}$  and  $\Delta G_{min}$  are the maximum and minimum of  $\Delta G$ , respectively;  $\Delta G_c = \frac{\Delta G_{max} + \Delta G_{min}}{2}$  represents the center of  $\Delta G$  range; and  $\gamma$  controls the weighting strength, set to 0.8 in this work. The difference between the groups' weighted means is then:

$$\overline{\Delta G}(x < 0) - \overline{\Delta G}(x > 0) = \frac{\sum_{i,x<0} w_i \Delta G_i}{\sum_{i,x<0} w_i} - \frac{\sum_{j,x>0} w_j \Delta G_j}{\sum_{j,x>0} w_j}$$

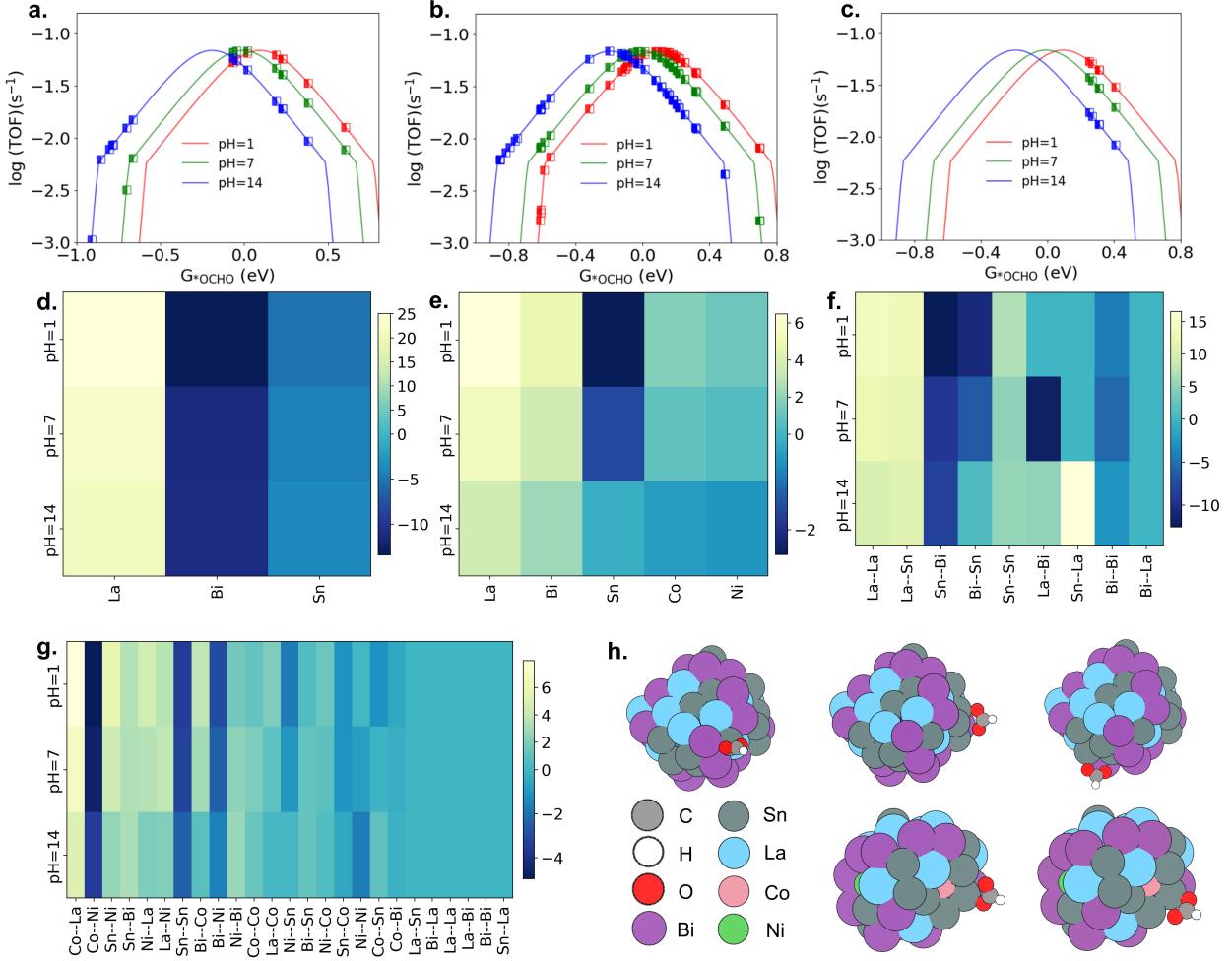
where  $x \in \{\alpha_{ij}, \delta_j\}$ . For each SRO descriptor, we compute the weighted-mean difference in  $\Delta G$  (eV). The results for all descriptors were combined and visualized as score heat maps (**Fig. 5f-i**). **Fig. 5f-g** shows the impacts of the LDD parameters  $\delta_j$  and the W-C parameters  $\alpha_{ij}$  on  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  selectivity; **Fig. 5h-i** shows the corresponding impacts for  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$ . Positive values (dark blue) indicate that the  $x < 0$  group has larger  $\Delta G$ , i.e., stronger preference for the \*OCHO pathway; negative values (light yellow) indicate that the  $x > 0$  group is more selective. Larger absolute values are shown by colors closer to the ends of the scale, indicating a stronger influence of the SRO descriptor on selectivity.

Synthesizing the site-specific comparisons (**Fig. 5b-e**) with the SRO-selectivity analysis (**Fig. 5f-i**) reveals some key trends governing the catalytic selectivity. For the  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  catalyst, enhanced selectivity is associated with adsorption sites where La is the nearest

atomic neighbor, a local environment enriched in La and Sn but depleted in Bi, and the aggregation of Sn–Sn and Sn–La pairs alongside the dispersion of Bi–Sn and Bi–Bi pairs. A similar pattern emerges in the  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$  system, where proximity to a La atom enhances selectivity while proximity to a Co atom suppresses it. In this system, a local environment enriched in La and Ni and depleted in Co is favorable, as is the aggregation of Co–La, Ni–La, Co–Ni, and Ni–Ni pairs coupled with the dispersion of Co–Bi and Co–Co pairs. Critically, a comparison of the two systems highlights a unifying principle: the presence of La, both as the nearest neighbor to the adsorption site and as an enriching element in the local chemical environment, is a key determinant of selectivity. In other words, beyond its experimentally reported role in La-doped  $\text{SnO}_2$ , where La enhances reaction selectivity *via* pinning effects and water activation[55], we also observe a pronounced influence of La on selectivity in our MEA/HEA catalysts. These findings further underscore the substantial application potential of La elements in CO<sub>2</sub>RR catalysis.

## E. Activity Analysis and the Relationship with SRO

Structures with different adsorbates but similar local chemical environments are matched through the cosine similarity of the short-range order vector  $\mathbf{X} = \text{vec}(\alpha_{ij}) \oplus \delta_j$ , and the \*OCHO favorable sites are identified by comparing  $G_{*OCHO}$  with  $G_{*COOH/*H}$ . Subsequently, the TOFs are predicted using our recently developed pH-dependent microkinetic modeling for CO<sub>2</sub>RR under the RHE scale,[35, 67] by comprehensively considering the electric field effects,[68] potential of zero-charges under explicit solvent conditions,[69] and solvation corrections via the H-bonding. Notably, this pH-dependent modeling method has excellent agreement with experimental results, in particular for those with CO<sub>2</sub>RR-to-formate[35, 67] and CO<sub>2</sub>RR-to-CO.[68] Details of the microkinetic modeling methods, the considered elementary steps, and all relevant parameters can be found in the **Supplementary Information**. These pH-dependent microkinetic volcano models are shown in **Fig. 6a-c** for  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$ ,  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$ , and the corresponding elemental catalysts, respectively. The activity of M/HEA catalysts is clearly higher than that of elemental catalysts, as several points are located close to the peak of the volcano models for both  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$  and  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$ , while no such points appear for the single elemental catalysts. However, the number of constituent elements does not monotonically determine catalytic activity.



**FIG. 6. Catalytic activity analysis and the role of short-range order.** (a-c) pH-dependent microkinetic models for  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$  (a),  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  (b) and elementary substances catalysts at an applied potential of -1.0 V vs. RHE. (c). (d-g) Heat maps showing the contribution of short-range order on activity, quantified as  $\overline{\Delta \log(\text{TOF})}_{x<0} - \overline{\Delta \log(\text{TOF})}_{x>0}$ , where  $x \in \{\alpha_{ij}, \delta_j\}$ . (d) and (e) report the Local Density Deviation parameter  $\delta_j$  contributions for  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  and  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$ , respectively; (f) and (g) report the Warren-Cowley parameter  $\alpha_{ij}$  contributions for  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  and  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$ , respectively. (i) Structures with the optimal active adsorption sites for  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  and  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$  under different pH conditions.

Both  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$  and  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  exhibit adsorption sites near the volcano peak, suggesting that they each possess high-activity sites for  $\text{CO}_2\text{RR}$ -to-formate.

To further investigate the influence of SRO on catalytic performance, we define an activity descriptor as  $\Delta \log(\text{TOF})_i = |\log(\text{TOF})_i - \max[\log(\text{TOF})]|$ , where TOF is measured in  $s^{-1}$ .

A larger value of this descriptor indicates poorer activity of the adsorption site. Structures with different adsorption sites of  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  or  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$  under different pH conditions are divided into two groups based on each SRO descriptor  $x \in \{\alpha_{ij}, \delta_j\}$ . The average difference in activity descriptors,  $\overline{\Delta \log(TOF)}_{x<0} - \overline{\Delta \log(TOF)}_{x>0}$ , is calculated between the two groups as the score of the descriptor. The results for all descriptors were combined and reported as the score heatmaps in **Fig. 6d-g**, where **Fig. 6d-e** denote the effect of  $\delta_j$  for  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  and  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$ , respectively, and **Fig. 6f-g** denote the effect of  $\alpha_{ij}$  for  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  and  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$ , respectively. Negative values (dark blue) indicate that the  $x < 0$  group exhibits higher catalytic activity, while positive values (light yellow) indicate stronger activity for the  $x > 0$  group. Larger absolute values are shown by colors closer to the ends of the scale, indicating a stronger influence of the SRO descriptor on activity. **Fig. 6h** depicts the adsorption structures corresponding to the highest activity for the two catalyst compositions across three pH levels. As  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$  exhibits the same optimal active site at pH = 1 and 7, five representative adsorption configurations are presented in total.

The SRO–activity analysis presented in **Fig. 6d-g** reveals that the structure–performance relationships for the MEA/HEA catalysts are approximately consistent across different pH conditions. For the  $\text{La}_{23}\text{Sn}_{23}\text{Bi}_{22}$  system, enhanced catalytic activity correlates with a local adsorbate environment enriched in Bi and Sn and depleted in La. This is structurally accompanied by the aggregation of Sn–Bi and Bi–Bi pairs and the dispersion of La–La and La–Sn pairs. Similarly, for the  $\text{La}_{14}\text{Sn}_{14}\text{Bi}_{14}\text{Co}_{13}\text{Ni}_{13}$  catalyst, higher activity is linked to a local environment enriched in Sn but depleted in La and Bi. The corresponding favorable structural motifs include the aggregation of Co–Ni, Sn–Sn, and Bi–Ni pairs, along with the dispersion of Co–La, Sn–Ni, Sn–Bi, and Ni–La pairs. A comparison between the two systems indicates that the SRO–activity relationships share key similarities, as catalytic performance in both is enhanced by a local environment that is enriched in Sn and depleted in La. The activity-enhancing role of Sn in both catalyst systems is readily understood, as Sn-mediated promotion of CO<sub>2</sub>RR-to-formate has been extensively documented[70, 71]. However, Bi, although a common catalyst component, exhibits a pronounced promotional effect only in the MEA catalyst. In the HEA catalyst, its impact on activity is markedly weaker than that of Sn and can even be slightly detrimental. This phenomenon reflects the unique complexity of high-entropy alloys (e.g., the cocktail effect and complex site distributions), where the

catalytic behavior of individual elements is strongly modulated by their distinct local environments and interaction with other elements, thus warranting careful discrimination.[6, 9] In addition, La shows a modest suppressive effect on activity, which may partly explain the historically limited exploration of La in CO<sub>2</sub>RR-to-formate catalysts; nevertheless, its strong enhancement of selectivity leads to an overall positive contribution to catalytic performance.

### III. Discussion

In this work, we have introduced ARCADE, a hierarchical and fully integrated computational framework designed to bridge the gap between the realistic structural complexity of high-entropy alloys and the rational design of their catalytic performance. This paradigm is built upon four synergistic pillars: i) the updated SRO-constrained ApolloX engine for generating thermodynamically stable nanoparticles; ii) an autonomous, geometry-aware search algorithm for identifying comprehensive adsorption ensembles; iii) state-of-the-art pH-dependent microkinetic modeling for the precise evaluation of theoretical activity under realistic operating conditions; and iv) crucially, a quantitative interpretability methodology utilizing SRO–performance score heatmaps. This final component successfully decouples the intricate relationships between local SRO environments and multiple performance metrics within complex catalytic ensembles.

Applying this paradigm to the design of MEA and HEA nanoparticle catalysts for CO<sub>2</sub>RR-to-formate, we utilized these SRO-guided score heatmaps to quantitatively decode the distinct roles of constituent elements and atomic pairs. Our analysis reveals general design principles where different performance metrics are governed by distinct structural features. Catalyst stability is primarily governed by specific segregation behaviors driven by high-entropy effects: the enrichment of Bi at the surface and Co/Ni in the bulk. Regarding this, trends are consistent across both systems, where the aggregation of Bi–La pairs at the surface and Co–Co/Co–Ni pairs in the core align perfectly with the observed elemental enrichment. Thus, while LDD analysis provides a clear understanding of individual elemental contributions, W-C parameters further offer granular structural insights, such as Bi’s specific tendency to aggregate with La rather than itself.

Distinct from the consistent trends observed for stability, the influence of SRO on catalytic function (selectivity and activity) exhibits specific structural dependencies between

the two catalysts. Based on LDD analysis, La and Sn are identified as the primary drivers for selectivity and activity, respectively. Specifically, high selectivity correlates with La-enriched local environments (favored by Sn–La aggregation in MEA and Co–La/Ni–Ni pairs in HEA), while high activity is promoted by Sn-enriched but La-depleted environments (supported by Sn–Bi aggregation in MEA and Co–Ni/Sn–Sn aggregation in HEA). These results demonstrate that while Sn and La act as the primary functional drivers for CO<sub>2</sub>RR-to-formate, their performance is critically modulated by specific pairwise interactions defined by the surrounding matrix. Collectively, these findings underscore the necessity of our dual-descriptor approach: without LDD parameters, the individual contributions of each element would remain unclear; without W-C parameters, the detailed characteristics of pairwise interactions governing specific reaction pathways could not be revealed. This comprehensive interpretability enables the optimization of catalysts by systematically tuning both local elemental enrichment and atomic pairwise patterns.

Finally, this computational workflow is fully integrated with the *DigCat* platform, establishing a closed-loop ecosystem for catalyst design. By leveraging *DigCat* for initial composition screening and parent structure determination, followed by our automated generation-and-analysis cycle, we successfully derived quantitative relationships between structural features and multiple performance metrics. Crucially, the newly generated structures and insights are deposited back into the *DigCat* database. Consequently, subsequent optimization becomes more targeted: starting from known structural prototypes, one can generate models using SRO motifs associated with superior performance and proceed directly to analysis. Iterating this loop can markedly accelerate interpretable catalyst design.

Beyond the specific success demonstrated here, the ARCADE framework represents a significant leap in digital materials discovery. By treating structural complexity as a programmable design parameter rather than a hindrance, this protocol offers a generalized methodology applicable to diverse electrochemical processes operating under varying pH environments. This approach effectively closes the loop between atomic-level disorder and multiple catalytic performance metrics, providing a clear decoding of how local chemical environments dictate overall system behavior. The ARCADE framework thus provides a powerful and transferable paradigm for the discovery of complex catalyst materials for relatively complicated catalytic reactions, moving beyond black-box screening toward the rational, SRO-informed design of next-generation catalysts. By integrating these capabilities into the

*DigCat* platform, we empower the research community to navigate the vast compositional space of high-entropy materials with unprecedented efficiency and interpretability.

## IV. Methods

### A. Warren-Cowley parameter

A widely-used descriptor to capture the chemical short-range order of structure is the Warren-Cowley (W-C) parameter, which is defined as

$$\alpha_{ij} = 1 - \frac{P_{ij}}{c_j}$$

where i and j denote atomic species,  $P_{ij}$  is the conditional probability of encountering an atom of type j at a certain distance cutoff (or a given neighbor shell) from a central atom of type i, and  $c_j$  is the overall concentration of element j in the system.

When  $\alpha_{ij} = 0$ , the distribution of j around i follows the global composition, i.e., random mixing. Values  $\alpha_{ij} > 0$  indicate that j occurs less frequently near i than expected from randomness, i.e., j is depleted around i, i-j atom pairs tend to disperse; while  $\alpha_{ij} < 0$  signifies an overrepresentation of j near i, i.e., j is enriched around i, i-j atom pairs tend to aggregate. The deviation of  $\alpha_{ij}$  from zero quantitatively characterizes the aggregation or dispersion propensity of an i-j pair at separation r.

### B. Local Density Deviation parameter

The adsorption process on catalysts is strongly related to the local chemical environments near the adsorbates, especially the compositions of the local environments. To more directly and transparently capture short-range order in local elemental concentrations, we introduce, by analogy to the Warren-Cowley parameter, a new descriptor termed the Local Density Deviation (LDD) parameter. It is defined as

$$\delta_j = 1 - \frac{p_j}{c_j}$$

where j denotes the atomic species,  $p_j$  is the local concentration of element j within the chosen local region, and  $c_j$  is the overall concentration of element j in the system.

When  $\delta_j = 0$ , the local density of element j is equal to its overall concentration. Positive values ( $\delta_j > 0$ ) indicate that element j is locally depleted relative to its global concentration, whereas negative values ( $\delta_j < 0$ ) signify local enrichment of j.

Although the LDD parameter shows certain similarity to the W-C parameter in form, their meanings are obviously non-equivalent. The LDD parameters focus on the enrichment or depletion of elements, while W-C parameters show more about aggregation or dispersion relationship between atom pairs. So LDD offers practical advantages—direct interpretability, low dimensionality, and robustness, making it well suited for SRO analysis for MEA/HEA catalysts.

### C. Density functional theory (DFT) calculation method

The Vienna Ab initio Simulation Package (VASP)[72] was employed to perform density functional theory (DFT) calculations. The interaction between ions and electrons was described using the projector augmented-wave (PAW) method. Exchange-correlation effects were treated with the revised Perdew-Burke-Ernzerhof (RPBE) functional within the generalized gradient approximation (GGA)[73, 74], a choice informed by prior benchmarking comparisons between experimental data and theoretical predictions.[75, 76] The DFT-D3 method is employed to correct the van der Waals interactions of the system. A plane-wave kinetic energy cutoff of 500 eV was applied for the expansion of the Kohn–Sham wave functions. Electronic self-consistency was converged to  $10^{-5}$  eV, and structural relaxations were performed until the forces were below  $0.05 \text{ eV } \text{\AA}^{-1}$ . For Brillouin zone integration, a  $\Gamma$ -centered Monkhorst-Pack grid was employed, ensuring that the product of k-points and basis vector length ( $k \times a$ ) exceeded 20  $\text{\AA}$ . Additionally, a vacuum spacing of 15  $\text{\AA}$  was applied perpendicular to the surface to prevent interactions between periodic images.

## References

---

- [1] Z. W. Seh, J. Kibsgaard, C. F. Dickens, I. Chorkendorff, J. K. Nørskov, and T. F. Jaramillo, Combining theory and experiment in electrocatalysis: Insights into materials design, *Science*

- 355**, eaad4998 (2017).
- [2] Z.-J. Zhao, S. Liu, S. Zha, D. Cheng, F. Studt, G. Henkelman, and J. Gong, Theory-guided design of catalytic materials using scaling relationships and reactivity descriptors, *Nature Reviews Materials* **4**, 792 (2019).
- [3] B. Hammer and J. K. Norskov, Why gold is the noblest of all the metals, *Nature* **376**, 238 (1995).
- [4] X. Jia, T. Wang, D. Zhang, X. Wang, H. Liu, L. Zhang, and H. Li, Advancing electrocatalyst discovery through the lens of data science: State of the art and perspectives, *Journal of Catalysis* , 116162 (2025).
- [5] Y. Chen, X. Huang, Y. He, Q. Liu, J. Du, W. Yang, W. Wang, D. Zhang, X. Jia, H. Wang, *et al.*, Data-driven strategies for designing multicomponent molten catalysts to accelerate the industrialization of methane pyrolysis, *ACS Catalysis* **15**, 11003 (2025).
- [6] Y. Xin, S. Li, Y. Qian, W. Zhu, H. Yuan, P. Jiang, R. Guo, and L. Wang, High-entropy alloys as a platform for catalysis: progress, challenges, and opportunities, *Acs Catalysis* **10**, 11280 (2020).
- [7] H. Ma, Y. Jiao, W. Guo, X. Liu, Y. Li, and X. Wen, Machine learning predicts atomistic structures of multielement solid surfaces for heterogeneous catalysts in variable environments, *The Innovation* **5** (2024).
- [8] Y. Yao, Q. Dong, A. Brozena, J. Luo, J. Miao, M. Chi, C. Wang, I. G. Kevrekidis, Z. J. Ren, J. Greeley, *et al.*, High-entropy nanoparticles: Synthesis-structure-property relationships and data-driven discovery, *Science* **376**, eabn3103 (2022).
- [9] T. Löffler, A. Ludwig, J. Rossmeisl, and W. Schuhmann, What makes high-entropy alloys exceptional electrocatalysts?, *Angewandte Chemie International Edition* **60**, 26894 (2021).
- [10] P. I. Odetola, B. J. Babalola, A. E. Afolabi, U. S. Anamu, E. Olorundaisi, M. C. Umba, T. Phahlane, O. O. Ayodele, and P. A. Olubambi, Exploring high entropy alloys: a review on thermodynamic design and computational modeling strategies for advanced materials applications, *Heliyon* **10** (2024).
- [11] J. R. Boes, O. Mamun, K. Winther, and T. Bligaard, Graph theory approach to high-throughput surface adsorption structure generation, *The Journal of Physical Chemistry A* **123**, 2281 (2019).

- [12] T. A. Batchelor, J. K. Pedersen, S. H. Winther, I. E. Castelli, K. W. Jacobsen, and J. Rossmeisl, High-entropy alloys as a discovery platform for electrocatalysis, *Joule* **3**, 834 (2019).
- [13] P. Hohenberg and W. Kohn, Inhomogeneous electron gas, *Physical review* **136**, B864 (1964).
- [14] W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects, *Physical review* **140**, A1133 (1965).
- [15] G. D. Barmparis, Z. Lodziana, N. Lopez, and I. N. Remediakis, Nanoparticle shapes by using wulff constructions and first-principles calculations, *Beilstein journal of nanotechnology* **6**, 361 (2015).
- [16] E. Ringe, R. P. Van Duyne, and L. Marks, Wulff construction for alloy nanoparticles, *Nano letters* **11**, 3399 (2011).
- [17] Y. A. Alsunni and C. B. Musgrave, Effect of applied potential on metal surfaces: Surface energy, wulff shape and charge distribution, *Applied Surface Science* **610**, 155147 (2023).
- [18] A. O. Lyakhov, A. R. Oganov, H. T. Stokes, and Q. Zhu, New developments in evolutionary structure prediction algorithm uspex, *Computer Physics Communications* **184**, 1172 (2013).
- [19] H. Wang, Y. Wang, J. Lv, Q. Li, L. Zhang, and Y. Ma, Calypso structure prediction method and its wide application, *Computational Materials Science* **112**, 406 (2016).
- [20] S. Moniri, Y. Yang, J. Ding, Y. Yuan, J. Zhou, L. Yang, F. Zhu, Y. Liao, Y. Yao, L. Hu, *et al.*, Three-dimensional atomic structure and local chemical order of medium-and high-entropy nanoalloys, *Nature* **624**, 564 (2023).
- [21] C. Zeni, K. Rossi, T. Pavloudis, J. Kioseoglou, S. de Gironcoli, R. E. Palmer, and F. Baletto, Data-driven simulation and characterisation of gold nanoparticle melting, *Nature Communications* **12**, 6056 (2021).
- [22] S.-P. Ju, C.-W. Chen, H.-L. Chen, H.-T. Chen, and H.-Y. Chen, Tailoring mechanical performance in bulk nanoparticle-structured zno and alo: Insights from deep learning potential molecular dynamics simulations, *Materials Today Communications* **42**, 111161 (2025).
- [23] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, Python materials genomics ( pymatgen): A robust, open-source python library for materials analysis, *Computational Materials Science* **68**, 314 (2013).
- [24] M. K. Bisbo and B. Hammer, Efficient global structure optimization with a machine-learned surrogate model, *Physical review letters* **124**, 086102 (2020).

- [25] C. Shang and Z.-P. Liu, Stochastic surface walking method for structure prediction and pathway searching, *Journal of Chemical Theory and Computation* **9**, 1838 (2013).
- [26] J. Burkhardt, Y. Jia, and W.-L. Li, Structure search with the strategic escape algorithm, *Journal of Chemical Theory and Computation* **21**, 3765 (2025).
- [27] K. Tran and Z. W. Ulissi, Active learning across intermetallics to guide discovery of electrocatalysts for co<sub>2</sub> reduction and h<sub>2</sub> evolution, *Nature Catalysis* **1**, 696 (2018).
- [28] Q. Wang and Y. Yao, Harnessing machine learning for high-entropy alloy catalysis: a focus on adsorption energy prediction, *npj Computational Materials* **11**, 91 (2025).
- [29] J. K. Nørskov, J. Rossmeisl, A. Logadottir, L. Lindqvist, J. R. Kitchin, T. Bligaard, and H. Jonsson, Origin of the overpotential for oxygen reduction at a fuel-cell cathode, *The Journal of Physical Chemistry B* **108**, 17886 (2004).
- [30] J. K. Nørskov, F. Abild-Pedersen, F. Studt, and T. Bligaard, Density functional theory in surface chemistry and catalysis, *Proceedings of the National Academy of Sciences* **108**, 937 (2011).
- [31] S. R. Kelly, C. Kirk, K. Chan, and J. K. Nørskov, Electric field effects in oxygen reduction kinetics: rationalizing ph dependence at the pt (111), au (111), and au (100) electrodes, *The Journal of Physical Chemistry C* **124**, 14581 (2020).
- [32] H. Liu, D. Zhang, Y. Wang, and H. Li, Reversible hydrogen electrode (rhe) scale dependent surface pourbaix diagram at different ph, *Langmuir* **40**, 7632 (2024).
- [33] D. Zhang, Z. Wang, F. Liu, P. Yi, L. Peng, Y. Chen, L. Wei, and H. Li, Unraveling the ph-dependent oxygen reduction performance on single-atom catalysts: from single-to dual-sabatier optima, *Journal of the American Chemical Society* **146**, 3210 (2024).
- [34] S. Ye, Y. Wang, H. Liu, D. Zhang, X. Jia, L. Zhang, Y. Zhang, A. Kumatori, H. Shiku, and H. Li, Decoding ph-dependent electrocatalysis through electric field models and microkinetic volcanoes, *Journal of Materials Chemistry A* **13**, 37821 (2025).
- [35] Y. Wang, D. Zhang, B. Sun, X. Jia, L. Zhang, H. Cheng, J. Fan, and H. Li, Divergent activity shifts of tin-based catalysts for electrochemical co<sub>2</sub> reduction: ph-dependent behavior of single-atom versus polyatomic structures, *Angewandte Chemie International Edition* **64**, e202418228 (2025).
- [36] G. Chen, R. Lu, C. Ma, X. Zhang, Z. Wang, Y. Xiong, and Y. Han, A long-range disordered ruo<sub>2</sub> catalyst for highly efficient acidic oxygen evolution electrocatalysis, *Angewandte Chemie*

- International Edition **63**, e202411603 (2024).
- [37] Y. Yang, Z. Jia, X. Zhang, Y. Liu, Q. Wang, Y. Li, L. Shao, S. Di, J. Kuang, L. Sun, *et al.*, Chemical short-range order in multi-principal element alloy with ordering effects on water electrolysis performance, *Materials Today* **72**, 97 (2024).
  - [38] X. Wang, Z. Li, H. Jang, C. Chen, S. Liu, L. Wang, M. G. Kim, J. Cho, Q. Qin, and X. Liu, RuO<sub>2</sub> with short-range ordered tantalum single atoms for enhanced acidic oxygen evolution reaction, *Advanced Energy Materials* **15**, 2403388 (2025).
  - [39] D. Zhang and H. Li, Digital catalysis platform (digcat): a gateway to big data and ai-powered innovations in catalysis, (2024).
  - [40] M. Grasemann and G. Laurenczy, Formic acid as a hydrogen source—recent developments and future trends, *Energy & Environmental Science* **5**, 8171 (2012).
  - [41] J. Eppinger and K.-W. Huang, Formic acid as a hydrogen energy carrier, *ACS Energy Letters* **2**, 188 (2017).
  - [42] N. Metropolis and S. Ulam, The monte carlo method, *Journal of the American statistical association* **44**, 335 (1949).
  - [43] H. Li, C. Liu, Y. Guo, X. Luo, Y. Chen, G. Liu, Y. Li, R. Wang, Z. Wang, J. Wu, *et al.*, Conditional generative modeling for amorphous multi-element materials, arXiv preprint arXiv:2503.07043 (2025).
  - [44] J. Kennedy and R. Eberhart, Particle swarm optimization, in *Proceedings of ICNN'95-international conference on neural networks*, Vol. 4 (ieee, 1995) pp. 1942–1948.
  - [45] F. Aurenhammer, Voronoi diagrams—a survey of a fundamental geometric data structure, *ACM computing surveys (CSUR)* **23**, 345 (1991).
  - [46] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, Spatial tessellations: concepts and applications of voronoi diagrams, (2009).
  - [47] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, *et al.*, The atomic simulation environment—a python library for working with atoms, *Journal of Physics: Condensed Matter* **29**, 273002 (2017).
  - [48] J. M. Cowley, An approximate theory of order in alloys, *Physical Review* **77**, 669 (1950).
  - [49] J. Cowley, Short-and long-range order parameters in disordered solid solutions, *Physical Review* **120**, 1648 (1960).

- [50] J. Cowley, Short-range order and long-range order parameters, *Physical Review* **138**, A1384 (1965).
- [51] Y. Du, X. Meng, J. Qi, B. Gong, Y. Ma, J. Li, H. Zou, and J. Qiu, Engineering efficient self-supporting sno<sub>2</sub>/carbon nanofiber electrode for electrochemical co<sub>2</sub> reduction, *Chemical Engineering Science* , 122147 (2025).
- [52] K. Shi, D. Le, T. Panagiotakopoulos, T. S. Rahman, and X. Feng, Effect of ammonium-based cations on co<sub>2</sub> electroreduction, *ACS Catalysis* **15**, 3647 (2025).
- [53] J. Zhu, G. Zhou, Y. Tong, L. Chen, and P. Chen, Vanadium oxide clusters mediated bismuth-tin alloy for accelerated dynamics of electrocatalytic co<sub>2</sub> conversion, *Advanced Functional Materials* **35**, 2420177 (2025).
- [54] X. Wang, Y. Zhang, S. Wang, Y. Li, Y. Feng, Z. Dai, Y. Chen, X. Meng, J. Xia, and G. Zhang, Steering geometric reconstruction of bismuth with accelerated dynamics for co<sub>2</sub> electroreduction, *Angewandte Chemie* **136**, e202407665 (2024).
- [55] Y. Wang, G. Li, J. Feng, X. Wang, Q. Xue, A. Cheng, B. Liu, S. Zeng, M. Wang, and X. Zhang, Boosting electrochemical co<sub>2</sub> reduction to formate over la-doped sno<sub>2</sub> via pinning effect and water activation, *Journal of the American Chemical Society* (2025).
- [56] B. Cantor, I. T. Chang, P. Knight, and A. Vincent, Microstructural development in equiatomic multicomponent alloys, *Materials Science and Engineering: A* **375**, 213 (2004).
- [57] J.-W. Yeh, S.-K. Chen, S.-J. Lin, J.-Y. Gan, T.-S. Chin, T.-T. Shun, C.-H. Tsau, and S.-Y. Chang, Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes, *Advanced engineering materials* **6**, 299 (2004).
- [58] G. R. Dey, C. R. McCormick, S. S. Soliman, A. J. Darling, and R. E. Schaak, Chemical insights into the formation of colloidal high entropy alloy nanoparticles, *ACS nano* **17**, 5943 (2023).
- [59] B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel, and G. Ceder, Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling, *Nature Machine Intelligence* **5**, 1031 (2023).
- [60] F. Li, G. H. Gu, C. Choi, P. Kolla, S. Hong, T.-S. Wu, Y.-L. Soo, J. Masa, S. Mukerjee, Y. Jung, *et al.*, Highly stable two-dimensional bismuth metal-organic frameworks for efficient electrochemical reduction of co<sub>2</sub>, *Applied Catalysis B: Environmental* **277**, 119241 (2020).

- [61] N. Han, Y. Wang, H. Yang, J. Deng, J. Wu, Y. Li, and Y. Li, Ultrathin bismuth nanosheets from in situ topotactic transformation for selective electrocatalytic co<sub>2</sub> reduction to formate, *Nature communications* **9**, 1320 (2018).
- [62] N. Sabouhian, J. Lipkowski, and A. Chen, Unveiling the potential of bismuth-based catalysts for electrochemical co<sub>2</sub> reduction, *Industrial Chemistry & Materials* (2025).
- [63] Z. Yao, Z. Cheng, and J. Wang, Heteroatom-doped bismuth-based electrocatalysts for co<sub>2</sub> reduction to formic acid: Advancement and perspective, *cMat* **2**, e70016 (2025).
- [64] Y. Luo, J. Chen, N. Han, and Y. Li, Bismuth-catalyzed electrochemical carbon dioxide reduction to formic acid: Material innovation and reactor design, *Accounts of Materials Research* **6**, 462 (2025).
- [65] J. T. Feaster, C. Shi, E. R. Cave, T. Hatsukade, D. N. Abram, K. P. Kuhl, C. Hahn, J. K. Nørskov, and T. F. Jaramillo, Understanding selectivity for the electrochemical reduction of carbon dioxide to formic acid and carbon monoxide on metal electrodes, *Acs Catalysis* **7**, 4822 (2017).
- [66] R. Kortlever, J. Shen, K. J. P. Schouten, F. Calle-Vallejo, and M. T. Koper, Catalysts and reaction pathways for the electrochemical reduction of carbon dioxide, *The journal of physical chemistry letters* **6**, 4073 (2015).
- [67] Y. Wang, Z. Wu, Y. Jiang, D. Zhang, Q. Wang, C. Wang, H. Li, X. Jia, J. Fan, and H. Li, Bridging theory and experiment: Machine learning potential-driven insights into ph-dependent co reduction on sn-based catalysts, *Advanced Functional Materials* **35**, e06314 (2025).
- [68] Y. Chu, Y. Wang, D. Zhang, X. Song, C. Yu, and H. Li, Data-driven discovery of single-atom catalysts for co<sub>2</sub> reduction considering the ph-dependency at the reversible hydrogen electrode scale, *The Journal of Chemical Physics* **162** (2025).
- [69] D. Zhang and H. Li, The potential of zero charge and solvation effects on single-atom m–n–c catalysts for oxygen electrocatalysis, *Journal of Materials Chemistry A* **12**, 13742 (2024).
- [70] Y. Hori, Electrochemical co<sub>2</sub> reduction on metal electrodes, *Modern aspects of electrochemistry* , 89 (2008).
- [71] J. T. Feaster, C. Shi, E. R. Cave, T. Hatsukade, D. N. Abram, K. P. Kuhl, C. Hahn, J. K. Nørskov, and T. F. Jaramillo, Understanding selectivity for the electrochemical reduction of carbon dioxide to formic acid and carbon monoxide on metal electrodes, *Acs Catalysis* **7**, 4822 (2017).

- [72] G. Kresse and J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, Computational materials science **6**, 15 (1996).
- [73] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, Physical review letters **77**, 3865 (1996).
- [74] B. Hammer, L. B. Hansen, and J. K. Nørskov, Improved adsorption energetics within density-functional theory using revised perdew-burke-ernzerhof functionals, Physical review B **59**, 7413 (1999).
- [75] R. B. Araujo, G. L. Rodrigues, E. C. Dos Santos, and L. G. Pettersson, Adsorption energies on transition metal surfaces: towards an accurate and balanced description, Nature Communications **13**, 6853 (2022).
- [76] J. Wellendorff, T. L. Silbaugh, D. Garcia-Pintos, J. K. Nørskov, T. Bligaard, F. Studt, and C. T. Campbell, A benchmark database for adsorption bond energies to transition metal surfaces and comparison to selected dft functionals, Surface Science **640**, 36 (2015).

## Acknowledgements

The calculations were partly performed using the San Diego Supercomputer Center (SDSC) Expanse at UC San Diego through allocation MAT240028 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program. Additionally, H. Li acknowledges the Center for Computational Materials Science, Institute for Materials Research, Tohoku University for the use of MASAMUNE-IMR (No. 202412-SCKXX0211) and the Institute for Solid State Physics (ISSP) at the University of Tokyo for the use of their supercomputers. W.-L. Li gratefully acknowledges support from the ACS Petroleum Research Fund under Doctoral New Investigator Grant 69037-DNI10. W.-L. Li and Y. Ding acknowledge the Jacobs School of Engineering Early Career Faculty Development Award, University of California San Diego. W.-L. Li also acknowledges support of the Hellman Fellowship.

## Data availability

Adsorption energy data, as well as data used for selectivity and activity analysis, are available in the paper. All structural data have been uploaded to *DigCat* (<http://www.DigCat.org>).

`digcat.org`). Additional data supporting this project are available from the corresponding author upon reasonable request.

### **Code availability**

The entire workflow can be implemented through *DigCat* (<http://www.digcat.org>). The code for ApolloX is also available on GitHub (<https://github.com/FNC001/ApolloX>).