Research Article | Open Access

Journal of Materials Informatics

Tang *et al. J. Mater. Inf.* **2026**, *6*, 47 **DOI:**10.20517/jmi.2026.63

Reaction-anchored generative informatics for million-scale CO_2 catalyst screening via multi-adsorbate constraints

Dongjin Tang[1,2,#], Ruili Li[3,4,#], Zhidi Lei[5], Qingqing Mao[6], Ritankar Das[6], Beien Zhu[1,7], Yi Gao[1,7], Rui Qi[1,7,*]

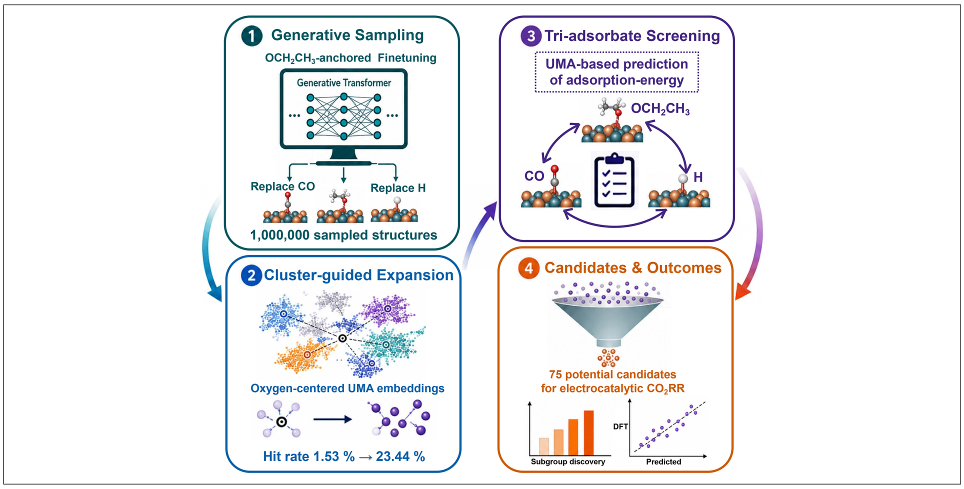
Keywords:
Materials informatics, generative Transformer, million-scale screening, multi-adsorbate constraints, CO_2 reduction reaction

Citation: Tang, D.; Li, R.; Lei, Z.; Mao, Q.; Das, R.; Zhu, B.; Gao, Y.; Qi, R. Reaction-anchored generative informatics for million-scale CO_2 catalyst screening via multi-adsorbate constraints. *J. Mater. Inf.* **2026**, *6*, 47. https://dx.doi.org/10.20517/jmi.2026.63

Received: 28 Jul 2026
First Decision: 18 Aug 2026
Revised: 5 Sep 2026
Accepted: 21 Sep 2026
Published: 29 Sep 2026

Academic Editor: Yaqiong Su
Copy Editor: Pei-Yun Wang
Production Editor: Pei-Yun Wang

Check for updates



Abstract

The rational design of electrocatalysts for multi-electron reaction networks hinges on navigating vast, multidimensional surface-environment spaces defined by coupled variations in composition, coordination topology, and adsorbate configurations. Taking electrochemical carbon dioxide reduction to ethanol as a model challenge, traditional screening methodologies struggle to efficiently represent and search these combinatorial landscapes. Here, we present a reaction-constrained generative materials informatics framework anchored by late-stage oxygenated intermediates ($^*OCH_2CH_3$). A fine-tuned generative Transformer model efficiently sampled 1,000,000 adsorption configurations. To efficiently navigate this massive space, adsorption configurations were compressed into latent topological representations via oxygen-centered local node embeddings derived from the universal models for atoms (UMA) platform. Cluster-guided spatial navigation

[1]Photon Science Research Center for Carbon Dioxide, Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China.
[2]School of Materials Science and Engineering, Xiangtan University, Xiangtan 411105, Hunan, China.
[3]Key Laboratory of Interfacial Physics and Technology, Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201800, China.
[4]University of Chinese Academy of Sciences, Beijing 100049, China.
[5]Shanghai Synchrotron Radiation Facility, Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201204, China.
[6]Titan Holdings, Hayward, CA 94541, USA.
[7]State Key Laboratory of Low Carbon Catalysis and Carbon Dioxide Utilization, Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China.
[#]These authors contributed equally to this work.

[*]**Correspondence to:** Dr. Rui Qi, Photon Science Research Center for Carbon Dioxide, Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China. E-mail: qir@sari.ac.cn

based on these UMA-derived representations achieved a 15-fold enrichment in positive-candidate discovery over naive representative sampling. By jointly applying multi-adsorbate thermodynamic constraints ($*OCH_2CH_3$, *CO, and *H) within the CO_2RR network, the workflow down-selected 75 promising material candidates. Density functional theory (DFT) calculations benchmarked representative systems, validating the screening-level fidelity of UMA-predicted binding energies. Furthermore, subgroup discovery extracted interpretable, physically intuitive elemental rules governing candidate enrichment. CO-CO coupling calculations further identified Ga-Pt and Al-Cu-Pd as kinetically favored over Cu. This framework successfully recovers 13 reported CO_2RR-related materials while uncovering 62 unexplored compositional domains, providing a transferable paradigm for reaction-constrained generative surface informatics.

INTRODUCTION

Electrochemical CO_2 reduction reaction (CO_2RR) powered by renewable electricity offers a sustainable pathway to convert greenhouse gases into valuable fuels and chemicals, thereby linking carbon recycling with green-energy storage[1-4]. Among various reduction products, ethanol is exceptionally promising due to its high energy density as a liquid fuel, ease of transport, and seamless compatibility with existing fueling infrastructure[1,5]. However, steering CO_2RR toward multi-carbon products like ethanol remains a formidable challenge because product selectivity is governed by multiple coupled elementary steps, including CO formation, C-C coupling, oxygenated C_2 intermediate stabilization, protonation, and competition with hydrogen evolution[6-10]. Consequently, catalyst discovery for ethanol synthesis cannot rely on optimizing a single adsorption descriptor. Instead, it demands the precise engineering of local surface environments that can simultaneously retain oxygenated C_2 intermediates, maintain local CO availability, and suppress the competing hydrogen evolution reaction (HER).

High-throughput computation and machine learning have greatly accelerated catalytic materials discovery by enabling rapid evaluations of adsorption energies, reaction descriptors, and composition-property relationships across broad candidate spaces[6,11-29]. Recent multiparameter machine-learning studies have further linked catalytic performance to coupled electronic and geometric descriptors[30]. However, conventional workflows predominantly explore catalyst spaces using predefined material libraries, idealized surface prototypes, or simple substitutional alloy models. These rigid baselines inherently limit the capacity to sample highly diverse and unconventional local surface environments. This constraint is particularly problematic for multicomponent catalysts, where adsorption behavior is governed by coupled variations in local composition, coordination topology, and adsorbate configuration. For ethanol-oriented CO_2RR, such local-environment sensitivity is critical because late-stage oxygenated C_2 intermediates are exceptionally site-dependent[7,31-33]. Expanding the searchable surface-structure space while retaining reaction-relevant constraints thus represents a major frontier in computational catalyst discovery.

Generative models offer a transformative route to transcend predefined structural templates by directly sampling uncharted material and surface configurations[34,35]. Pioneering architectures including variational autoencoders (VAEs), generative adversarial networks (GANs), and diffusion models have demonstrated a profound ability to sample periodic structures directly from latent spaces[36-38]. Furthermore, Transformer-based large language models have been increasingly adopted in chemistry for reasoning, prediction, and autonomous workflows[39-41], exemplified by CrystaLLM for direct crystal-structure generation from text[42]. In heterogeneous catalysis, frameworks such as CatGPT and MAGECS have connected structure generation with surface and property optimization for catalytic applications[43,44]. Our recent distributed generative-Transformer framework further showed that adsorption-structure generation can be scaled to tens of millions of candidates using a single reaction descriptor[45]. These advances indicate that large-scale structure generation is becoming technically feasible. However, for complex electrocatalytic reactions, the key challenge has shifted: it is no longer merely about how many structures can be generated, but how such a vast surface space can be compressed into meaningful representations, efficiently navigated under reaction-relevant constraints, and translated into interpretable physical knowledge.

Herein, we report a reaction-constrained generative materials informatics workflow tailored for ethanol-pathway-oriented CO_2RR catalyst screening. Rather than enumerating predefined adsorption sites, the workflow explicitly employs ethoxy (*OCH_2CH_3) as a late-stage oxygenated C_2 anchor. Its retained C-O functionality and strong sensitivity to the local surface environment make it well suited for guiding generative sampling toward ethanol-relevant adsorption environments[32,33]. A fine-tuned generative Transformer model samples 1,000,000 adsorption configurations, establishing a million-scale reaction-anchored surface pool. To effectively navigate and compress this massive topological space, oxygen-centered local node embeddings are extracted from a pretrained universal models for atom (UMA) platform. Crucially, cluster-guided navigation based on UMA-derived representations achieves a 15-fold enrichment in positive-candidate discovery over naive representative sampling. Thermodynamic filtration is then enforced via a tri-adsorbate gate. By jointly applying the adsorption-energy windows for *OCH_2CH_3, *CO, and *H, the workflow successfully isolates candidate surfaces that balance C_2 retention and CO supply while inhibiting hydrogen evolution. Representative density functional theory (DFT) calculations support the screening-level reliability of UMA-predicted binding energies, while subgroup discovery (SGD) extracts interpretable, physically intuitive elemental rules from the positive ensemble. Additional CO-CO coupling calculations provide an initial kinetic assessment of selected candidates. Overall, the screening yielded 75 material-level candidates: 13 correspond to experimentally reported CO_2RR-related systems, while the remaining 62 represent less-explored multicomponent candidate spaces. This workflow successfully converts a million-scale generative space into reaction-constrained catalyst regions, providing a highly transferable materials informatics strategy for unexplored surface discovery.

MATERIALS AND METHODS

Transformer-based distributed generative framework

To enable parallelized structure generation, we developed a distributed, Transformer-based framework built upon the core architecture of GPT-2[43]. The underlying generative engine employs a decoder-only configuration featuring 12 stacked Transformer blocks, 8 attention heads, and a 512-dimensional embedding space, trained with a batch size of 144. Each block contains causal masked multi-head self-attention, layer normalization, and a position-wise feed-forward network, and the resulting hidden representations are used for autoregressive prediction of the structural token sequence. Surface environments involving lattice parameters, elemental identities, and atomic positions were mapped into discrete tokens via coordinate-level tokenization. Broad structural priors were initially established by pretraining the framework on 2 million entries from the OC20-S2EF dataset, with architectural and tokenization consistency maintained throughout validation against the OC20-S2EF Val-ID split[46]. Tailored exploration of the reaction space was achieved by fine-tuning the pretrained model on 14,000 *OCH_2CH_3 adsorption structures, split into 90% training and 10% validation subsets, for 14 epochs. This specialized adaptation allowed the network to transition from general geometric representations to focused sampling of intermediate-specific potential energy surfaces. Autoregressive sampling was performed at a temperature of 1.0 on a single NVIDIA H100 GPU. A random seed of 42 was used for both fine-tuning and sampling. This highly optimized execution pipeline generated a target pool of 1 million candidate structures, serving as the foundational dataset for downstream deduplication, clustering, and surrogate-property evaluation.

UMA encoding and optimization

Architectural embedding extraction

Structural characterization and energy evaluations were accelerated using the pretrained uma-s-1p2 checkpoint within the UMA framework[47]. For each screened surface environment, forward inference was performed under the standard OC20 configuration to extract the normalized, zero-angular-momentum (l=0) scalar node embeddings from the network backbone. These representations were decoupled into two analytical tiers, where the scalar node representations of all non-adsorbate atoms were pooled into a unified 128-dimensional vector to provide a global substrate-level descriptor for characterizing and visualizing the overall generated surface space. Concurrently, the scalar node representation of the specific oxygen atom anchoring the ethoxy intermediate was isolated as a local adsorption environment descriptor for spherical k-means clustering, representative sampling, local-space visualization, and neighborhood expansion. Further details regarding the UMA architecture and implementation are provided in the Supplementary Materials.

Geometry optimization

Structural optimizations were initiated from these representative $*OCH_2CH_3$ configurations using the UMA architecture. To compute competitive binding energies for the accompanying intermediates, the ethoxy group was systematically removed upon geometric convergence of the $*OCH_2CH_3$ surface, and the original oxygen position was used only to initialize *CO and *H adsorption. No adsorbate coordinates were constrained during the subsequent UMA relaxation, while the bottommost slab atoms remained fixed. Structural minimization was driven by a batched L-BFGS algorithm with a maximum allocation of 500 ionic steps, proceeding until the maximum residual force on all unconstrained atoms fell below 0.05 $eV \cdot Å^{-1}$.

The adsorption energies for the three intermediate species were calculated using:

$$\Delta E_{\text{ads}} = E_{\text{sys}} - E_{\text{slab}} - E_{\text{ref}} \tag{1}$$

where E_{sys}, E_{slab}, E_{ref} represent the total energy of the adsorption system, the bare surface energy, and the adsorbate reference energy, respectively.

DFT calculations

Electronic structure evaluations of the screened structural motifs were conducted via spin-polarized DFT simulations implemented within the Vienna Ab initio Simulation Package (VASP) package[48,49]. Exchange-correlation interactions were described using the revised Perdew-Burke-Ernzerhof (RPBE) functional[50], and core-valence interactions were treated using the projector augmented-wave (PAW) method[51]. A plane-wave kinetic-energy cutoff of 350 eV was employed, and metallic occupations were treated using first-order Methfessel-Paxton smearing with a width of 0.2 eV. Slab atoms located within 2.0 Å below the uppermost slab atom along the third lattice-vector direction were allowed to relax, whereas deeper slab atoms were fixed; all adsorbate atoms were allowed to relax. Reciprocal-space integration over the Brillouin zone was performed using Γ-centered $k_1 \times k_2 \times 1$ meshes, where $k_i = \max[1, \text{round}(40/|a_i|)]$ and $|a_i|$ is the length of the corresponding in-plane lattice vector in Å. Geometry optimizations were considered converged when the residual force on each relaxed atom was below 0.05 $eVÅ^{-1}$ and the electronic self-consistency threshold was set to 1×10^{-4} eV. The same DFT settings were applied consistently to all 20 representative benchmark systems. The adsorption energies were evaluated according to Equation (1), where the adsorbate reference energies were derived via linear combinations of the standard OC20 atomic reference energies[46]. Each production DFT relaxation was performed on CPUs using 12 MPI processes. Representative CPU/GPU wall-clock benchmarks for UMA and VASP calculations are provided in Supplementary Table 1.

SGD

SGD was employed to identify interpretable elemental descriptor conditions correlated with target labels within the structural candidate pools[52]. The search for subgroup rules was executed independently across the binary and ternary compositional spaces, using a feature space constructed from intrinsic properties of the constituent elements [Supplementary Table 2]: atomic radius (R), electron affinity (EA), ionization potential (IP), cohesive energy (EC), the number of d-valence electrons (Nd), the number of group valence electrons (Nv), electronegativity (EN), Mendeleev number (MN), and relative atomic mass (M). Implemented via the beam search algorithm within the pysubgroup framework[53], the search targeted the positive structural class as a binary objective variable with the quality function set to StandardQF (a = 0.5) and a maximum rule depth of 5. The resulting rules were ranked by their quality function values, and statistical metrics including total sample coverage, positive class coverage, subgroup positive ratio, baseline positive ratio, and enrichment factor were systematically recorded to evaluate the enrichment capabilities of varying elemental descriptor combinations for the positive structures.

Next, we introduced a post hoc Shapley-based condition-attribution method to quantify the contribution of individual conditions within each discovered rule[54]. For each subset S of the conditions, the StandardQF score of the subgroup defined by S was used as the characteristic-function value $v(S)$. For a rule containing n conditions, all 2^n possible condition subsets were enumerated, and the Shapley value of each condition was calculated as the weighted average of its marginal contribution to the StandardQF score over all subsets not containing that condition. Additional methodological details for SGD rule search and Shapley attribution can be found in the Supplementary Materials.

RESULTS AND DISCUSSION

Reaction-anchored generative materials informatics framework

To bypass the combinatorial limitation of brute-force site enumeration on multicomponent surfaces, we established a reaction-anchored generative materials informatics framework organized into four integrated stages [Figure 1]. In the first stage, the search space is chemically oriented by defining ethoxy ($*OCH_2CH_3$) as a late-stage oxygenated C_2 reaction anchor for CO_2RR-to-ethanol pathways. Rather than serving as a complete descriptor of ethanol selectivity, $*OCH_2CH_3$ is employed to bias the generated structures toward local surface environments intrinsically capable of stabilizing oxygenated C_2 intermediates. These surface environments are subsequently evaluated following adsorbate replacement with *CO and *H, establishing a tri-adsorbate constraint framework that simultaneously considers oxygenated C_2 binding, CO binding, and H adsorption as screening-level descriptors.

In the second stage, this reaction anchor is mapped into a generative structure space via a Transformer-based generative model. The model was pretrained on the OC20 dataset and fine-tuned on $*OCH_2CH_3$ adsorption configurations, enabling the direct generation of 1,000,000 candidate adsorption structures. In the third stage, oxygen-centered local node embeddings extracted from a pretrained UMA model are used to represent the local microenvironment around the ethoxy O atom. These embeddings are grouped via spherical k-means clustering based on cosine similarity to guide representative sampling and visualized within a Uniform Manifold Approximation and Projection (UMAP) latent space[55] to identify high-hit clusters under the joint $*OCH_2CH_3$/*CO/*H adsorption-energy windows.

High-hit clusters identified from the local environment space were subjected to spatial neighborhood expansion and cascade thermodynamic filtering, yielding 1,122 structure-level positive candidates, which were further reduced to 1,098 after elemental safety filtering and consolidated into 75 distinct material-level systems. These 75 systems should therefore be regarded as candidates satisfying the imposed late-stage adsorption constraints, while the energetic accessibility of preceding elementary steps requires further

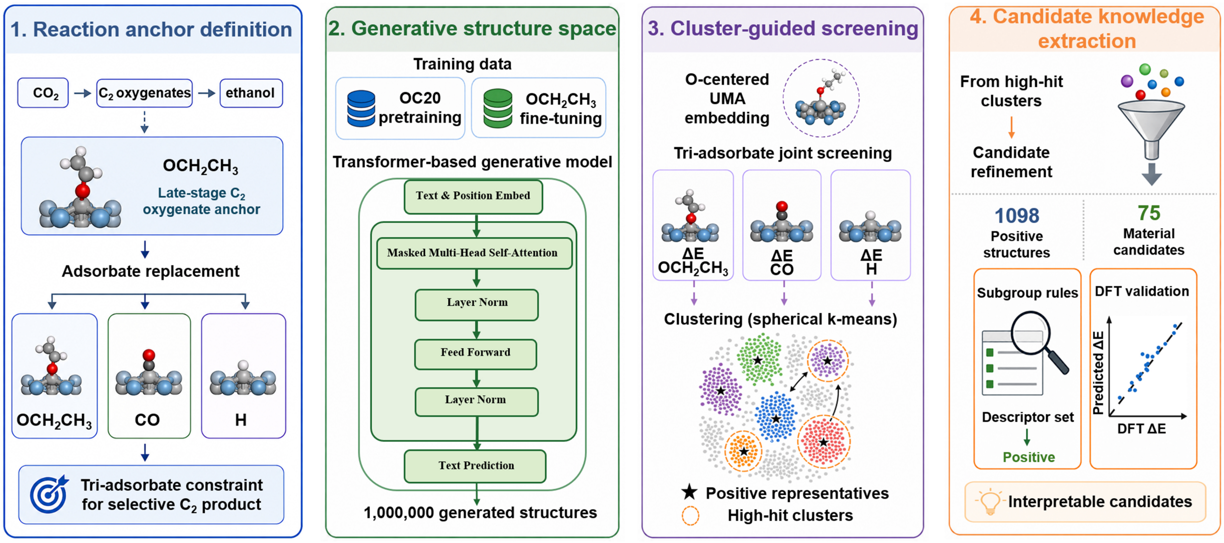


Figure 1. Reaction-anchored generative materials informatics workflow for CO_2RR ethanol-pathway-oriented surface discovery. CO_2RR: CO_2 reduction reaction; UMA: universal models for atom; DFT: density functional theory.

pathway-level evaluation. Representative DFT calculations for $^*OCH_2CH_3$, *CO, and *H adsorption benchmarked the screening-level reliability of UMA-predicted binding energies, while SGD identified interpretable elemental rules governing positive-candidate enrichment. Overall, this workflow effectively translates a million-scale generative surface space into a reaction-constrained, physics-informed candidate pool for ethanol-oriented CO_2RR catalyst discovery.

Cluster-guided expansion of the reaction-anchored surface space

To construct a reaction-anchored surface library for ethanol-oriented CO_2RR discovery, the fine-tuned Transformer model was deployed to generate $^*OCH_2CH_3$ adsorption configurations at the million scale. Among 1,000,000 generated configurations, 999,996 were successfully decoded into atomic structures. After filtering out unphysical geometries with interatomic distances shorter than 0.7 Å or incorrect adsorbate stoichiometry, 839,237 valid $^*OCH_2CH_3$ adsorption systems were retained for downstream analysis [Figure 2A]. The resulting structural library spans 55 elements across the main-group and transition-metal regions, indicating broad compositional coverage of the generated surface space [Figure 2B]. Compared with the elemental distribution of the fine-tuning dataset [Supplementary Figure 1], the generated set largely preserves the learned chemical domain while exhibiting moderate shifts in relative elemental abundance. UMA-derived substrate embeddings further show that the generated structures substantially overlap with the fine-tuning domain while forming a denser, more continuous coverage within the UMAP space [Figure 2C]. This confirms that the generator preserves the learned adsorption-structure domain while enhancing both sampling density and structural diversity.

To organize this million-scale structural space into locally coherent adsorption environments, the oxygen-centered UMA embeddings were partitioned into 2,000 clusters. The resulting clusters generally exhibited high intra-cluster cosine similarity across varying cluster sizes [Supplementary Figure 2]. Five representative structures were then selected from each cluster according to the 0th, 25th, 50th, 75th, and 95th percentiles of cosine similarity, yielding 10,000 representative candidates for downstream UMA evaluations. For each representative $^*OCH_2CH_3$ structure, the corresponding *CO and *H adsorption configurations were constructed at the identical local site and relaxed. Pairwise Spearman correlation analysis revealed that the three adsorption descriptors are only partially coupled, supporting the necessity of joint multi-adsorbate

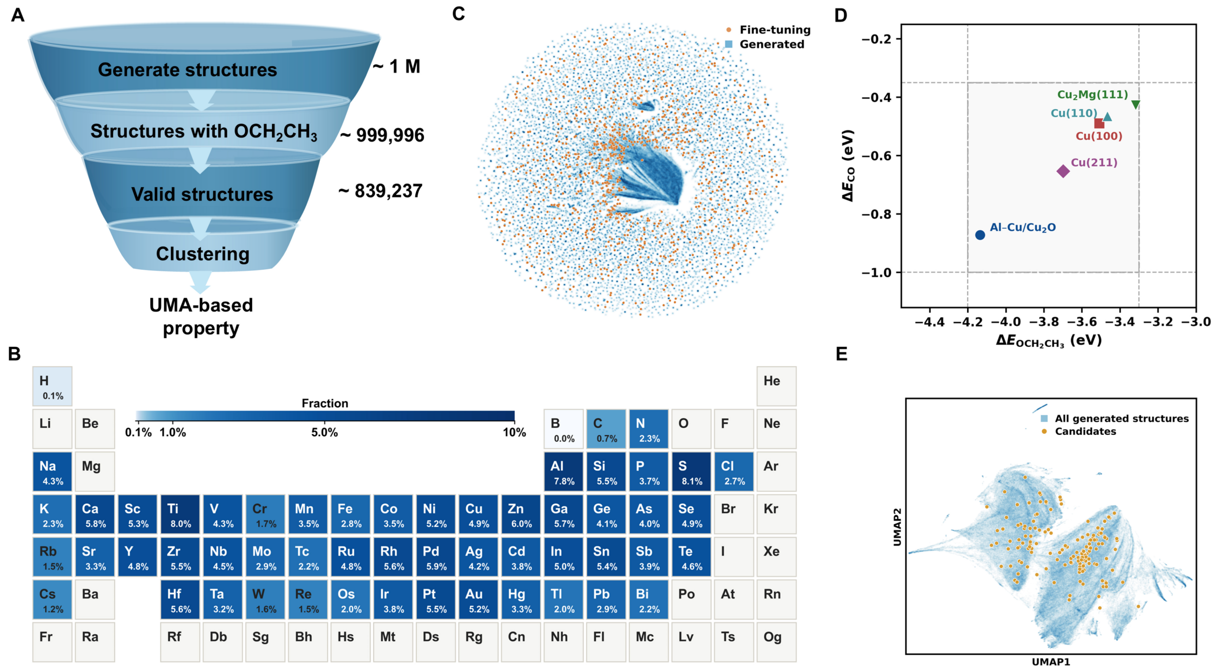


Figure 2. Million-scale generation and reference-calibrated local-environment compression. (A) Sequential validity filtering of the million-scale generated adsorption structures; (B) Elemental distribution of the generated $^*OCH_2CH_3$ adsorption systems; (C) UMAP projection of UMA-extracted substrate embeddings for fine-tuning and generated structures; (D) Reference-calibrated thermodynamic adsorption-energy windows for $^*OCH_2CH_3$, *CO, and *H. E: O-centered UMAP/high-hit structures. UMAP: Uniform Manifold Approximation and Projection; UMA: universal models for atom.

($^*OCH_2CH_3$/CO/H) screening rather than relying on a single adsorption descriptor [Supplementary Figure 3].

To establish a chemically meaningful candidate region, reference-calibrated adsorption-energy windows were constructed using established CO_2RR-active surfaces including Cu (100), Cu (110), Cu (211), Al-Cu/Cu_2O, and Cu_2Mg (111) as benchmarks [Figure 2D][56,57]. The working ranges thus were set to -4.2 eV ≤ $\Delta E_{OCH2CH3}$ ≤ -3.3 eV and -1.0 eV ≤ ΔE_{CO} ≤ -0.35 eV, while ΔE_H > -0.2 eV enforced as a first-order gate to exclude overly strong hydrogen binding. These windows are not intended to fully determine ethanol selectivity, but to define a reaction-relevant local surface region for candidate prioritization. Applying the joint $^*OCH_2CH_3$/*CO/*H windows to the 10,000 representative candidates identified high-hit structures in specific regions of the O-centered UMAP space [Figure 2E]. These localized regions were subsequently utilized to define high-hit clusters for downstream spatial neighborhood expansion.

Having identified localized high-hit regions within the oxygen-centered UMAP space, we next utilized these regions to guide neighborhood expansion. Positive structures were not uniformly distributed across the oxygen-centered embedding space, but instead formed localized, candidate-enriched microenvironments. A proximity-resolved analysis further confirmed this spatial enrichment: within clusters containing at least two positive candidates, the hit rate decreased systematically with distance from the nearest positive seed, from 77.4% for the 1-5 nearest neighbors to 62.2% for the 11-20 nearest neighbors and 29.0% beyond the 200th neighbor [Figure 3A]. This distance-dependent decay indicates that the O-centered UMA embedding captures chemically meaningful local similarity. Guided by this non-uniform enrichment pattern, we adopted a cluster-guided expansion strategy. Specifically, all members of clusters containing at least two positive representative structures were included in the expansion pool. For clusters containing only one positive representative, the positive seed and its 20 nearest cluster members, ranked by cosine similarity in

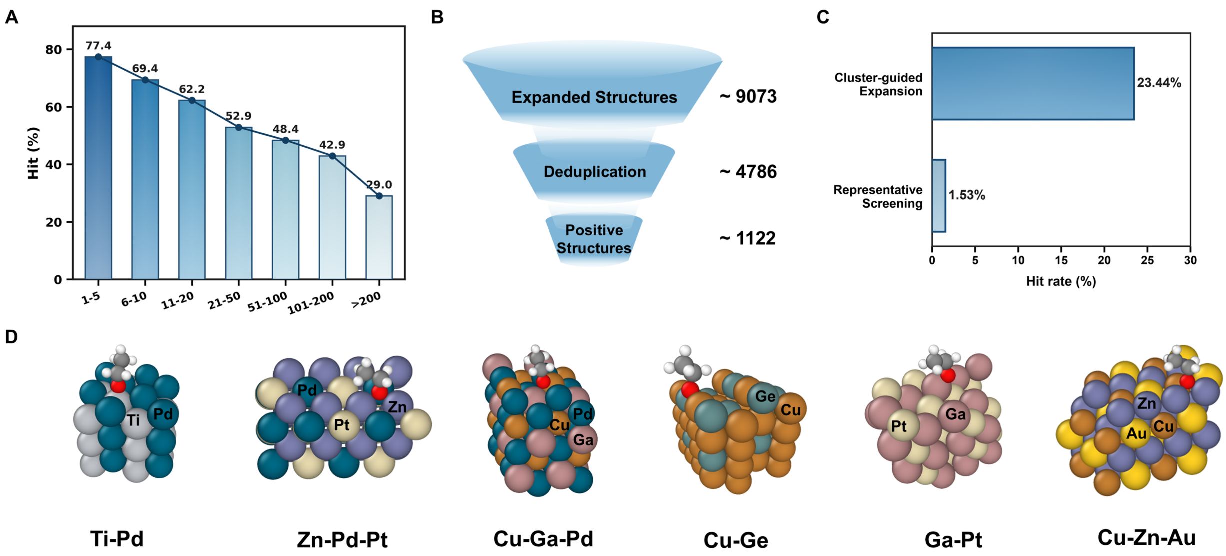


Figure 3. Cluster-guided expansion of high-hit local environments. (A) Candidate hit rate as a function of neighbor rank relative to positive seed structures within local UMA embedding clusters; (B) Workflow from neighborhood expansion to deduplicated positives; (C) Quantitative hit-rate comparison between the representative-sampling baseline and cluster-guided expansion; (D) Representative positive structures showing diverse local adsorption environments and multicomponent surface motifs. UMA: Universal models for atom.

the O-centered embedding space, were retained for subsequent UMA-based property evaluation.

The expanded structures were subjected to UMA-based tri-adsorbate evaluation. For each selected $*OCH_2CH_3$ configuration, the corresponding *CO and *H configurations were constructed at the same oxygen-anchored local site and relaxed using the unified UMA protocol. After applying the joint adsorption-energy windows for $*OCH_2CH_3$, *CO, and *H, the cluster-guided expansion yielded 1,122 unique positive structures after geometric deduplication [Figure 3B]. Compared with the representative-sampling baseline, the cluster-guided expansion strategy elevated the discovery hit rate from 1.53% to 23.44%, achieving an approximate 15-fold enrichment in candidate discovery efficiency [Figure 3C]. Representative structures further show that the enriched candidates span diverse local adsorption motifs and multicomponent surface environments rather than collapsing into a single structural prototype [Figure 3D].

Multi-adsorbate screening and composition-space convergence

The joint $*OCH_2CH_3$/*CO/*H thermodynamic adsorption-energy windows were applied across the expanded structural pool to define the positive candidate space [Figure 4]. Within the $*OCH_2CH_3$ *vs.* *CO adsorption-energy landscape, the screened structures span a broad energetic distribution, reflecting the diverse local environments sampled by the generative model. The reference-calibrated windows delimit a moderate-binding region for $*OCH_2CH_3$ and *CO, while the color-coded *H adsorption energy introduces an additional constraint to exclude structures with overly strong H binding and potentially enhanced HER competition [Figure 4A]. Overlap analysis further shows that the three adsorption descriptors impose complementary rather than redundant constraints. Although each single descriptor selects a relatively broad subset, including 2,399 structures satisfying the $*OCH_2CH_3$ window, 2,248 satisfying the *CO window, and 3,698 satisfying the *H window, only 1,122 structures simultaneously satisfy all three thermodynamic constraints [Figure 4B]. The adsorption-energy distributions confirm this contraction of the candidate space: after multi-adsorbate filtering, the originally broad generated landscape is narrowed into a chemically constrained region characterized by moderate $*OCH_2CH_3$ and *CO binding together with weak *H adsorption [Figure 4C].

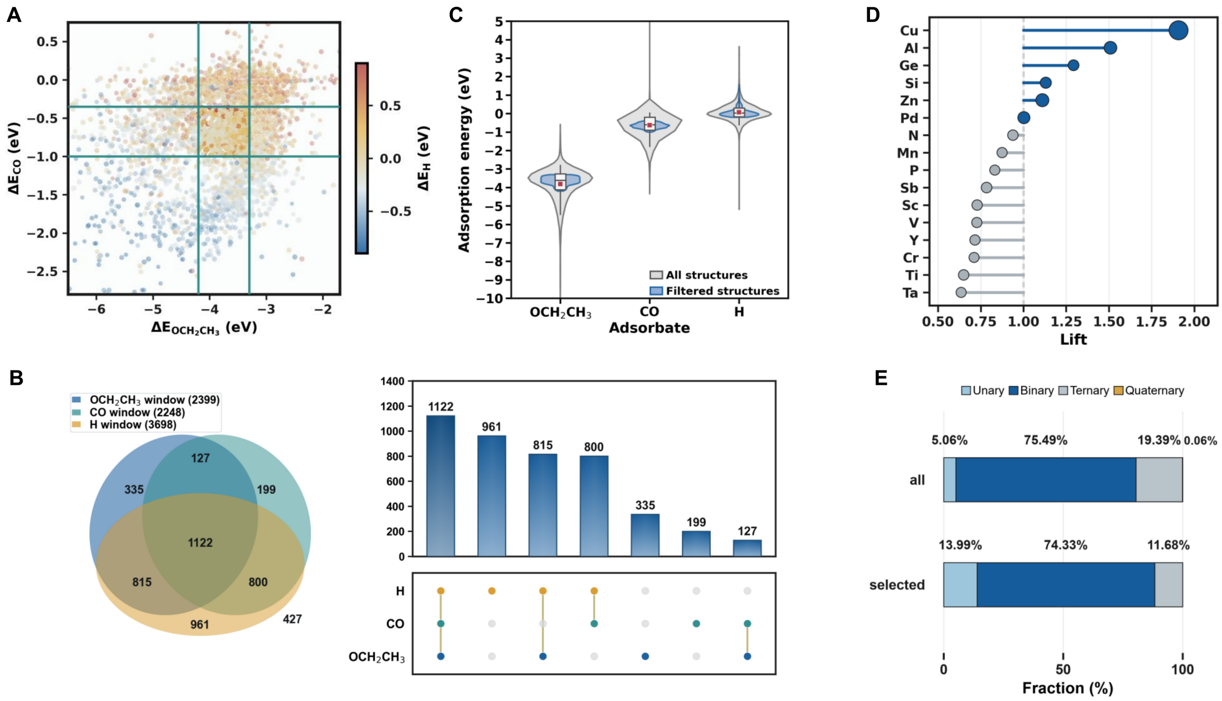


Figure 4. Multi-adsorbate thermodynamic filtration and compositional convergence. (A) $^*OCH_2CH_3$-$^*CO$ adsorption-energy landscape with $^*H$ adsorption-energy coding and the target adsorption-energy window marked by solid lines; (B) Overlap and UpSet analysis of structures satisfying the individual $^*OCH_2CH_3$, $^*CO$, and $^*H$ adsorption-energy criteria; (C) Contraction of adsorption-energy distributions before and after multi-adsorbate filtering; (D) Element-level enrichment of positive structures relative to the full structure set, expressed as lift; (E) Unary, binary, ternary, and quaternary fractions before and after screening.

The resulting positive structures show clear compositional convergence under the joint adsorption constraints. Element-level enrichment analysis identifies Cu as the most strongly enriched element, followed by Al, Ge, Si, Zn, and Pd, suggesting that Cu-centered and compositionally modulated local environments are preferentially selected for balancing oxygenated C_2 retention, CO availability, and suppressed H adsorption [Figure 4D]. Conversely, elements exhibiting enrichment factors (lift values) near or below unity are depleted following tri-adsorbate screening. Compositional breakdown reveals that binary systems remain predominant, shifting only slightly from 75.49% to 74.33% of the candidate pool, whereas unary systems increase markedly from 5.06% to 13.99% and ternary systems decrease from 19.39% to 11.68%. The minor quaternary fraction, accounting for 0.06% before screening, is absent after screening [Figure 4E]. This shift demonstrates that increasing elemental complexity does not inherently guarantee multi-adsorbate thermodynamic compatibility. After excluding systems containing toxic or radioactive elements, the candidate pool was further refined from 1,122 to 1,098 structure-level positive candidates. Overall, multi-adsorbate screening combined with toxicity filtration does not merely reduce candidate counts, but reorganizes the generative surface space into a realistic, high-confidence candidate region for material-level catalyst discovery.

The structure-level positive candidates were further consolidated into material-level systems to map their broad distribution across compositional space [Figure 5]. The binary hit-fraction matrix shows localized hotspots rather than uniform enrichment across all element pairs, with Cu-containing combinations forming the major candidate regions [Figure 5A]. Experimentally validated CO_2RR-active systems, including representative Al-Cu, Cu-Ag, Cu-Zn, and Si-Cu compositions, concentrate squarely within these enriched domains, corroborating the chemical validity of our tri-adsorbate screening framework[57-67]. Meanwhile, additional high-hit binary combinations suggest that the workflow can extend beyond known CO_2RR material families and identify less-explored Cu-centered composition spaces.

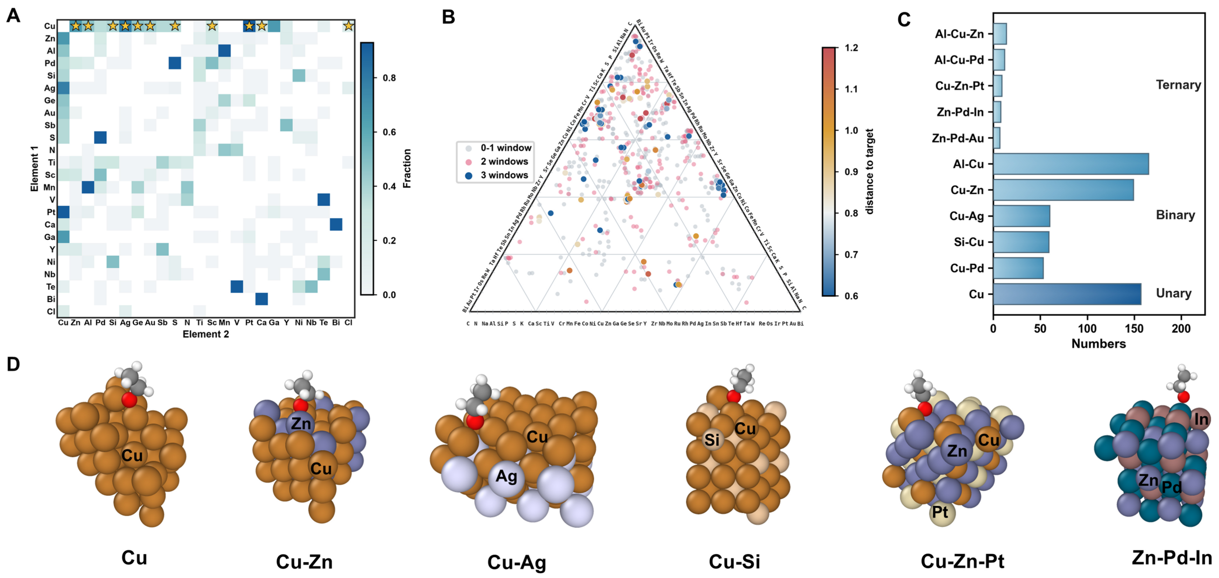


Figure 5. Material-level organization of candidate composition space. (A) Binary composition hit-fraction matrix showing localized enrichment of element pairs under the joint $^*OCH_2CH_3$/*CO/*H adsorption-energy constraints; (B) Pseudo-ternary map of ternary candidate systems, where point categories denote the number of satisfied adsorption-energy criteria and, for candidates satisfying all three criteria, the continuous color scale represents the normalized distance to the target adsorption-energy region; (C) Major unary, binary, and ternary material-level candidates ranked by the number of positive structures; (D) Representative unary, binary, and ternary materials.

The pseudo-ternary composition map shows that ternary candidates are more dispersed, with only a subset satisfying all three adsorption-energy windows or approaching the target region [Figure 5B]. The experimentally reported Al-Cu-Zn ternary composition space is also recovered, indicating that the workflow can capture known composition families beyond binary systems[68]. Final 75 materials satisfy the joint $^*OCH_2CH_3$/*CO/*H constraints [Supplementary Table 3], including 13 experimentally reported CO_2RR-related systems and 62 less-explored candidates identified in this work. Material-level ranking identifies Cu as the leading unary candidate, with Al-Cu and Cu-Zn dominating the binary space and several ternary systems also retained [Figure 5C]. Representative unary, binary, and ternary materials are further displayed in Figure 5D. Overall, these results show that the joint adsorption constraints reorganize the generated surface space into interpretable composition regions, recovering key experimental CO_2RR families while uncovering additional candidate spaces for further mechanistic and experimental validation.

Knowledge extraction and DFT validation of candidate environments

To convert the screened candidate set into interpretable materials knowledge, SGD was applied to identify descriptor-defined regions associated with positive-candidate enrichment [Figure 6]. Unlike multiparameter predictive models that establish global descriptor-performance relationships, SGD identifies compact combinations of descriptor conditions that define candidate-enriched subspaces. Feature-correlation analysis shows that the selected descriptors retain complementary elemental information despite partial correlations, allowing SGD to identify compact combinations that are not simply driven by descriptor redundancy [Supplementary Figure 4]. The top-rule regions identified for both binary and ternary systems exhibit substantially higher pass rates than their corresponding baseline populations, indicating that positive candidates can be concentrated within simple elemental-property-defined regions [Figure 6A]. To further quantify the relative importance of individual conditions within each rule, we performed a post hoc Shapley-based condition-attribution analysis using the StandardQF score as the characteristic function [Supplementary Figure 5].

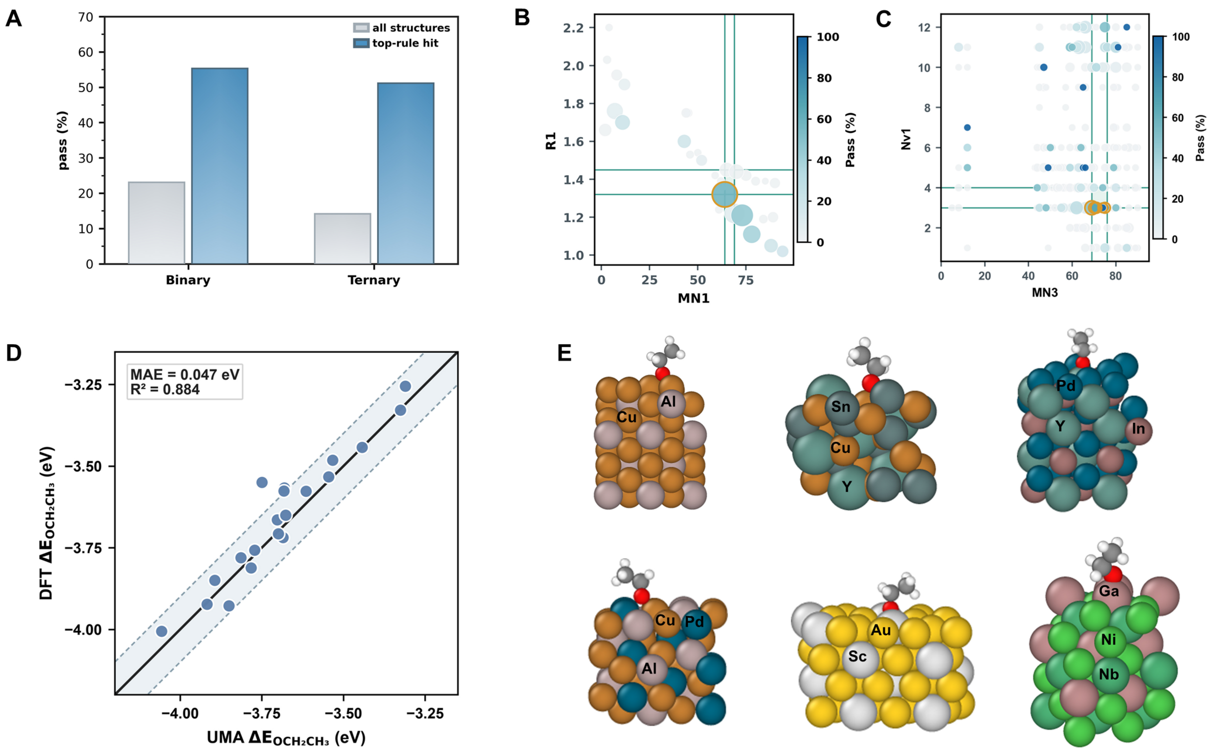


Figure 6. Interpretable rule extraction and DFT benchmarking of candidate environments. (A) Comparison of pass rates between all structures and top-rule regions identified by SGD for binary and ternary candidates; (B and C) Descriptor-space maps showing enriched regions associated with positive-candidate formation in binary and ternary systems. The color represents pass rate, and the highlighted boundaries indicate the selected top-rule regions; (D) Parity plot comparing UMA-predicted and DFT-calculated $^*OCH_2CH_3$ adsorption energies for representative candidate structures. The shaded region indicates ±0.10 eV; (E) Representative DFT-optimized adsorption structures for $^*OCH_2CH_3$. DFT: Density functional theory; SGD: subgroup discovery; UMA: universal models for atom; MAE: mean absolute error; R^2: coefficient of determination.

For binary candidates, the top rule identifies 309 positives among 559 structures, corresponding to a pass rate of 55.28% compared with a baseline of 23.08%, yielding an enrichment factor of 2.40. Shapley attribution ranks the Mendeleev-number condition of the first element, 64 ≤ MN1 < 69, as the dominant contributor, followed by the atomic-radius condition of the first element, 1.32 ≤ R1 < 1.45, and the d-valence-electron condition of the second element, Nd2 ≥ 10, with respective contributions to Q of 0.0620, 0.0369, and 0.0277 [Supplementary Figure 5A]. Accordingly, Figure 6B visualizes the binary candidate landscape in the two highest-ranked descriptor dimensions, MN1 and R1, with the top-rule boundaries highlighted. These results suggest that favorable binary environments are primarily associated with a specific Mendeleev-number regime and moderate atomic-size range, while the nearly filled d-shell condition further refines the local chemical environment required to jointly balance $^*OCH_2CH_3$, *CO, and *H adsorption. To provide physical support for the SGD-derived binary rule, representative rule-matching and rule-mismatching systems were further compared at the electronic-structure level. Cu-Pd, which satisfies all three SGD conditions (MN1 = 64, Nd2 = 10, R1 = 1.32), exhibits a balanced adsorption profile for $^*OCH_2CH_3$, *CO, and *H, whereas Tc-Ru, lying outside the descriptor-defined region (MN1 = 53, Nd2 = 7, R1 = 1.47), binds *CO and *H substantially more strongly while providing weaker stabilization of $^*OCH_2CH_3$ resulting in an imbalanced multi-adsorbate profile. In Cu-Pd, the Cu 3d and Pd 4d states are concentrated mainly below the Fermi level, with limited H 1s-metal d-state overlap but appreciable overlap with the C 2p and O 2p states of *CO and $^*OCH_2CH_3$. By contrast, Tc-Ru exhibits broader Tc/Ru d-state distributions near the Fermi level and stronger overlap with H 1s and C 2p states, consistent with its adsorption imbalance. The more localized charge redistribution on Cu-Pd further indicates more moderate adsorbate-metal coupling [Supplementary Figures 6 and 7].

For ternary candidates, the top rule yields a pass rate of 51.11%, compared with a baseline of 14.12%, corresponding to a 3.62-fold enrichment [Figure 6A]. Shapley analysis identifies the group-valence-electron condition of the first element, Nv1 ∈ [3,4), as the most influential contributor, followed by the Mendeleev-number condition of the third element, MN3 ∈ [69,76), and the d-valence-electron condition of the second element, Nd2 ≥ 10, with respective contributions of 0.033, 0.027, and 0.021 to the StandardQF score. Figure 6C therefore visualizes the candidate distribution using Nv1 and MN3, the two highest-ranked conditions, while Nd2 acts as an additional constraint that further enriches the positive region. The recurrence of Nd2 ≥ 10 in both binary and ternary rules points to the broader relevance of filled-d components, whereas the dominant Nv1 and MN3 conditions in ternary systems indicate that the additional compositional degree of freedom is associated with low-valence and chemically tunable elemental components. Together, the SGD and Shapley analyses reveal that candidate enrichment is governed by a hierarchy of complementary elemental descriptors rather than by any single property, providing an interpretable basis for prioritizing composition spaces beyond direct adsorption-energy filtering.

A subsequent synthesizability assessment identified experimental synthesis precedents for 72 of the 75 final candidate systems, with the remaining three evaluated using Materials Project thermodynamic data [Supplementary Tables 4 and 5]. Sc-Pd-Hf shows a high E_{hull} of 2.925 eV·atom^{-1}, indicating limited thermodynamic accessibility. The 20-system UMA-DFT benchmark included four experimentally established CO_2RR-to-ethanol reference systems, together with 16 less-explored screened candidates. Representative unary, binary, and ternary systems with feasible compositional motifs were then subjected to DFT calculations and the results were shown in Supplementary Table 6. The UMA and DFT adsorption energies for the reaction anchor (*OCH_2CH_3) display exceptional agreement, yielding a mean absolute error (MAE) of 0.047 eV and a coefficient of determination (R^2) of 0.884 across the candidate region [Figure 6D]. Consistent agreement is also obtained for *CO and *H adsorption, with MAEs of 0.029 and 0.024 eV and R^2 values of 0.922 and 0.959, respectively [Supplementary Figure 8]. Representative optimized structures further illustrate that the selected candidates cover diverse unary, binary, and ternary local adsorption environments, rather than collapsing into a single structural prototype [Figure 6E and Supplementary Figures 9-11]. Because CO-CO coupling is a key C-C bond-forming step toward C_2 products in CO_2RR[7], we first evaluated the reaction energy difference between the initial and final states for five representative systems, using Cu as a reference. Cu-Pd-Sn exhibited a substantially higher reaction energy and was therefore excluded from further kinetic analysis [Supplementary Table 7]. Activation barriers were subsequently calculated for Ga-Pt, Sc-Cu-Ga, Al-Cu-Pd, and Cu [Figure 7 and Supplementary Figure 12]. The resulting barriers are 0.613, 1.478, 0.995, and 1.172 eV, respectively. Notably, both Ga-Pt and Al-Cu-Pd exhibit lower barriers than Cu, indicating more favorable CO-CO coupling kinetics. These results provide an additional kinetic criterion for refining the thermodynamically screened candidates. Together, these results demonstrate that the workflow not only identifies candidate surfaces but also extracts interpretable composition rules and provides first-principles validation for representative adsorption environments. The present work establishes a uniform high-throughput thermodynamic screening framework under solvent-free conditions, while solvent, mass-transport, particle-aggregation, and other electrochemical environmental effects will be considered in future validation and refinement under realistic catalyst and electrode conditions.

CONCLUSIONS

In summary, this work establishes a reaction-constrained generative materials informatics framework for ethanol-pathway-oriented CO_2RR catalyst discovery. Using *OCH_2CH_3 as a late-stage oxygenated C_2 reaction anchor, the generative workflow sampled one million adsorption configurations and retained 839,237 valid structures for UMA-based representation and cluster-guided exploration. This strategy increased the positive-candidate hit rate from 1.53% to 23.44%. Subsequent tri-adsorbate screening and elemental safety filtering yielded 1,098 positive structures, which were consolidated into 75 material-level systems, including

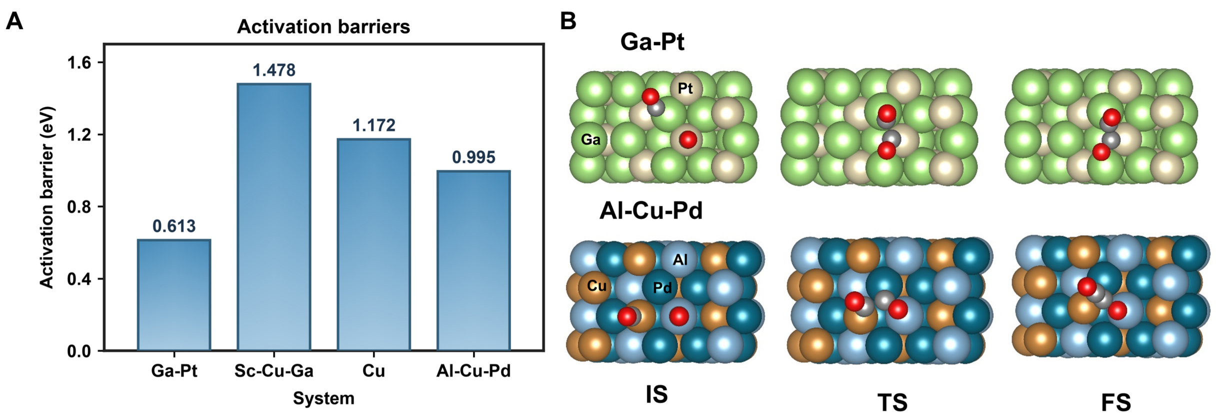


Figure 7. Calculated activation barriers and representative structures for CO-CO coupling. (A) Activation barriers for Ga-Pt, Sc-Cu-Ga, Cu, and Al-Cu-Pd; (B) IS, TS, and FS structures for representative Ga-Pt and Al-Cu-Pd systems. IS: Initial-state; TS: transition-state; FS: final-state.

13 materials previously reported for CO_2RR and 62 less-explored candidates. Composition-space analysis revealed Cu-centered regions that recover established CO_2RR families and broaden the accessible compositional landscape. SGD further identified interpretable descriptor hierarchies associated with candidate enrichment, highlighting the contributions of specific Mendeleev-number ranges, atomic-size regimes, low-valence components, and filled-d configurations. Representative DFT calculations confirmed the screening reliability of UMA predictions for $^*OCH_2CH_3$, *CO, and *H adsorption. Additional CO-CO coupling calculations showed that Ga-Pt and Al-Cu-Pd exhibit lower activation barriers than Cu, providing an initial kinetic criterion for candidate prioritization. Further refinement under realistic electrochemical environments will enable more comprehensive assessment of the screened candidates. Overall, this framework provides a transferable materials informatics strategy for converting large generative surface spaces into reaction-constrained candidate regions and interpretable compositional guidance for multistep catalytic reactions.

DECLARATIONS

Acknowledgments

The authors acknowledge the National Supercomputing Center in Shanghai for providing computational resources.

Authors' contributions

Project conception and initiation: Qi, R.
Supervision: Qi, R.
Calculations and data analysis: Tang, D.
Manuscript drafting: Qi, R.
Manuscript review and editing: Gao, Y.
Discussion and approval of the final manuscript: Tang, D.; Li, R.; Lei, Z.; Mao, Q.; Das, R.; Zhu, B.; Gao, Y.; Qi, R.

Availability of data and materials

Some results supporting the study are presented in the Supplementary Materials. Other raw data that support the findings of this study are available from the corresponding author upon reasonable request.

AI and AI-assisted tools statement

During the preparation of this manuscript, the AI tool ChatGPT (GPT-5.5, released 2026-04-23) was used solely for language editing. The tool did not influence the study design, data generation, analysis, interpretation, or the scientific content of the work. All authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

Financial support and sponsorship

Qi, R. thanks the Innovation Program of Shanghai Advanced Research Institute, CAS (2025CP007).

Conflicts of interest

Mao, Q. and Das, R. are affiliated with Titan Holdings, while the other authors have declared that they have no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Supplementary Materials

Supplementary Materials

REFERENCES

1. Bushuyev, O. S.; De Luna, P.; Dinh, C. T.; et al. What should we make with CO_2 and how can we make it? *Joule* **2018**, *2*, 825-32. DOI
2. De Luna, P.; Hahn, C.; Higgins, D.; Jaffer, S. A.; Jaramillo, T. F.; Sargent, E. H. What would it take for renewably powered electrosynthesis to displace petrochemical processes? *Science* **2019**, *364*, eaav3506. DOI PubMed
3. Jouny, M.; Luc, W.; Jiao, F. General techno-economic analysis of CO_2 electrolysis systems. *Ind. Eng. Chem. Res.* **2018**, *57*, 2165-77. DOI
4. Nitopi, S.; Bertheussen, E.; Scott, S. B.; et al. Progress and perspectives of electrochemical CO_2 reduction on copper in aqueous electrolyte. *Chem. Rev.* **2019**, *119*, 7610-72. DOI PubMed
5. Zhu, P.; Wang, H. High-purity and high-concentration liquid fuels through CO_2 electroreduction. *Nat. Catal.* **2021**, *4*, 943-51. DOI
6. Tran, K.; Ulissi, Z. W. Active learning across intermetallics to guide discovery of electrocatalysts for CO_2 reduction and H_2 evolution. *Nat. Catal.* **2018**, *1*, 696-703. DOI
7. Zhao, F.; Huang, B.; Zhang, Y.; Wei, T.; Zhang, J.; Zhao, D. A review of CO_2 electroreduction to ethanol: C-C coupling mechanistic insights and catalyst design. *Nanomicro. Lett.* **2026**, *18*, 313. DOI PubMed PMC
8. Li, F.; Li, Y. C.; Wang, Z.; et al. Cooperative CO_2-to-ethanol conversion via enriched intermediates at molecule–metal catalyst interfaces. *Nat. Catal.* **2020**, *3*, 75-82. DOI
9. Luo, M.; Wang, Z.; Li, Y. C.; et al. Hydroxide promotes carbon dioxide electroreduction to ethanol on copper via tuning of adsorbed hydrogen. *Nat. Commun.* **2019**, *10*, 5814. DOI PubMed PMC
10. Zhan, C.; Dattila, F.; Rettenmaier, C.; et al. Key intermediates and Cu active sites for CO_2 electroreduction to ethylene and ethanol. *Nat. Energy.* **2024**, *9*, 1485-96. DOI PubMed PMC
11. Liu, X.; Liang, J.; Wang, Z.; Li, Q.; Deng, Y.; Wang, H. Building catalyst exploration highways by integrating high-throughput and machine learning technologies. *Adv. Energy. Mater.* **2026**, *16*, e05497. DOI
12. Levy, O.; Hart, G. L.; Curtarolo, S. Uncovering compounds by synergy of cluster expansion and high-throughput methods. *J. Am. Chem. Soc.* **2010**, *132*, 4830-3. DOI PubMed
13. Greeley, J.; Jaramillo, T. F.; Bonde, J.; Chorkendorff, I. B.; Nørskov, J. K. Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. *Nat. Mater.* **2006**, *5*, 909-13. DOI PubMed
14. Curtarolo, S.; Hart, G. L.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O. The high-throughput highway to computational materials design. *Nat. Mater.* **2013**, *12*, 191-201. DOI PubMed
15. Pan, Y.; Shan, X.; Cai, F.; Gao, H.; Xu, J.; Zhou, M. Accelerating the discovery of oxygen reduction electrocatalysts: high-throughput screening of element combinations in Pt-based high-entropy alloys. *Angew. Chem. Int. Ed. Engl.* **2024**, *63*, e202407116. DOI PubMed
16. Ivanciuc, O. Applications of support vector machines in chemistry. In *Reviews in computational chemistry*, Vol 23; Wiley, 2007; pp. 291-400. DOI

17. Machado Cavalcanti, F.; Emilia Kozonoe, C.; André Pacheco, K.; Maria de Brito Alves, R. Application of artificial neural networks to chemical and process engineering. In *Deep learning applications*. IntechOpen; 2021. DOI
18. Myles, A. J.; Feudale, R. N.; Liu, Y.; Woody, N. A.; Brown, S. D. An introduction to decision tree modeling. *J. Chemom.* **2004**, *18*, 275-85. DOI
19. Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547-55. DOI PubMed
20. Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting reaction performance in C-N cross-coupling using machine learning. *Science* **2018**, *360*, 186-90. DOI PubMed
21. Reid, J. P.; Sigman, M. S. Holistic prediction of enantioselectivity in asymmetric catalysis. *Nature* **2019**, *571*, 343-8. DOI PubMed PMC
22. Huang, B.; von Lilienfeld, O. A. Ab initio machine learning in chemical compound space. *Chem. Rev.* **2021**, *121*, 10001-36. DOI PubMed PMC
23. Han, Z.; Gao, R.; Wang, T.; et al. Machine-learning-assisted design of a binary descriptor to decipher electronic and structural effects on sulfur reduction kinetics. *Nat. Catal.* **2023**, *6*, 1073-86. DOI
24. Esterhuizen, J. A.; Goldsmith, B. R.; Linic, S. Interpretable machine learning for knowledge generation in heterogeneous catalysis. *Nat. Catal.* **2022**, *5*, 175-84. DOI
25. Song, Z.; Wang, X.; Liu, F.; et al. Distilling universal activity descriptors for perovskite catalysts from multiple data sources via multi-task symbolic regression. *Mater. Horiz.* **2023**, *10*, 1651-60. DOI PubMed
26. Weng, B.; Song, Z.; Zhu, R.; et al. Simple descriptor derived from symbolic regression accelerating the discovery of new perovskite catalysts. *Nat. Commun.* **2020**, *11*, 3513. DOI PubMed PMC
27. Zhong, M.; Tran, K.; Min, Y.; et al. Accelerated discovery of CO_2 electrocatalysts using active machine learning. *Nature* **2020**, *581*, 178-83. DOI PubMed
28. Ulissi, Z. W.; Tang, M. T.; Xiao, J.; et al. Machine-learning methods enable exhaustive searches for active bimetallic facets and reveal active site motifs for CO_2 reduction. *ACS. Catal.* **2017**, *7*, 6600-8. DOI
29. Sun, J.; Tu, R.; Xu, Y.; et al. Machine learning aided design of single-atom alloy catalysts for methane cracking. *Nat. Commun.* **2024**, *15*, 6036. DOI PubMed PMC
30. Zhang, Q.; Dong, Z.; Liu, X.; et al. Multiparameter machine learning quantifies electronic dominance in Pd-catalyzed formic acid dehydrogenation. *Nano. Lett.* **2026**, *26*, 7927-36. DOI PubMed
31. Bertheussen, E.; Verdaguer-Casadevall, A.; Ravasio, D.; et al. Acetaldehyde as an intermediate in the electroreduction of carbon monoxide to ethanol on oxide-derived copper. *Angew. Chem. Int. Ed. Engl.* **2016**, *55*, 1450-4. DOI PubMed PMC
32. Piqué, O.; Low, Q. H.; Handoko, A. D.; Yeo, B. S.; Calle-Vallejo, F. Selectivity map for the late stages of CO and CO_2 reduction to C_2 species on copper electrodes. *Angew. Chem. Int. Ed. Engl.* **2021**, *60*, 10784-90. DOI PubMed
33. Vasileff, A.; Zhu, Y.; Zhi, X.; et al. Electrochemical reduction of CO_2 to ethane through stabilization of an ethoxy intermediate. *Angew. Chem. Int. Ed. Engl.* **2020**, *59*, 19649-53. DOI PubMed
34. Goodfellow; IJ; Pouget-Abadie; J; Mirza, M.; et al. Generative adversarial nets. In *Proceedings of the 28th International Conference on Neural Information Processing Systems*, Montréal, Canada, December 8-13, 2014; Ghahramani, Z.; Welling, M.; Cortes, C.; Lawrence, N. D.; Weinberger, K. Q., Eds.; Curran Associates, Inc.: Red Hook, USA, 2014; Vol. 2, pp 2672-80. DOI
35. Yan, D.; Smith, A. D.; Chen, C. C. Structure prediction and materials design with generative neural networks. *Nat. Comput. Sci.* **2023**, *3*, 572-4. DOI PubMed
36. Zhao, Y.; Al-Fahdi, M.; Hu, M.; et al. High-throughput discovery of novel cubic crystal materials using deep generative neural networks. *Adv. Sci.* **2021**, *8*, e2100566. DOI PubMed PMC
37. Yao, Z.; Sánchez-Lengeling, B.; Bobbitt, N. S.; et al. Inverse design of nanoporous crystalline reticular materials with deep generative models. *Nat. Mach. Intell.* **2021**, *3*, 76-86. DOI
38. Zeni, C.; Pinsler, R.; Zügner, D.; et al. A generative model for inorganic materials design. *Nature* **2025**, *639*, 624-32. DOI PubMed PMC
39. Bran, A. M.; Cox, S.; Schilter, O.; Baldassari, C.; White, A. D.; Schwaller, P. Augmenting large language models with chemistry tools. *Nat. Mach. Intell.* **2024**, *6*, 525-35. DOI PubMed PMC
40. Jablonka, K. M.; Schwaller, P.; Ortega-Guerrero, A.; Smit, B. Leveraging large language models for predictive chemistry. *Nat. Mach. Intell.* **2024**, *6*, 161-9. DOI
41. Boiko, D. A.; MacKnight, R.; Kline, B.; Gomes, G. Autonomous chemical research with large language models. *Nature* **2023**, *624*, 570-8. DOI PubMed PMC
42. Antunes, L. M.; Butler, K. T.; Grau-Crespo, R. Crystal structure generation with autoregressive large language modeling. *Nat. Commun.* **2024**, *15*, 10570. DOI PubMed PMC

43. Mok, D. H.; Back, S. Generative pretrained transformer for heterogeneous catalysts. *J. Am. Chem. Soc.* **2024**, *146*, 33712-22. DOI PubMed
44. Song, Z.; Fan, L.; Lu, S.; Ling, C.; Zhou, Q.; Wang, J. Inverse design of promising electrocatalysts for CO_2 reduction via generative models and bird swarm algorithm. *Nat. Commun.* **2025**, *16*, 1053. DOI PubMed PMC
45. Li, R.; Zhang, S.; Tang, Q.; et al. Generative intelligence explores the chemical space of ten million catalysts. *Chem. Sci.* **2026**, *17*, 12996-3006. DOI PubMed PMC
46. Chanussot, L.; Das, A.; Goyal, S.; et al. Open catalyst 2020 (OC20) dataset and community challenges. *ACS. Catal.* **2021**, *11*, 6059-72. DOI
47. Wood, B. M.; Dzamba, M.; Fu, X.; et al. UMA: a family of universal models for atoms. *arXiv* **2025**, arXiv:2506.23971. Available online: https://doi.org/10.48550/arXiv.2506.23971. (accessed 2026-09-23).
48. Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **1996**, *6*, 15-50. DOI
49. Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B. Condens. Matter.* **1996**, *54*, 11169-86. DOI PubMed
50. Hammer, B.; Hansen, L. B.; Nørskov, J. K. Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Phys. Rev. B.* **1999**, *59*, 7413-21. DOI
51. Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B.* **1999**, *59*, 1758-75. DOI
52. Atzmueller, M. Subgroup discovery. *WIREs. Data. Min. Knowl. Discov.* **2015**, *5*, 35-49. DOI
53. Lemmerich, F.; Becker, M. pysubgroup: Easy-to-use subgroup discovery in Python. In *Machine learning and knowledge discovery in databases*. Cham: Springer International Publishing, 2019; Vol. 11053, pp. 658-62. DOI
54. Shapley, L. S. 17. A VALUE for n-person games. In *Contributions to the theory of games (AM-28)*, Volume II; Princeton University Press, 1953; pp. 307-18. DOI
55. Mcinnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: uniform manifold approximation and projection. *J. Open. Source. Softw.* **2018**, *3*, 861. DOI
56. Peng, C.; Ma, J.; Luo, G.; et al. (111) Facet-oriented Cu_2Mg intermetallic compound with Cu_3-Mg sites for CO_2 electroreduction to ethanol with industrial current density. *Angew. Chem. Int. Ed. Engl.* **2024**, *63*, e202316907. DOI PubMed
57. Zhang, L.; Feng, J.; Wu, L.; et al. Oxophilicity-controlled CO_2 electroreduction to C_{2+} alcohols over Lewis acid metal-doped $Cu^{\delta+}$ catalysts. *J. Am. Chem. Soc.* **2023**, *145*, 21945-54. DOI PubMed
58. Hou, Y.; Li, P.; Wang, Y.; et al. Steering acidic CO_2 electroreduction to multicarbon alcohols with high efficiency and selectivity over calcium-induced bicrystalline Cu architecture. *ACS. Catal.* **2025**, *15*, 19227-37. DOI
59. Zhu, Y.; Zhu, J.; Li, H.; et al. Confinement effect and hydrogen species modulation toward enhanced electrochemical CO_2 reduction to ethanol. *Research* **2025**, *8*, 0796. DOI PubMed PMC
60. Hoang, T. T. H.; Verma, S.; Ma, S.; et al. Nanoporous copper-silver alloys by additive-controlled electrodeposition for the selective electroreduction of CO_2 to ethylene and ethanol. *J. Am. Chem. Soc.* **2018**, *140*, 5791-7. DOI PubMed
61. Kuang, S.; Su, Y.; Li, M.; et al. Asymmetrical electrohydrogenation of CO_2 to ethanol with copper-gold heterojunctions. *Proc. Natl. Acad. Sci. U. S. A.* **2023**, *120*, e2214175120. DOI PubMed PMC
62. Zhou, X.; Zheng, Z.; Zhang, J.; Ji, J.; Zhang, X. Electronic state modulation of Cu by Ge for post C–C steps in electrochemical CO_2 reduction. *Adv. Funct. Mater.* **2026**, *36*, e75284. DOI
63. Bae, S.; Yun, G.; Gwon, Y.; Kim, S. Y.; Sohn, Y. Interface engineering of Pt-deposited Cu electrodes via laser ablation for enhanced electrochemical CO_2 reduction to multi-carbon products. *Adv. Ind. Eng. Chem.* **2025**, *1*, 19. DOI
64. Varandili, S. B.; Stoian, D.; Vavra, J.; et al. Elucidating the structure-dependent selectivity of CuZn towards methane and ethanol in CO_2 electroreduction using tailored Cu/ZnO precatalysts. *Chem. Sci.* **2021**, *12*, 14484-93. DOI PubMed PMC
65. Guo, C.; Guo, Y.; Shi, Y.; et al. Electrocatalytic reduction of CO_2 to ethanol at close to theoretical potential via engineering abundant electron-donating $Cu^{\delta+}$ species. *Angew. Chem. Int. Ed. Engl.* **2022**, *61*, e202205909. DOI PubMed
66. Crandall, B. S.; Qi, Z.; Foucher, A. C.; et al. Cu based dilute alloys for tuning the C_{2+} selectivity of electrochemical CO_2 reduction. *Small* **2024**, *20*, e2401656. DOI PubMed
67. Li, J.; Ozden, A.; Wan, M.; et al. Silica-copper catalyst interfaces enable carbon-carbon coupling towards ethylene electrosynthesis. *Nat. Commun.* **2021**, *12*, 2808. DOI PubMed PMC
68. Guzmán, H.; Albo, J.; Irabien, A.; Castellino, M.; Hernández, S. Role of electrochemical cell configuration on the selectivity of CuZnAl-oxide-based electrodes for the continuous CO_2 conversion: aqueous electrolyte vs. catholyte-less configuration. *Discov. Chem. Eng.* **2024**, *4*, 49. DOI

Disclaimer/Publisher's Note: All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.