Generative Intelligence Explores Chemical Space at Ten Million Catalysts

Ruili Li, [a,f] Shuoqi Zhang, [a,f] Qingli Tang, [b,c] Qingqing Mao, [d,e] Ritankar Das, [e] Rui Qi,*[b,c] Beien Zhu,*[b,c] and Yi Gao*[b,c]

a. Key Laboratory of Interfacial Physics and Technology, Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201800, China

b. Photon Science Research Center for Carbon Dioxide, Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China

c. State Key Laboratory of Low Carbon Catalysis and Carbon Dioxide Utilization, Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China

d. Incept Labs, Hayward, CA 94541, United States

e. Titan Holdings, Hayward, CA 94541, United States

f. University of Chinese Academy of Sciences, Beijing 100049, China

Email: Rui Qi: qir@sari.ac.cn, Beien Zhu: zhube@sari.ac.cn, Yi Gao: gaoyi@sari.ac.cn

Methods

Structures Generation

We adopted a Transformer-based framework [1] as the generative model for heterogeneous catalyst structures. The model is based on a GPT-2 decoder-only architecture with 12 self-attention layers, 8 attention heads, an embedding dimension of 512, and a batch size of 144. The model was trained for 10 epochs. Catalyst surface structures were represented as token sequences encoding lattice parameters, atomic species, and atomic coordinates via coordinate-level tokenization. Pretraining was performed on 2 million structures from the OC20-S2EF 2M dataset, and validation was conducted using the OC20-S2EF Val-ID dataset. To bias the generative distribution toward a specific adsorbate, the pretrained model was further fine-tuned on 12,000 catalyst structures containing only CH_3 adsorbates, without modifying the model architecture or tokenizer. This fine-tuning step selectively reshaped the generation space while preserving the structural priors learned during large-scale pretraining.

To enable large-scale structure generation, we implemented a GPU-based parallel generation framework that performs batched inference directly on the device, maximizing GPU computational throughput while minimizing GPU-CPU communication. Using NVIDIA H100 GPUs, the framework achieves a sustained generation rate of up to approximately 2.5 million structures per day. This strategy enables the practical production of tens of millions of catalyst structures at manageable computational cost, thereby supporting subsequent large-scale statistical analysis and screening workflows.

Structure featurization

For each generated structure, we extracted a set of 49 numeric descriptors to represent its composition and geometric/structural characteristics. The feature vector includes: (i) system size and composition statistics (e.g., number of atoms, number of elements, major-element fraction, fractions of the top-k most abundant elements frac1–frac5); (ii) geometric spans along Cartesian directions (e.g., span_x, span_y, span_z); (iii) local coordination statistics computed from neighbor lists (e.g., CN_mean, CN_std, and fractions of atoms above coordination thresholds such as CN_ge3_frac, CN_ge4_frac, CN_ge6_frac); and (iv) short-range structural fingerprints derived from distance/radial distribution summaries (e.g., nn_d_mean, nn_d_std, and binned RDF-like terms). In structures containing the adsorbate CH_3, for feature extraction focused on the substrate, those adsorbate atoms were excluded before computing the above descriptors.

Data cleaning and normalization

All features were assembled into a tabular matrix $X \in R^{N \times 49}$. Structures failing parsing or producing invalid values were filtered out (e.g., any NaNs). Prior to dimensionality reduction, each feature was standardized to zero mean and unit variance (z-score normalization) to prevent large-scale features (e.g., natoms) from dominating distance-based methods.

PCA preprocessing

Because UMAP is sensitive to noise and can become slow/noisy on high-dimensional inputs, we applied principal component analysis (PCA) as a preprocessing step to compress the standardized feature vectors. We retained 40 principal components. This choice is also consistent with the practical constraint that PCA dimensionality cannot exceed the number of input features (here 49). The PCA-transformed representation $Z \in R^{N \times 40}$was then used as input to UMAP.

UMAP-based dimensionality reduction and visualization

UMAP was employed to obtain two-dimensional embeddings of the generated catalyst structures for qualitative inspection of their global organization in feature space. Given that the total number of successfully generated structures exceeds several million, direct embedding of the full dataset is computationally impractical and unnecessary for exploratory visualization. Instead, UMAP models were fitted using randomly sampled subsets of the generated structures, while the fitted models were subsequently used to transform the remaining structures. In practice, subsets of 300,000 structures for global catalyst descriptors and 200,000 structures for local adsorption environment of CH_3 were sampled proportionally from all feature files to ensure representative coverage of the generated structural space and to avoid bias toward specific compositions or structural motifs. Only physically valid structures without atomic overlap (minimum interatomic distance ≥ 0.7 Å) were included in the sampling.

Two closely related feature representations were considered within the same UMAP framework. In the first case, structures were represented using global catalyst descriptors capturing overall geometric and compositional characteristics. In the second case, the feature construction was centered on the local adsorption environment of CH_3, incorporating local geometric descriptors (e.g., nearest-neighbor distances, coordination numbers, and adsorbate-surface metrics), composition-related quantities (such as dominant elemental fraction and number of unique elements), and discretized radial distribution function (RDF) features. To emphasize the shape of the local atomic environment rather than absolute atom counts, RDF vectors were normalized on a per-structure basis, and all features were standardized to zero mean and unit variance prior to dimensionality reduction.

UMAP embeddings were generated using a fixed set of hyperparameters, with the number of nearest neighbors set to 30 and the minimum inter-point distance set to 0.05, and the Euclidean distance metric employed throughout. These parameters balance preservation of local neighborhood structure with global continuity of the embedding. The resulting two-dimensional projections were used exclusively for qualitative visualization, providing an overview of whether the generated structures populate a relatively continuous and compact region of feature space or exhibit pronounced fragmentation into isolated regions.

Density rendering of large-scale embeddings

To visualize nearly one million points without overplotting, we rendered the embedding as a pixel-wise density map. Specifically, the 2D plane was discretized into a canvas of size 1500 × 700 pixels (width/height). For each pixel, we accumulated the count of points falling into that pixel. To improve readability and avoid a few extremely dense pixels dominating contrast, we clipped densities above a high percentile (clip_percentile=99.9), and removed ultra-low-density background by applying a minimum pixel count threshold (min_count=1). Color intensity in the figure therefore reflects the local point density per pixel (after clipping and thresholding), not a physical property.

Hash

Deduplication was performed based on the local atomic environment around the CH_3. For each structure, the C atom of the CH_3 group was identified and used as the reference center. Distances between this C atom and all catalyst atoms were computed, and neighboring atoms within a cutoff radius of 5.0 Å were considered. The local coordination environment was characterized by the coordination number, the sorted nearest-neighbor distances, and element-resolved neighbor counts within discretized radial distance bins. To construct a stable and permutation-invariant representation, neighbor distances were assigned to predefined radial bins, and the resulting element-bin pairs were counted. The nearest-neighbor distances were further quantized with a fixed

resolution to reduce numerical noise. These descriptors were concatenated into a canonical string representation encoding the local chemical environment, which was subsequently mapped to a unique hash using a cryptographic hash function. Structures sharing identical hashes were considered locally equivalent and only one representative was retained.

Large-scale clustering and representative structure selection

To enable clustering and organization of millions of generated catalyst structures, a two-stage dimensionality reduction and clustering strategy was employed. Because direct clustering in the original high-dimensional feature space is computationally prohibitive, a representative subset of 200 000 deduplicated structures was randomly selected to construct the clustering model. Specifically, each structure was encoded as a sparse vector consisting of (i) hashed element-distance-bin tokens derived from the local atomic environment within a cutoff radius of 5.0 Å, using distance bins of [0, 2, 3, 4, 5] Å and a hashing dimension of 2048, and (ii) a set of continuous descriptors including coordination number, minimum adsorbate–surface distance, and the distances to the five nearest neighboring atoms. Continuous features were scaled by a factor of 0.2 to balance their contribution relative to the hashed categorical components. This representation preserves local geometric information while maintaining a fixed dimensionality suitable for large-scale processing. Prior to clustering, the high-dimensional sparse feature matrix was projected onto a 64-dimensional latent space using truncated singular value decomposition (SVD). This linear projection reduces noise and redundancy in the original feature space while substantially lowering the computational cost of clustering. K-means clustering was then performed in the reduced latent space with the number of clusters set to 2,000. The resulting SVD/K-means model was saved and subsequently used to assign cluster labels to all generated structures by applying the same feature extraction and SVD transformation, followed by K-means prediction.

To enable efficient property-based screening while preserving structural diversity, strategy was adopted for machine-learning potential (MLP) evaluation. From each of the 2,000 clusters, a random subset corresponding to 1% of the structures was selected, yielding approximately 70,000 structures in total for initial MLP calculations. For each cluster, the fraction of sampled structures exhibiting negative CH_3 binding energies was computed based on the MLP predictions. Clusters were then ranked according to this fraction, and the top 400 clusters were selected. All structures belonging to these selected clusters (approximately 1.2 million structures in total) were subsequently subjected to full MLP evaluation for downstream analysis.

Local-environment similarity and extrapolation analysis

For each CH_3-adsorbed structure, the carbon atom of the CH_3 adsorbate was identified, and all atoms within a radial cutoff of 5.0 Å from this carbon center were extracted to define a CH_3-centered local environment. The extracted local environment retained the atomic species and coordinates of all atoms within the cutoff sphere and was treated as the basic object for subsequent similarity analysis. This procedure was applied to both the fine-tuning set and the generated structures.

Each extracted local environment was encoded using the Smooth Overlap of Atomic Positions (SOAP) descriptor. SOAP provides a continuous vector representation of the local atomic arrangement by capturing both elemental identities and geometric distributions of neighboring atoms around the local region. In this work, SOAP descriptors were computed with a radial cutoff of 5.0 Å, $n_{max} = 6$, $l_{max} = 4$, and $\sigma = 0.3$. The descriptor was evaluated for each extracted local environment and averaged over the atoms within that local environment to obtain one fixed-length vector per structure. All SOAP vectors were subsequently L2-normalized before similarity

comparison.

Similarity between local environments was quantified in SOAP space using cosine distance,

$$d(x,y) = 1 - \frac{x \cdot y}{\| x \| \| y \|},$$

where x and y are the normalized SOAP vectors of two local environments. A smaller distance indicates a more similar local environment, whereas a larger distance indicates a stronger deviation in local atomic motif. For each generated local environment, we identified its nearest neighbor in the fine-tuning set and recorded the corresponding nearest-neighbor distance, yielding the generated-to-fine-tuning distance distribution, $d^{gen \to FT}$. In parallel, for each local environment in the fine-tuning set, we identified its nearest neighbor among the remaining fine-tuning structures, yielding the within-fine-tuning distance distribution, $d^{FT \to FT}$. The latter serves as a data-driven baseline for the normal similarity scale of the fine-tuning local-environment space.

To determine whether a generated local environment remained within the normal similarity range of the fine-tuning set or extended beyond it, we used percentile-based thresholds derived from the $d^{FT \to FT}$ distribution. In particular, the 95th and 99th percentiles of $d^{FT \to FT}$, denoted $P_{95}(d^{FT \to FT})$ and $P_{99}(d^{FT \to FT})$, were used as data-driven reference values. Distances exceeding these thresholds indicate that a generated local environment is farther from its nearest fine-tuning neighbor than most local environments are from one another within the fine-tuning set itself. Thus, such structures can be interpreted as lying outside the typical local similarity range of the fine-tuning distribution. Because the within-fine-tuning distance distribution was highly concentrated near zero, the percentile-based thresholds were correspondingly small. To provide a more interpretable stratification of extrapolation extent beyond the fine-tuning boundary, we additionally introduced two empirical reference levels, 0.05 and 0.10, in SOAP cosine distance. These values were not treated as universal or theory-derived novelty criteria, but rather as descriptive thresholds to separate generated local environments into moderate and more pronounced extrapolation regimes after comparison to the fine-tuning baseline.

SGD

The SGD was performed with the RealKD code (https://bitbucket.org/realKD/realkd/). We applied the 15-means algorithm to cluster each feature into 15 groups. The resulting inter-cluster boundaries (a1, a2, ...) formed the basis for constructing inequalities (feature1 < a1), (feature2 ≥ a2), and so forth. While the final results can vary with the number of clusters, prior research indicates that employing a relatively high number yields broadly consistent findings.[2] Candidate subgroups are constructed as combinations of simple inequalities. The kernel of SGD is that subgroups are distinctive when the data distribution within them significantly deviates from the whole sampling. In this work, the data distribution refers to the distribution of a target property (binding energy). The uniqueness is assessed using a quality function for minimizing binding energy:

$$Q_{max}(S) = \frac{s(S)}{s(P)}\left(\frac{\min(S) - \min(P)}{\max(P) - \min(P)}\right)\theta(\min(S) - c)$$

$$Q_{min}(S) = \frac{s(S)}{s(P)}\left(\frac{\max(S) - \max(P)}{\max(P) - \min(P)}\right)(c - \max(S))$$

The variables used in the function include S for subgroup, P for the entire sample, s for size, max and min denote the maximum and minimum values of target property, c is the cutoff of the target property, and θ is the Heaviside function. A Monte Carlo scheme, customized for this task, was utilized to identify subgroups. The procedure initiated with the generation of a specific quantity of trial conjunctions (seeds). Subsequently, the quality function was evaluated for each seed, incorporating the pruning of inequalities. Based on this evaluation, subgroups arising from 6,400,000 seeds that showed high convergence precision were chosen for subsequent analysis. The selection of appropriate primary features is essential, as it underpins the robustness and effectiveness of the descriptors discovered through SGD.[3] Primary features associated with the metal atom and bulk were considered based on previous studies and shown in Table S1.[4] Additional information and specific values for all primary features can be found in the Table S1. Notably, individual elemental features exhibit only weak correlations with CH_3 binding energy (Fig. S4), highlighting the necessity of multivariate subgroup analysis to capture nontrivial structure-property relationships.

DFT Methods

All spin-polarized density functional theory (DFT) calculations were performed using the Vienna ab initio Simulation Package (VASP, version 5.4.4).[5, 6] The Perdew-Burke-Ernzerhof (PBE) exchange–correlation functional within the generalized gradient approximation (GGA) was employed.[7, 8] The interactions between core and valence electrons were described using the projector-augmented wave (PAW) method.[9, 10] A plane-wave kinetic energy cutoff of 520 eV was used throughout all calculations. Periodic slab models were used to represent the catalyst surfaces. Since the generated structures in this work contain vacuum layers larger than 15 Å, this vacuum separation was retained in the DFT calculations to avoid spurious interactions between periodic images along the surface-normal direction. During structural relaxation, the layer-fixing scheme depended on the slab thickness: for slabs containing more than 4 atomic layers, the bottom two layers were fixed, whereas for four-layer slabs, only the bottommost layer was fixed. All remaining atoms and adsorbates were fully relaxed. Brillouin zone sampling was carried out using a Monkhorst-Pack k-point mesh of 2 × 2 × 1.[11] Geometry optimizations were considered converged when the residual forces on each atom were below 0.02 eV $Å^{-1}$ and the total energy change was less than 1×10^{-5} eV.

DFT calculations were performed to validate representative candidate systems identified from the large-scale screening results. In particular, CH_3 adsorption energies were computed for selected systems, and the first methane dissociation energy (ΔE_{diss}) for $CH_4 \rightarrow CH_3^* + H^*$ was further examined. For representative candidates, minimum-energy pathways and activation barriers were determined using the climbing-image nudged elastic band (CI-NEB) method.[12] Transition states were confirmed by vibrational frequency analysis using the finite-difference method with a displacement step of 0.02 Å, where a single imaginary frequency along the reaction coordinate was identified.

Table S1. Reported generation scales of representative prior generative frameworks.

Frameworks	Reported generated structures	Representative reference
CatGPT	10^4	Ref 1
MatterGen	10^4	Ref 13
MAGECS	10^5	Ref 14
CrystalFlow	10^4	Ref 15
CrystaLLM	10^3	Ref 16
CatDRX	10^4	Ref 17
OM-Diff	10^4	Ref 18
This work*	**10^7**	/

Table S2. **Distribution of recovered experimental catalysts over CH_3-binding-energy ranges.**

Binding energy range of CH_3	Count	Representative reference
< -1.5 eV	2	Ref [19, 20]
-1.5~-0.5 eV	26	see main text
> -0.5 eV	2	Ref [21, 22]

Table S3. Considered features of element1, element2, and element3 for alloys.

Abbreviation	Name
R1	Atomic radius of element1 (Å)
EC1	Cohesive energy of element1 (eV)
EP1	Electronegativity potential of element1(/)
Nd1	Number of d valence electrons of element1 (/)
Nds1	Total number of d and s valence electrons of element1 (/)
M1	Relative atomic mass of element1 (a.m.u.)
IP1	Ionization potential of element1 (eV)
EA1	Electron affinity of element1 (eV)
BC1	d- or p- band center of element1 (eV)
R2	Atomic radius of element2 (Å)
EC2	Cohesive energy of element2 (eV)
EP2	Electronegativity potential of element2 (/)
Nd2	Number of d valence electrons of element2 (/)
Nds2	Total number of d and s valence electrons of element2 (/)
M2	Relative atomic mass of element2 (a.m.u.)
IP2	Ionization potential of element2 (eV)
EA2	Electron affinity of element2 (eV)
BC2	d- or p- band center of element2 (eV)
R3	Atomic radius of element3 (Å)
EC3	Cohesive energy of element3 (eV)
EP3	Electronegativity potential of element3 (/)
Nd3	Number of d valence electrons of element3 (/)
Nds3	Total number of d and s valence electrons of element3 (/)

M3	Relative atomic mass of element3 (a.m.u.)
IP3	Ionization potential of element3 (eV)
EA3	Electron affinity of element3 (eV)
BC3	d- or p- band center of element3 (eV)

Table S4. Experimental synthesis precedents for representative candidate alloy systems identified in this work.

System	Experimental synthesis precedent	Representative reference
Co-Ni-V	Yes	Ref 23
Mn-Ti-V	Yes	Ref 24
Ag-Au-Zn	Yes	Ref 25
Al-Ni	Yes	Ref 26
Hf-Pd	Yes	Ref 27
Ag-Al-Pd	Yes	Ref 28
Ag-Cd-Pd	Yes	Ref 29
Al-Au-Sc	Yes	Ref 30
Al-Au-Cu	Yes	Ref 31
Al-Cu-Pd	Yes	Ref 32
Al-Hf-Pd	Yes	Ref 33
Au-Cu-Zn	Yes	Ref 34
Au-Pd-Zn	Yes	Ref 35
Al-Pd	Yes	Ref 36
Ag-Pd	Yes	Ref 37
Ag-Pt	Yes	Ref 38
Cu-Y	Yes	Ref 39
Cu-Zn	Yes	Ref 40

Table S5. Thermodynamic accessibility indicators for representative candidate alloy systems identified in this work.

System	Formation Energy (eV)	E_{hull} (eV)
Ag-Au-Pd	-0.09	0.014
Ag-Cd-Y	-0.28	0.011
Au-Ir-Y	-0.80	0.029
Cu-Sc	-0.26	0.00

Table S6. Lists of MLP-predicted and DFT-calculated CH_3 binding energies for representative screened candidate systems.

System	MLP (eV)	DFT (eV)
Ag-Al-Pd	-0.72	-1.45
Ag-Au-Pd	-0.54	-1.38
Ag-Au-Zn	-0.74	-1.32
Ag-Cd-Pd	-0.74	-1.41
Al-Au-Cu	-1.12	-1.41
Al-Cu-Pd	-0.80	-1.51
Au-Cu-Zn	-1.13	-1.40
Au-Ir-Y	-0.76	-1.20
Au-Pd-Zn	-0.61	-1.36
Al-Pd	-1.04	-1.23
Ag-Pd	-0.82	-0.86
Ag-Pt	-0.99	-0.99

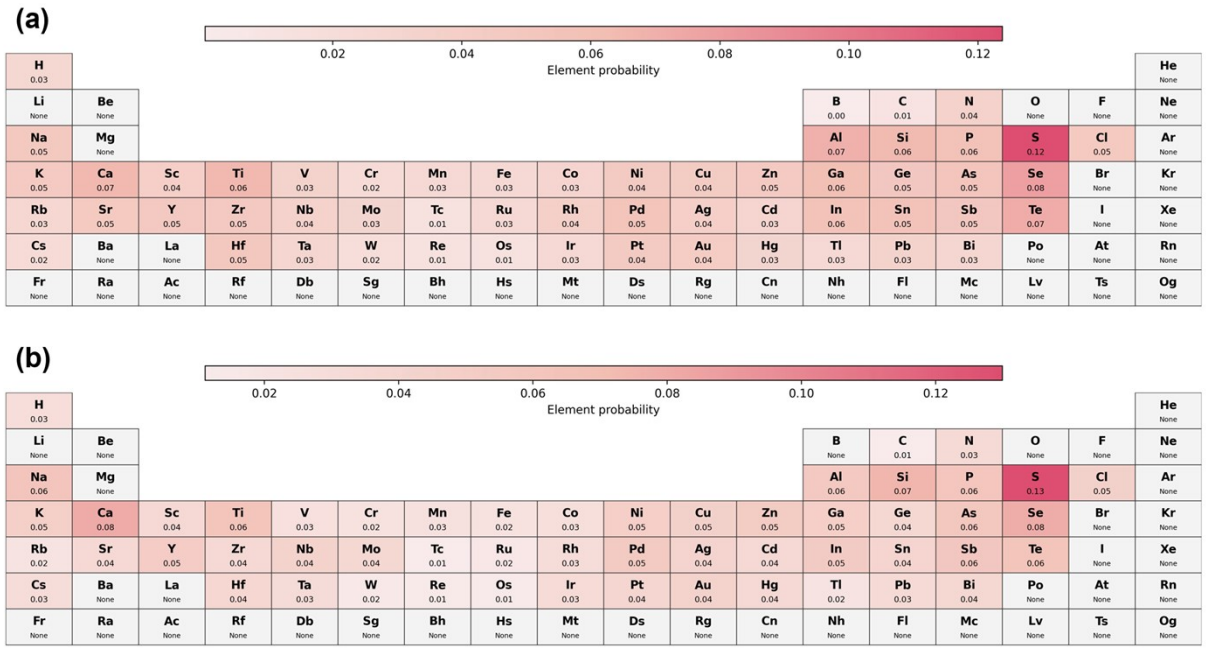

Fig. S1. Elemental composition of the datasets. **(a)** Periodic table heatmap showing the element occurrence frequency in the OC20-S2EF 2M dataset used for pretraining. **(b)** Periodic table heatmap showing the element occurrence frequency in the CH_3-adsorbed subset used for fine-tuning.

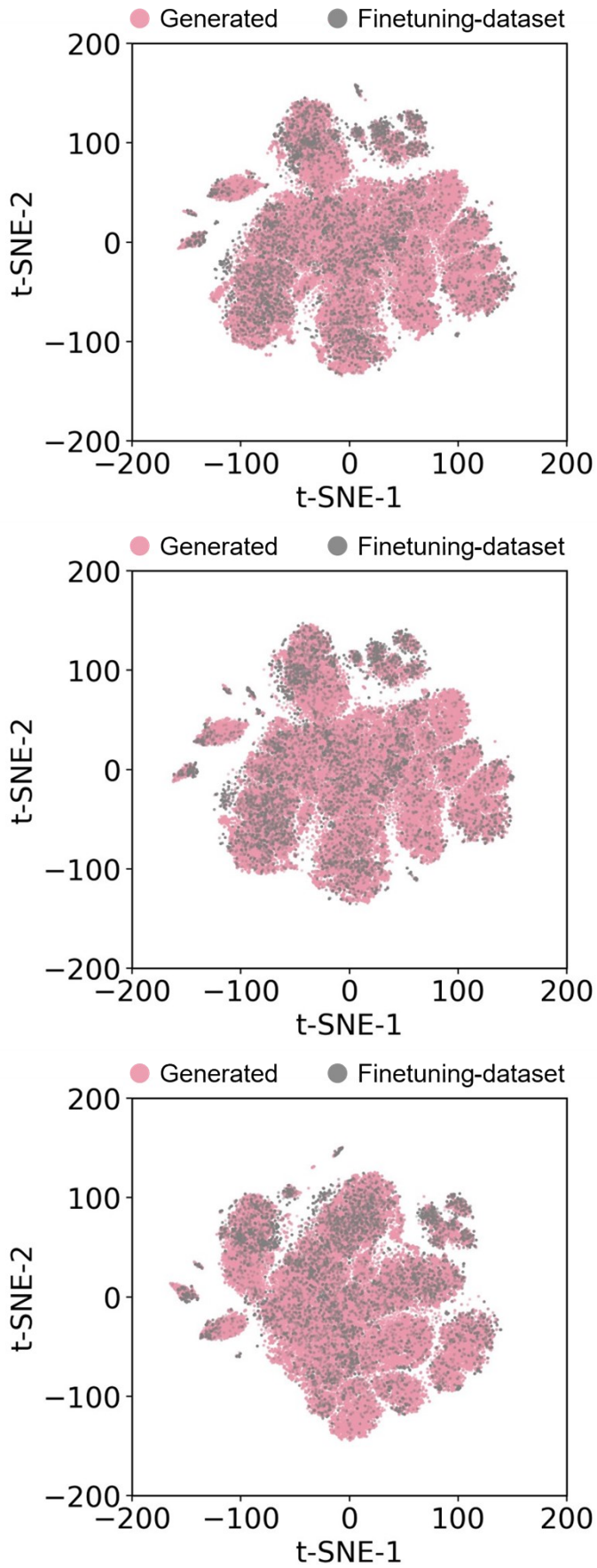

Fig. S2. The t-SNE embeddings of the fine-tuning and generated datasets based on structural descriptors. Three independent random subsets of the generated structures were analyzed to assess the robustness of the observed distribution patterns.

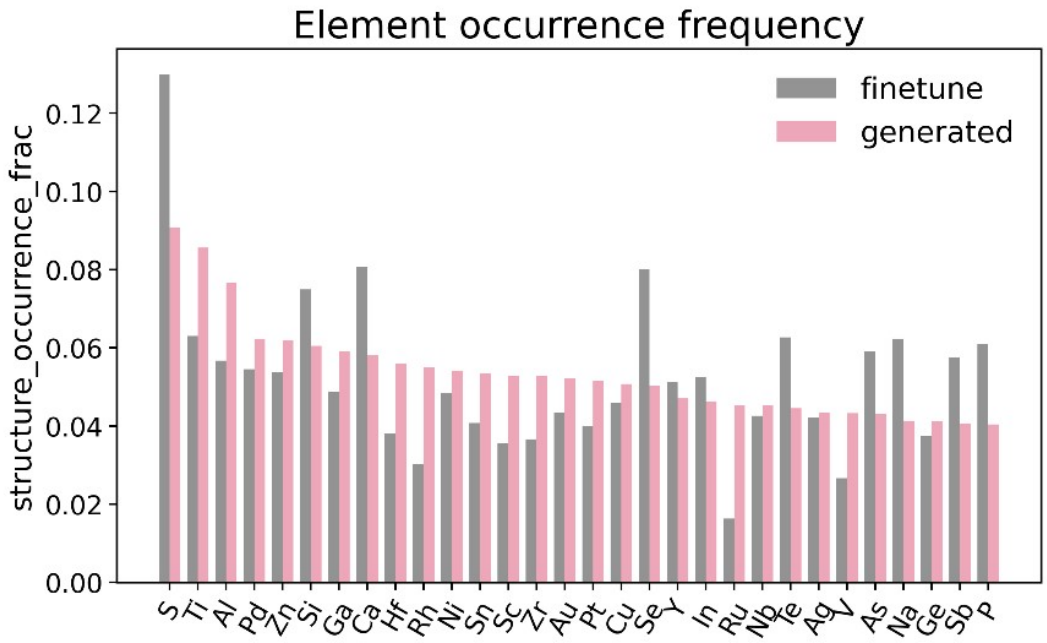

Fig. S3. Compositional comparison between the fine-tuning and generated datasets about the element occurrence frequencies.

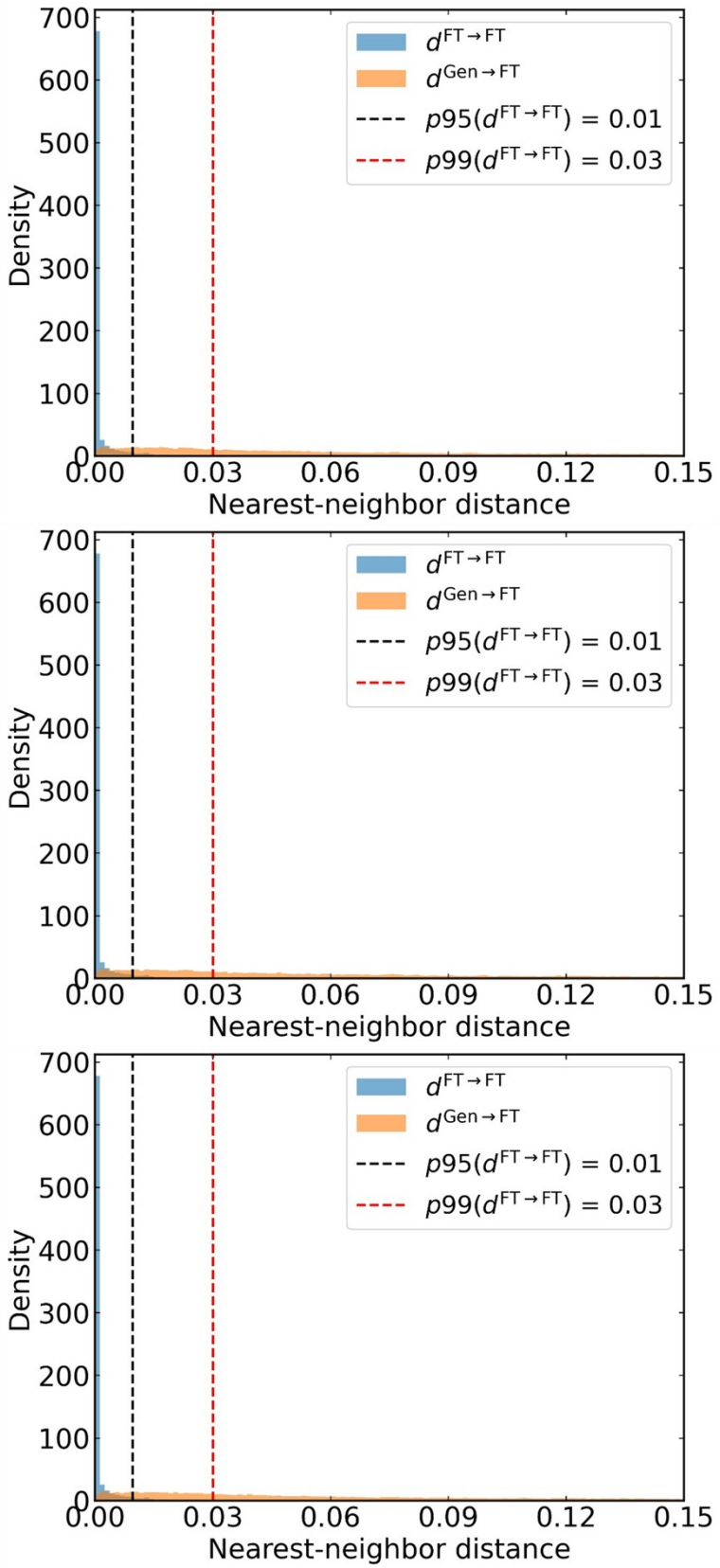

Fig. S4. SOAP nearest-neighbor distance distributions for three independently sampled generated subsets. In each panel, the within-fine-tuning distribution (blue) is compared with the generated-to-fine-tuning distribution (orange); dashed lines mark the p95 and p99 thresholds of the within-fine-tuning distances.

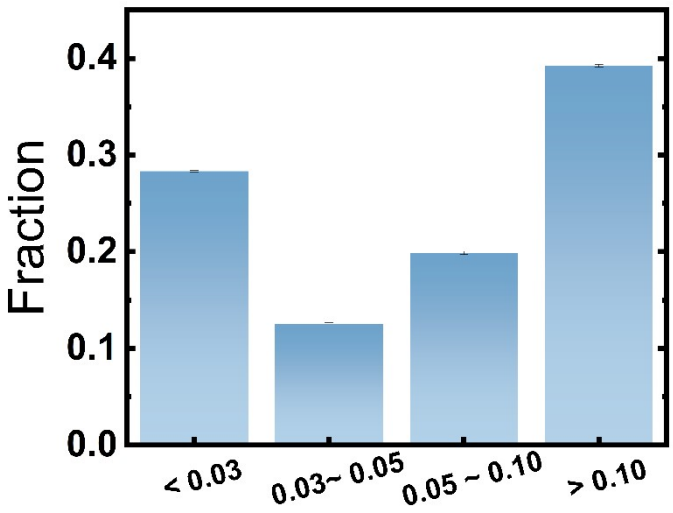

Fig. S5. Fractions of generated local environments in different SOAP-based extrapolation regimes. Error bars indicate the standard deviation across three independently sampled generated subsets.

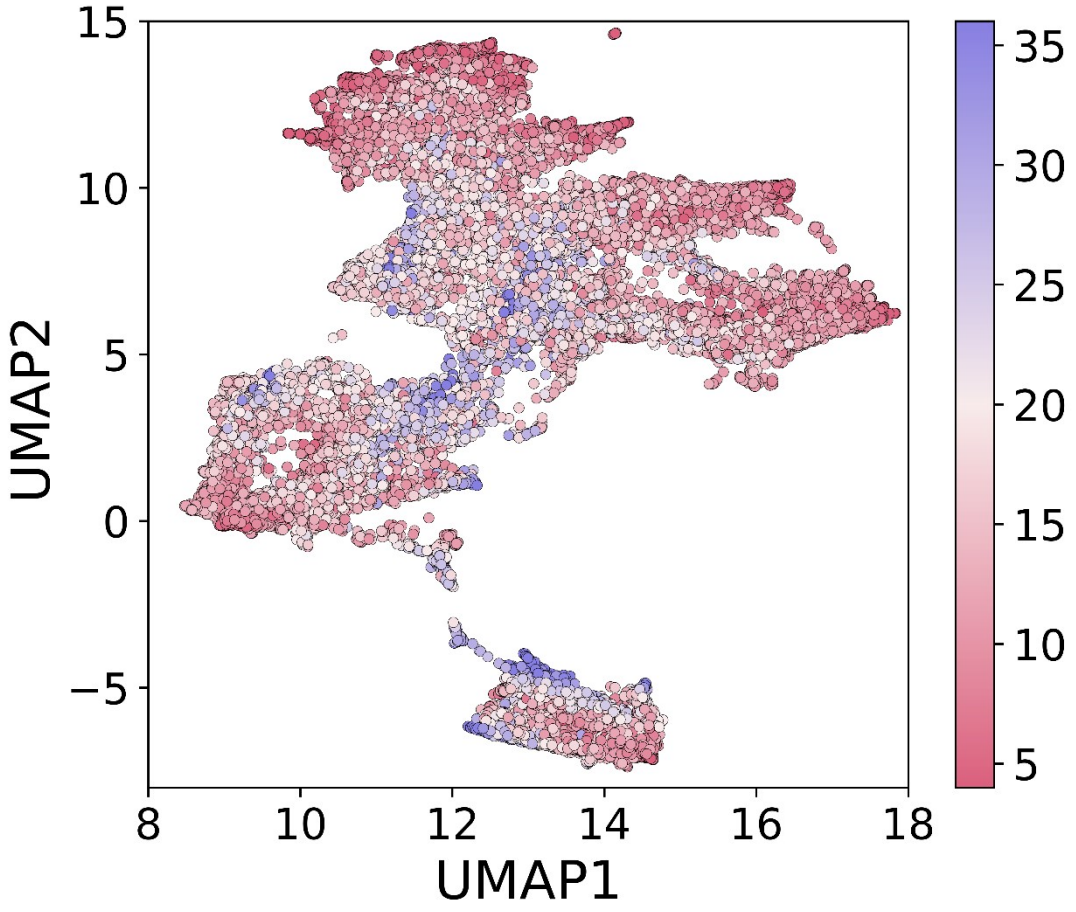

Fig. S6. Two-dimensional UMAP embedding of a randomly sampled subset of generated structures based on local structural descriptors colored by coordination number within 6 Å of the CH3 adsorption site.

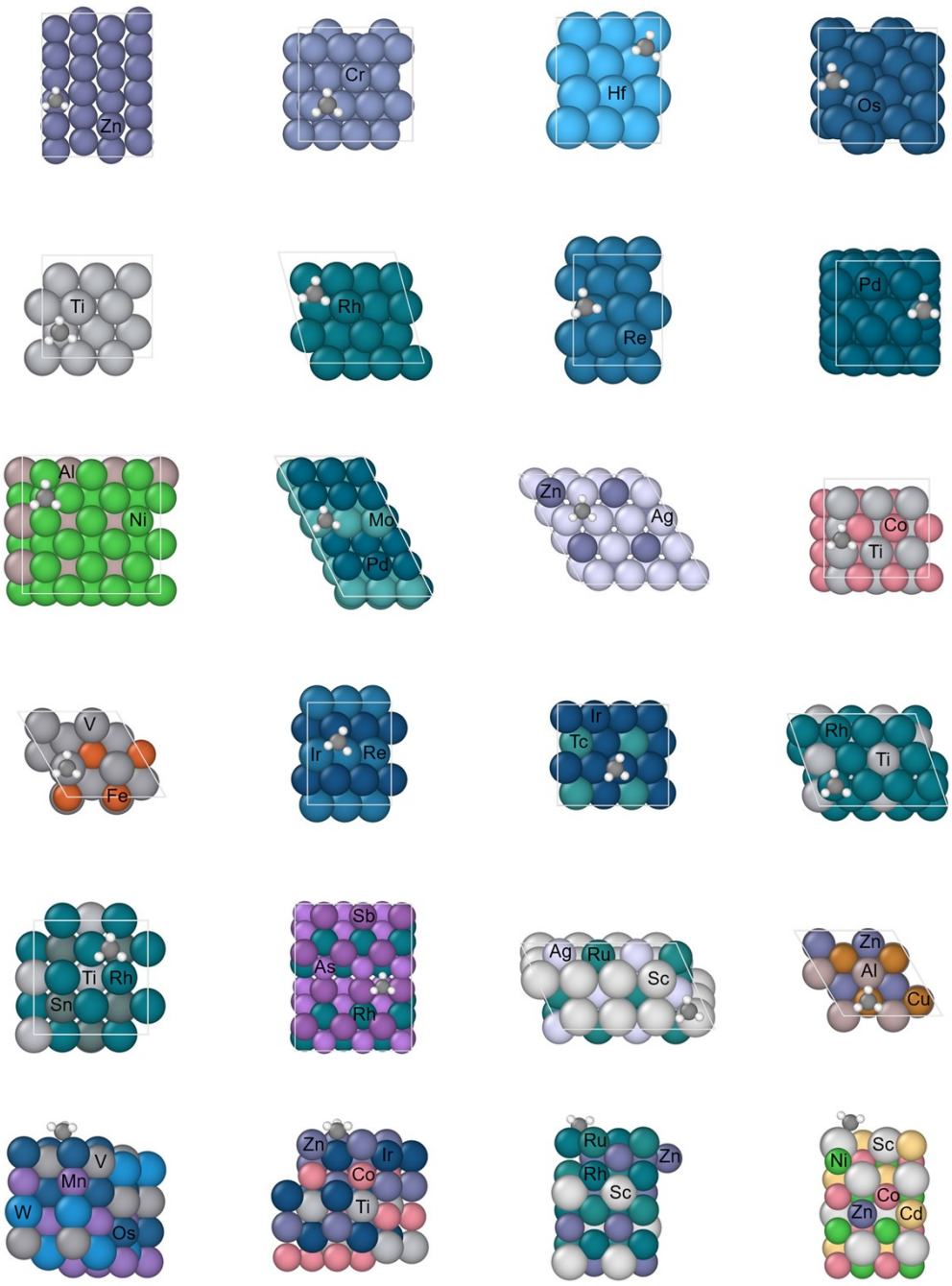

Fig. S7. Representative generated catalyst structures with varying compositional complexity. Representative snapshots of generated catalyst structures ranging from unary to quinary compositions. For clarity, atomic species are explicitly labeled within each individual structure panel, given the large number of elements involved.

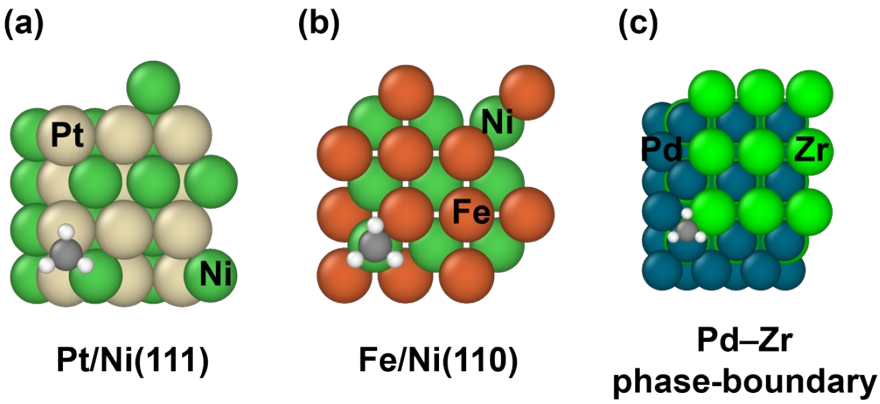

Fig. S8. Representative recovered catalyst systems showing literature-consistent local/surface motifs. **(a)** Ni-Pt with a Pt/Ni (111)-like mixed-metal surface environment around CH_3. **(b)** Fe-Ni with a (110)-type surface/facet motif. **(c)** Pd-Zr with CH_3 located near a heterometal interfacial boundary.

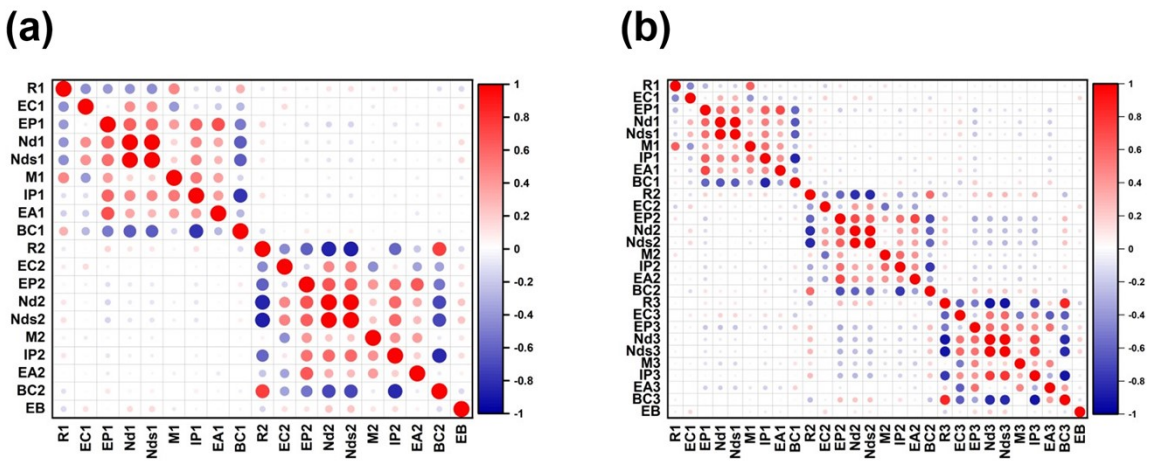

Fig. S9. Correlation heat map of features and properties for **(a)** binary and **(b)** trinary catalysts.

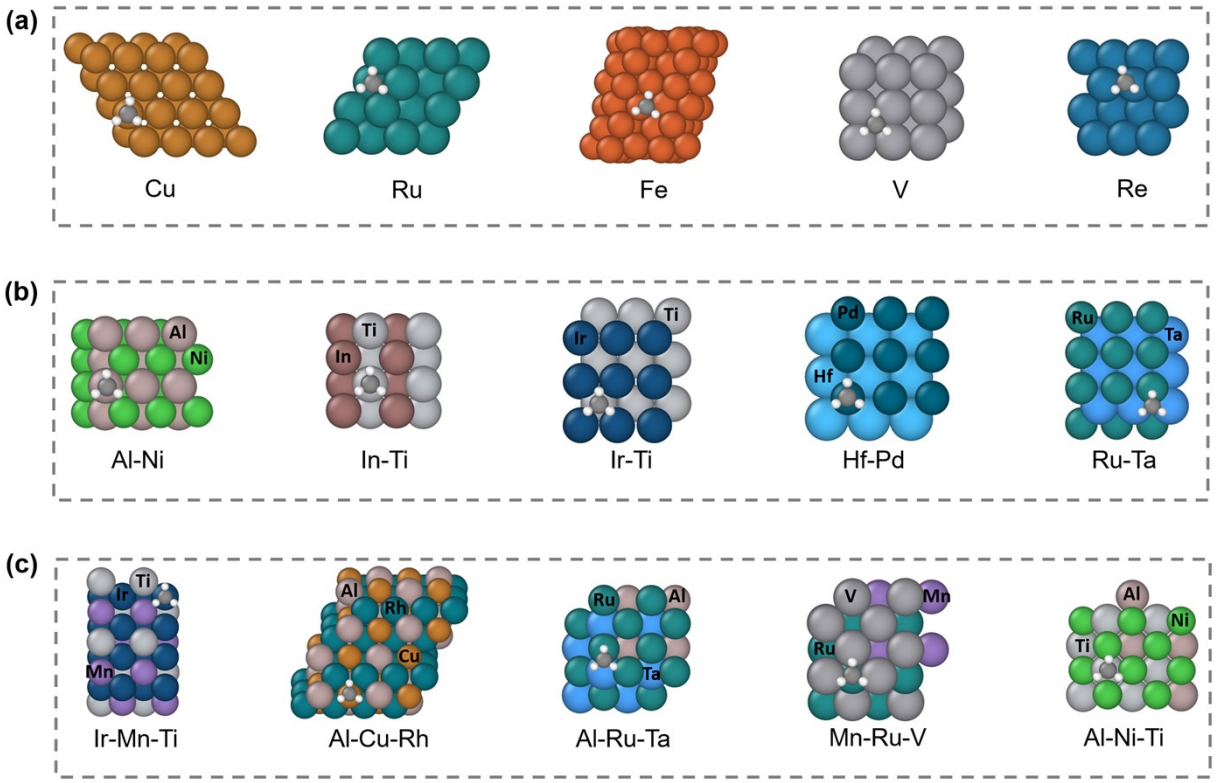

Fig. S10. Representative catalyst structures generated within the selected energy window. Atomic snapshots of representative unary **(a)**, binary **(b)**, and ternary **(c)** catalyst structures selected from the generated dataset. For unary catalysts, the elemental identity can be directly inferred. For binary and ternary catalysts, element labels are explicitly indicated in the snapshots to facilitate identification, as the large number of elements involved makes color-based distinction impractical.

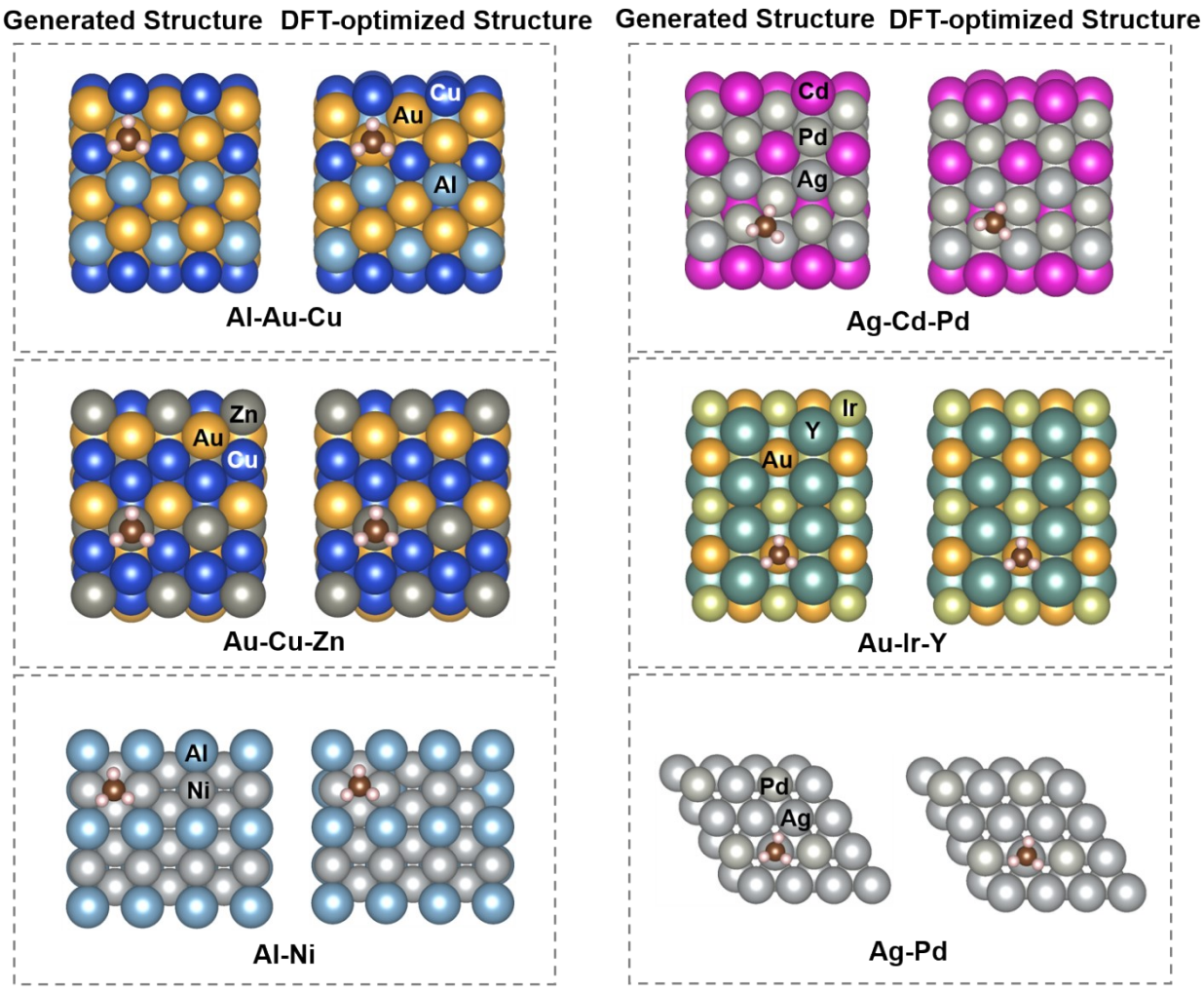

Fig. S11. Representative adsorption structures of generated and DFT-optimized.

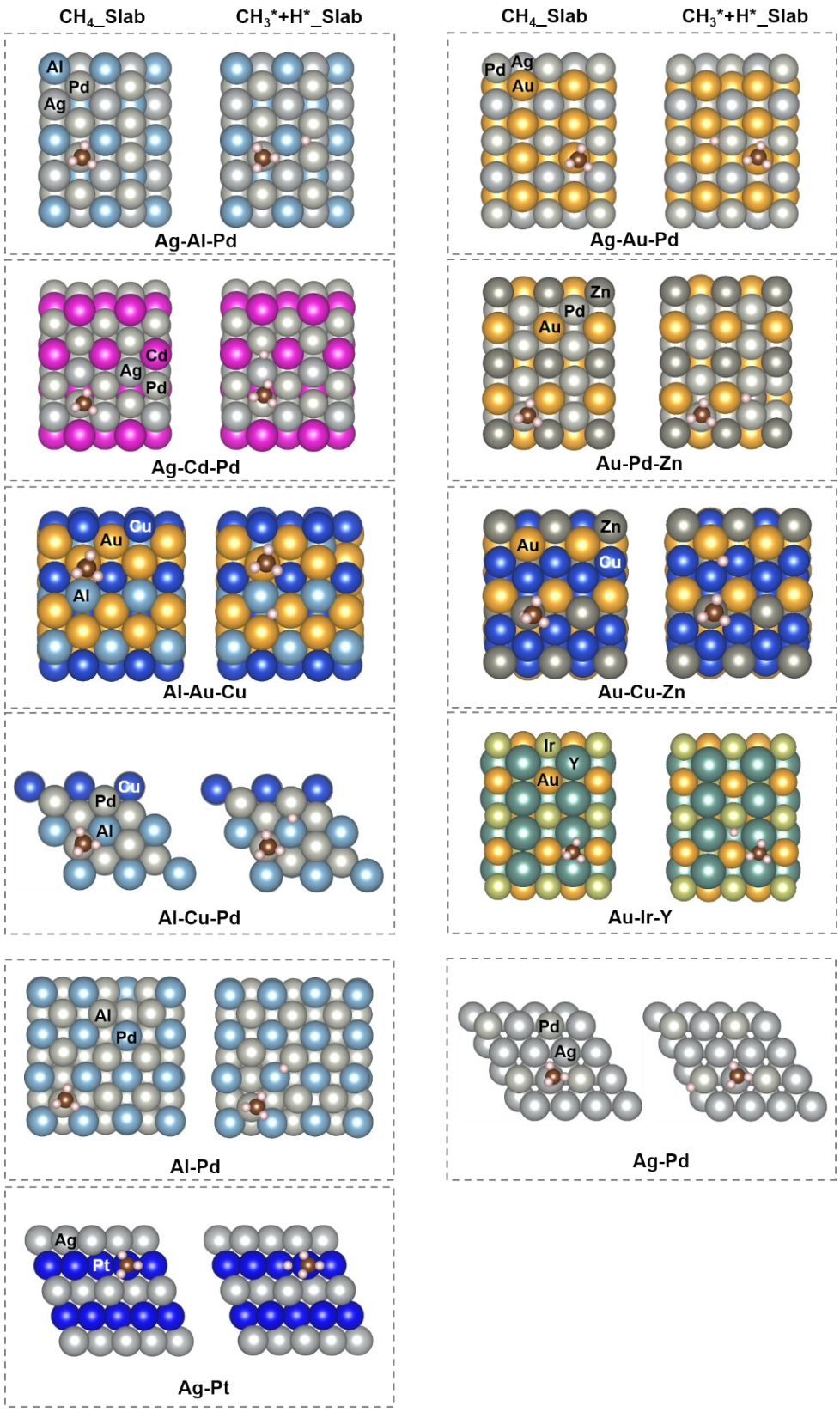

Fig. S12. The optimized initial state, and final state for CH_4 dissociation to CH_3* + H* on representative screened candidates.

Reference

[1] D. H. Mok, S. Back, *J. Am. Chem. Soc.* 2024, **146**, 33712-33722.

[2] A. Mazheika, Y.-G. Wang, R. Valero, F. Viñes, F. Illas, L. M. Ghiringhelli, S. V. Levchenko, M. Scheffler, *Nat. Commun.* 2022, **13,** 419.

[3] B. R. Goldsmith, M. Boley, J. Vreeken, M. Scheffler, L. M. Ghiringhelli, *New J. Phys.,* 2017, **19**, 013031.

[4] Z.-K. Han, D. Sarker, R. Ouyang, A. Mazheika, Y. Gao, S. V. Levchenko, *Nat. Commun.* 2021, **12**, 1833.

[5] G. Kresse, J. Furthmüller, *Comput. Mater. Sci.*, 1996, **6**, 15-50.

[6] G. Kresse, J. Furthmüller, *Phys. Rev. B*, 1996, **54**, 11169-11186.

[7] J. P. Perdew, *Phys. Rev. Lett.*, 2008, **100**, 136406.

[8] S. Grimme, J. Antony, S. Ehrlich and S. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

[9] P. E. Blöchl, *Phys. Rev. B,* 1994, **50**, 17953-17979.

[10] G. Kresse, D. Joubert, *Phys. Rev. B,* 1999, **59**, 1758-1775.

[11] H. J. Monkhorst, J. D. Pack, *Phys. Rev. B,* 1976, **13**, 5188.

[12] G. Henkelman, B. P. Uberuaga,H. J. Jónsson, *Chem. Phys.,* 2000, **113**, 9901-9904.

[13] C. Zeni, R. Pinsler, D. Zügner, A. Fowler, M. Horton, X. Fu, Z. Wang, A. Shysheya, J. Crabbé, S. Ueda, R. Sordillo, L. Sun, J. Smith, B. Nguyen, H. Schulz, S. Lewis, C.-W. Huang, Z. Lu, Y. Zhou, H. Yang, H. Hao, J. Li, C. Yang, W. Li, R. Tomioka, T. Xie, *Nature,* 2025, **639**, 624–632.

[14] Z. Song, L. Fan, S. Lu, C. Ling, Q. Zhou, J. Wang, *Nat. Commun.* 2025, **16,** 1053.

[15] X. Luo, Z. Wang, Q. Wang, J. Lv, L. Wang, Y. Wang, Y. Ma, *Nat. Commun.* 2025, **16**, 9267.

[16] L. M. Antunes, K. T. Butler, R. Grau-Crespo, *Nat. Commun.* 2024, **15**, 10570.

[17] A. Kengkanna, Y. Kikuchi, T. Niwa, M. Ohue, *Commun. Chem.* 2025, **8**, 314.

[18] F. Cornet, B. Benediktsson, B. Hastrup, M. N. Schmidt, A. Bhowmik, *Digit. Discov.* 2024, **3**, 1793–1811.

[19] J. Mao, H. Liu, Y. Li, M. Gao, Y. Zhang, Y. Song, M. Zhang, G. Xu, W. Zhou, L. Yu, X. Cui, D. Deng, *J. Am. Chem. Soc.* 2025, **147**, 14530–14540.

[20] G. Fu, X. Xu, X. Lu, H. Wan, *J. Am. Chem. Soc.* 2005, **127**, 3989–3996.

[21] C. Hammond, M. M. Forde, M. H. Ab Rahim, A. Thetford, Q. He, R. L. Jenkins, N. Dimitratos, J. A. Lopez-Sanchez, N. F. Dummer, D. M. Murphy, A. F. Carley, S. H. Taylor, D. J. Willock, E. E. Stangland, J. Kang, H. Hagen, C. J. Kiely, G. J. Hutchings, *Angew. Chem. Int. Ed.* 2012, **51**, 5129–5133.

[22] C. Palmer, D. Chester Upham, S. Smart, M. J. Gordon, H. Metiu, E. W. McFarland, *Nat. Catal.* 2020, **3**, 83–89.

[23] M. Hu, Q. Cao, X. Wang, D. Zhang, J.-Z. Jiang, *Thin Solid Films* 2021, **734**, 138866.

[24] Z. Dehouche, R. Djaozandry, J. Goyette, T. Bose, *J. Alloys Compd.* 2005, **400**, 276–280.

[25] Z. Shen, J. Zhong, S. Jiang, W. Xie, S. Zhan, K. Lin, L. Zeng, H. Hu, G. Lin, Y. Lin, S. Sun, Z. Shi, *ACS Appl. Mater. Interfaces,* 2022, **14**, 41022–41036.

[26] A. Malakhov, D. Shakhray, I. Denisov, F. Galiev, S. Seropyan, *Materials,* 2022, **15**, 6062.

[27] N. Selhaoui, J.-C. Gachon, J. Hertz, J.-Y. Nazare, J.-C. Joud, J.-P. Bonnelle, X. Ding, *J. Alloys Compd*. 1994, **204**, 69–77.

[28] T.-H. Chuang, S.-W. Hsu, C.-H. Chen, *IEEE Trans. Compon. Packag. Manuf. Technol.* 2020, **10**, 1657–1665.

[29] H. Wen and I. J. van Rooyen, *J. Eur. Ceram. Soc.,* 2017, 37, 3271–3284.

[30] E. C. J. Gießelmann, M. Radzieowski, S. F. Matar, O. Janka, *Inorg. Chem.* 2023, **62**, 9602–9617.

[31] P. J. S. Buenconsejo, A. Ludwig, *Acta Mater*. 2015, **85**, 378–386.

[32] A. B. Y. Lim, X. Long, L. Shen, X. Chen, R. V. Ramanujan, C. L. Gan, Z. Chen, *J. Alloys Compd.*, 2015, **628**, 107–115.

[33] B. Wiendlocha, M. J. Winiarski, M. Muras, C. E. Zvoriste-Walters, J.-C. Griveau, S. Heathman, M. Gazda, T. Klimczuk, *Phys. Rev. B*, 2015, **91**, 024509.

[34] Y. Jia, Y. Ma, W.-E. Yang, Y. Zhu, X. Zhang, S. Xie, T.-W. Kuo, S. Dai, K.-W. Wang, *Chem. Eng. J.*, 2024, **488**, 150946.

[35] M. W. Tew, H. Emerich, J. A. van Bokhoven, *J. Phys. Chem. C* 2011, **115**, 8457–8465.

[36] M. Khoudiakov, A. B. Ellis, J. H. Perepezko, S. Kim, K. D. Kepler, *Chem. Mater.,* 2000, **12**, 2008–2013.

[37] X. Wen, S. A. Nazemi, R. Rosa da Silva, K. Moth-Poulsen, *Langmuir,* 2023, **39**, 11268–11273.

[38] X. Yang, L. T. Roling, M. Vara, A. O. Elnabawy, M. Zhao, Z. D. Hood, S. Bao, M. Mavrikakis, Y. Xia, *Nano Lett.,* 2016, **16**, 6644–6649.

[39] S. Y. Yang, X. Liu, Q. Yao, J. Wang, Y. Du, L. Zhang, G. Rao, H. Zhou, *J. Phase Equilib. Diffus.*, 2025, **46**, 133–147.

[40] K. Schütte, H. Meyer, C. Gemel, J. Barthel, R. A. Fischer, C. Janiak, *Nanoscale* 2014, **6**, 3116–3126.