# Data-driven massive reaction networks reveal new pathways underlying catalytic $CO_2$ hydrogenation

Anand M. Verma,[†,‡,@] Shivam Chaturvedi,[†,@] Swastik Paul,[†,@] Srinibas Nandi,[†,¶] Rahul Sheshanarayana,[†,§] Kotni Santhosh,[‖] G Valavarasu,[‖] Ambedkar Dukkipati,[⊥] Chuandayani Gunawan Gwie,[#] Pei Ying Moo,[#] Chun Qi Joy Ng,[#] Amol Amrute,[#] and Ananth Govind Rajan[*,†]

†*Department of Chemical Engineering, Indian Institute of Science, Bengaluru, Karnataka 560012, India*

‡*Department of Chemical Engineering, Motilal Nehru National Institute of Technology Allahabad, Prayagraj, Uttar Pradesh 211004, India*

¶*Department of Computational and Data Sciences, Indian Institute of Science, Bengaluru, Karnataka 560012, India*

§*Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, New York 14853, United States*

‖*HP Green R&D Centre, Hindustan Petroleum Corporation Limited, Bengaluru, India*

⊥*Department of Computer Science and Automation, Indian Institute of Science, Bengaluru, Karnataka 560012, India*

#*Institute of Sustainability for Chemicals, Energy and Environment (ISCE$^2$), Agency for Science, Technology and Research (A*STAR), 1 Pesek Road, Jurong Island, Singapore 627833, Republic of Singapore*

@*Contributed equally to this work*

**Abstract**

Heterogeneous catalytic pathways for clean energy conversion involve thousands of elementary steps, but most quantum-mechanical models involve only a few dozen reactions. We combine extensive density functional theory (DFT) calculations, machine learning (ML) for activation barrier prediction, and human intelligence-inspired reaction enumeration and elementary reaction identification. This enables automated kinetic modeling of $CO_2$ hydrogenation on copper, a key process to produce fuels and

chemicals. We construct the largest dataset of 152 elementary $CO_2$ reduction reactions and experimentally determine $CO_2$ conversion, finding that even large networks with 100+ reactions are insufficient. In contrast, our approach reveals 9389 elementary reactions, reducing human bias in the reaction pathway. We unravel 40-fold higher $CO_2$ conversion rates, following experimental trends of methanol and CO production. We establish the crucial role of intermolecular hydrogen transfer and hydrogenation by molecular hydrogen, a surprising ML-enabled discovery validated post-facto. The proposed strategy to comprehensively model complex catalytic mechanisms will significantly advance catalysis research and carbon conversion processes.

# Introduction

Catalytic reaction pathways are invariably complex, involving thousands of elementary steps interconnecting hundreds of reaction intermediates.[1–3] However, computational studies often investigate only a few dozen elementary reactions and aim to explain experimental trends such as catalytic activity, yield, and selectivity.[3,4] The major factor necessitating the consideration of a limited number of reactions is that it is infeasible to simulate thousands of reactions even on state-of-the-art supercomputing facilities. Indeed, locating the transition state for a reaction is very expensive, requiring ≈10k CPU core hours.[5] Considering a network with 10,000 reactions, it would take ≈ 47 years to compute all activation barriers with 240 CPU cores. Although previous studies on large reaction networks used scaling relations for activation energies or relied on reaction energies,[2,6] developing frameworks that can correctly predict the outcomes of massive ($>$ 1k) reaction networks is an unsolved challenge. In addition to the challenges of accurately predicting activation barriers, there are no generalizable approaches available to accurately identify plausible elementary steps in a reaction mechanism apart from database-based mechanism generators.[7–9] Moreover, no studies have demonstrated the kinetic modeling of large reaction networks or compared their predictions carefully with experimental data.

In this regard, machine learning (ML) in conjunction with automation and high-throughput computation strategies can accelerate catalyst and mechanism discovery[10] for complex processes such as carbon dioxide ($CO_2$) conversion. Here, we formulate a multi-faceted, end-to-end, and data-driven strategy to tackle the chemical complexity in reaction mechanisms via ML and mechanism discovery. Copper (Cu) is an effective catalyst that hydrogenates $CO_2$ to hydrocarbons such as methanol and methane,[11–13] when combined with metals such as Fe,[14,15] demonstrating the potential to enable C-C coupling reactions that lead to $>C_2$-products.[16] We comprehensively investigate thermocatalytic $CO_2$ hydrogenation targeting several possible $C_1$-$C_4$ hydrocarbons/alcohols (e.g., methane, methanol, ethane, ethylene, acetylene, ethanol, propionaldehyde, propanol, butanol, etc.) on Cu(111) (Figure 1a) via extensive quantum-mechanical simulations. A subset of the manually drawn reactions studied using density functional theory (DFT) leading to only $C_1$ and $C_2$ products from $CO_2$ reduction can be seen in Figure 1b, and some exemplary transition states are shown in Figure 1c. The DFT calculations and reaction network for $C_3$ and $C_4$ products can be seen in Section S1. Overall, we examined 152 elementary reactions via DFT, which is significantly more comprehensive compared to previous computational studies and forms the largest database of activation barriers for elementary $CO_2$ hydrogenation reactions till date.[3,4,17–19]

We considered several key classes of reactions, e.g., 109 hydrogenation/dehydrogenation reactions, 11 C-C coupling, and 32 oxygenation/deoxygenation/dehydroxylation reactions. In reality, these numbers would be much higher, along with the possibility of other classes of reactions, e.g., isomerization and oxygen/hydrogen hopping. This is revealed by the automated reaction enumerator and elementary reaction identifier developed herein. For instance, $CO_2$ to $C_1$ products (Figure 1a) involves 386 surface reactions with 25 intermediates. Including both $C_1$ and $C_2$ products, this number increases to 1622 with 50 intermediates and would further grow exponentially upon including $C_3$-$C_4$ products. To make such a massive network computationally tractable, we develop a physics-inspired ML model to directly predict activation (free) energies of elementary reactions based on reaction energies,

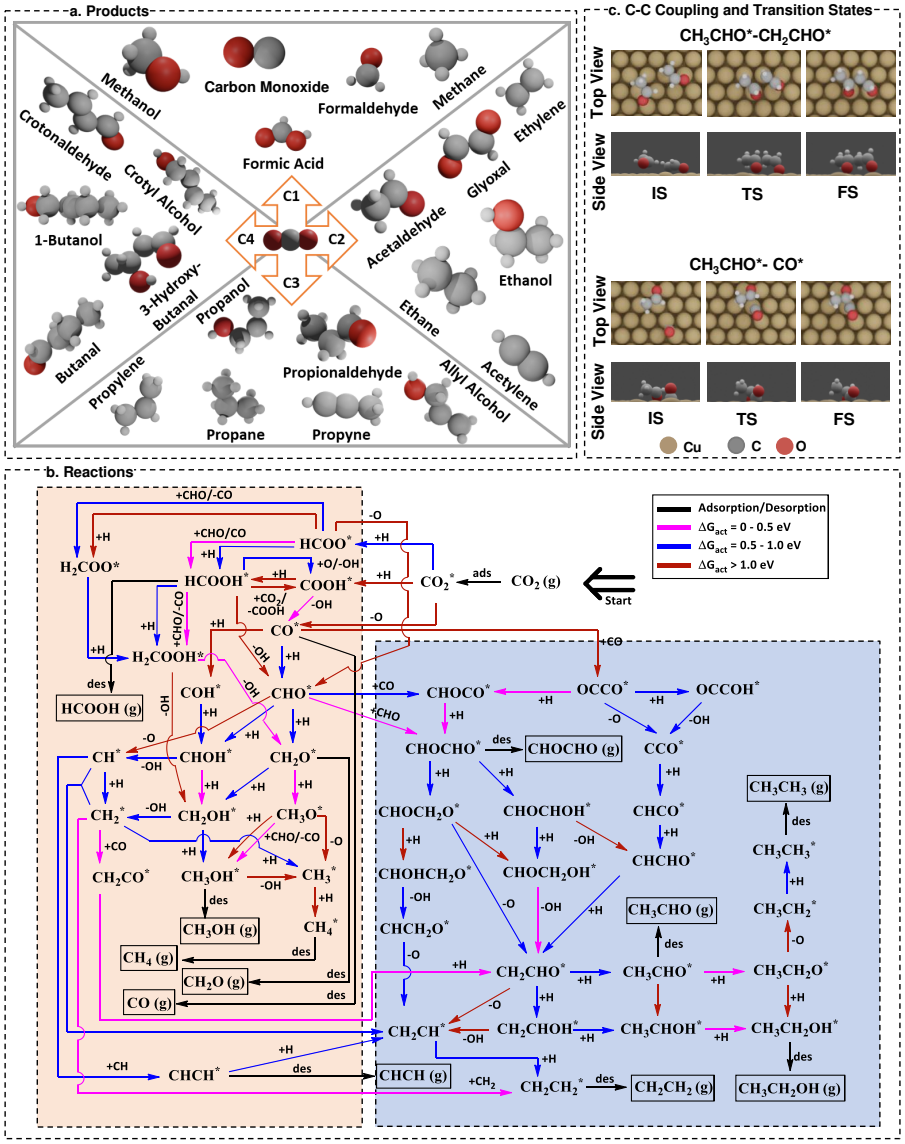


Figure 1: Complexity of the $CO_2$ hydrogenation mechanism considered in this work via comprehensive DFT calculations. (a) All $C_1$-$C_4$ products from $CO_2$ hydrogenation modelled in this work. (b) A manually curated reaction network of $CO_2$ hydrogenation, producing all aforementioned $C_1$ (orangish background) and $C_2$ products (bluish background). The reaction starts from the leftwards double arrow with the label "Start". The products are highlighted in squared boxes. The free energies of activation, computed at 298.15 K, are colour-coded and can be seen with the description given in the legend. The abbreviations "ads" and "des" stand for adsorption and desorption steps, respectively, and all such steps are depicted with black arrows. (c) Exemplary $CH_3CHO$*-$CH_2CHO$* and $CH_3CHO$*-CO* coupling steps with their transition states.

molecular fingerprints, and reaction features (Figure 2a-b). Indeed, although neural network potentials based on ML are being explored for structural optimization, they are currently not accurate enough to predict transition states for arbitrary reactions without quantum chemistry assistance,[20] requiring more physically grounded approaches. In addition, we develop an optimization framework that enumerates all possible surface reactions based on the number and types of adsorbates given (Figure 2c). Moreover, we advance a molecular similarity-based algorithm for elementary reaction screening, mimicking human reasoning (Figure 2d). Finally, we employ automated microkinetic modeling (MKM) to enable the accurate study of thousands of surface chemical reactions involved in thermocatalytic carbon dioxide hydrogenation on Cu(111) (Figure 2f). We compare our predictions with experimental measurements, demonstrating the criticality of our approach in predicting reaction outcomes, such as production rates and selectivities.

# Results

## Creation of the DFT database of reaction and activation free energies

We explored numerous $C_1$-$C_4$ products from $CO_2$ via various reaction pathways (Figure 1b, Section S1), with many explored here for the first time. Our thermodynamic and kinetic analyses support the formation of HCOO* (formate) compared to COOH* (carboxylate). We found that it is more favourable to form HCOOH* (adsorbed formic acid) than $H_2COO$* via direct hydrogenation. Interestingly, we determined that the cross-species reactions of HCOO* and HCOOH* with CHO* have low barriers, which encourages exploration of more such reactions. Our simulations predict that $H_2COOH$* would undergo dehydroxylation to directly form $CH_2O$* (adsorbed formaldehyde). Regarding the formation of CO*, our DFT analysis indicates the dehydrogenation of CHO* that forms upon dehydroxylation of HCOOH* to be the likely pathway. CO* hydrogenates more favourably to CHO* (than COH*), which is an important intermediate in the formation of various $C_1$-based prod-

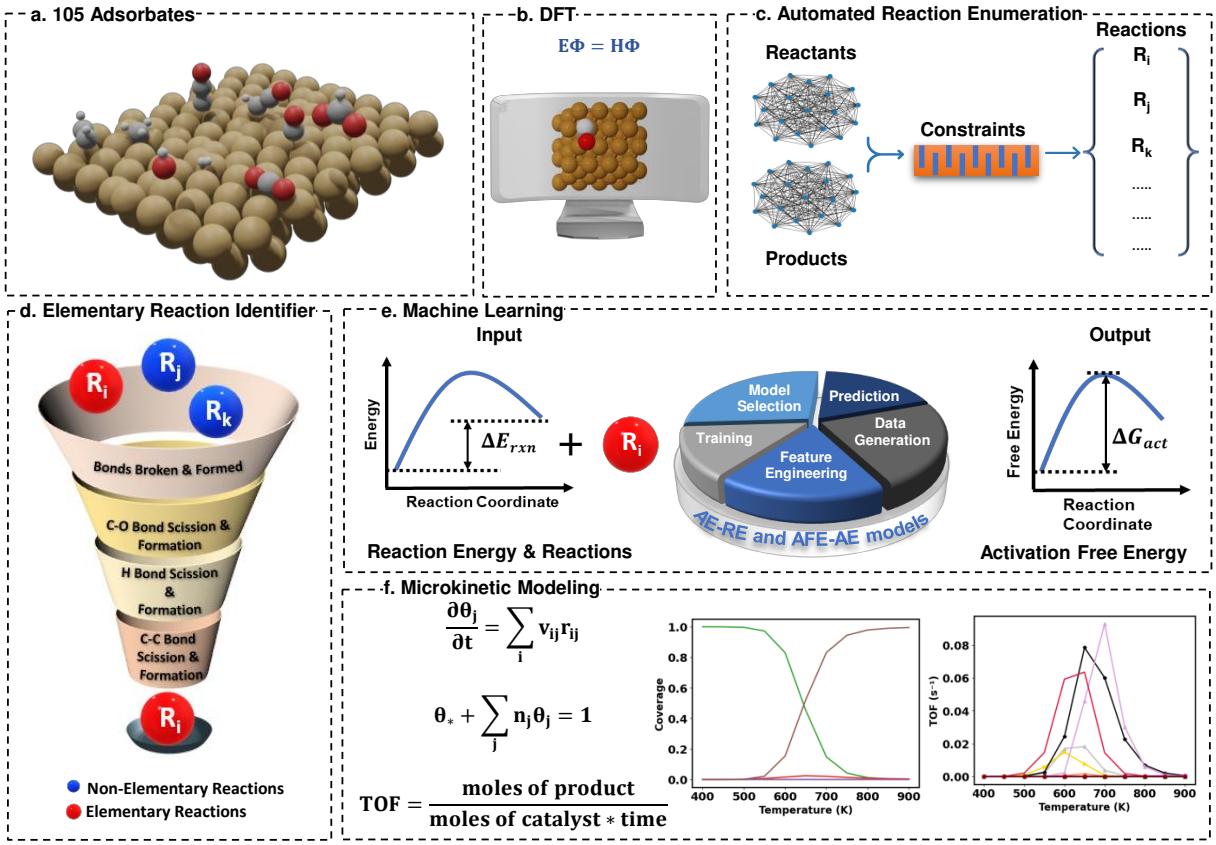


Figure 2: Automation workflow involving DFT, automated reaction enumeration, elementary reaction identification, ML, and MKM. (a) Generation of different adsorbates on the Cu(111) catalyst surface. (b) Geometry optimization of adsorbates on the catalyst surface and locating the transition-state structures for all the listed reactions using DFT. (c) Automated reaction enumeration using adsorbate structural information. (d) Identification of the elementary reactions from a pool of all possible reactions suggested by automated reaction enumeration. (e) Development of two different ML models to predict i. activation energy from reaction energy, and ii. activation free energy from activation energy. (f) MKM based on all possible elementary reactions suggested by the elementary reaction identification framework. The typical MKM workflow based on DFT-derived reactions involves steps a-b-f, while the newly proposed, automated MKM workflow based on DFT, automated reaction enumeration, elementary reaction identification, and ML involves steps a-b-c-d-e-f.

ucts via CH*, CHOH*, and $CH_2O$* (orangish background section in Figure 1b). Moreover, CHO* can favourably undergo self-condensation to produce CHOCHO* (adsorbed glyoxal), the first signature of a $C_2$ hydrocarbon. CHOCHO* acts as a building block for longer-chain molecules, hydrogenating to form a vinyloxy species ($CH_2CHO$*), which is a precursor to acetaldehyde ($CH_3CHO$). Vinyloxy follows sequential hydrogenation to yield ethanol (Section S1). We predict that ethylene ($CH_2CH_2$) could be formed via hydrogenation of CHCH*, which is formed by self-condensation of CH*. For the formation of $C_3$ and $C_4$ products, $CH_3CHO$ is the key intermediate as in electrochemical pathways.[18] $C_3$ products form by the carbonylation of acetaldehyde to $CH_3CHOCO$* (Figure 1c), leading to the synthesis of various $C_3$-products such as propanol, propane, propylene, propionaldehyde, allyl alcohol, and propyne. The activation free energies in the favourable pathways for most $C_3$-products, except for non-oxygenates, remain below 1 eV (Section S1). The formation of $C_4$ products can occur via the cross-condensation of $CH_3CHO$* and $CH_2CHO$* (Figure 1c), with a lower kinetic barrier than the formation of a $C_3$-backbone. Moving forward, the adsorbed $C_4$-backbone, i.e., $CH_3CHOCH_2CHO$*, undergoes hydrogenation to produce 3-hydroxybutanal ($CH_3CHOHCH_2CHO$). This molecule can favourably undergo dehydrogenation and dehydroxylation to produce crotonaldehyde ($CH_3CHCHCHO$), which can hydrogenate to produce butanal ($CH_3CH_2CH_2CHO$) and crotyl alcohol ($CH_3CHCHCH_2OH$). Both butanal and crotyl alcohol can be hydrogenated favourably to produce butanol. Overall, the proposed reaction network is the most comprehensive one for $C_1$ to $C_4$ pathways so far and forms the basis for our detailed study.

## Machine learning (ML) of activation (free) energies

Previously, the Brønsted–Evans–Polanyi (BEP) relationship has been utilized to correlate activation and reaction energies. However, this scaling relationship often fails when considering different types of bond scission/formation. For our dataset, the BEP relation performs poorly with low goodness of fit ($R^2$) and mean absolute error (MAE) of 0.48 and 0.22 eV,

respectively, when fitted over the complete dataset. Recently, some research groups have developed ML models for activation energies,[5,21–23] trained using data from the literature or open-source databases like Catalysis-Hub,[24] RMG-Cat,[25] Materials Project,[26] and the ChemCatBio Database, bringing non-uniformity in the dataset due to different DFT functionals, parameters, and structures. We use our comprehensive in-house DFT data of 304 activation barriers (including both directions of reactions) to develop ML models for the prediction of activation energy from reaction energy (AE-RE) and activation free energy from activation energy (AFE-AE) with a high degree of reliability and accuracy. Figure 3a shows the ML workflow, from getting the coordinate files from the DFT calculations to predicting activation free energies and carrying out MKM.

The AE-RE ML model predicts activation energy from reaction energy and various reaction/molecular features and Morgan fingerprints (Section S2). We mainly used decision tree-based models for the AE-RE ML model due to their interpretability and robustness. We compared several regression models, whose MAE and $R^2$ for both training and test sets after cross validation and hyperparameter tuning (Section S3) are illustrated in radial bar graphs in Figure 3b. CatBoost offered similar MAE and $R^2$ on the training set as the gradient boosting model but on the test set, CatBoost outperformed all other models with an MAE of 0.13 eV and $R^2$ value of 0.82 (parity plot in Figure 3e). This led us to choose the CatBoost algorithm for the AE-RE ML model. We examined the importance of the various features in the AE-RE ML model using SHAP (SHapley Additive exPlanations) values.[27] Figure 3c illustrates the essential features and Figure 3d depicts the visual representation of the important bits according to SHAP values, in which the reaction energy emerged to be the most important one for activation energy prediction. However, other features such as C-H, C-OH, and CO-H bonds also are essential. The AFE-AE model was trained using activation energy, temperature, molecular and reaction features, via a ridge regressor. Figure 3e depicts the parity plots for both ML algorithms used in this study. Clearly, the ridge regressor works well, as evidenced by the high $R^2$ values and low MAE values (0.06 eV for both training and

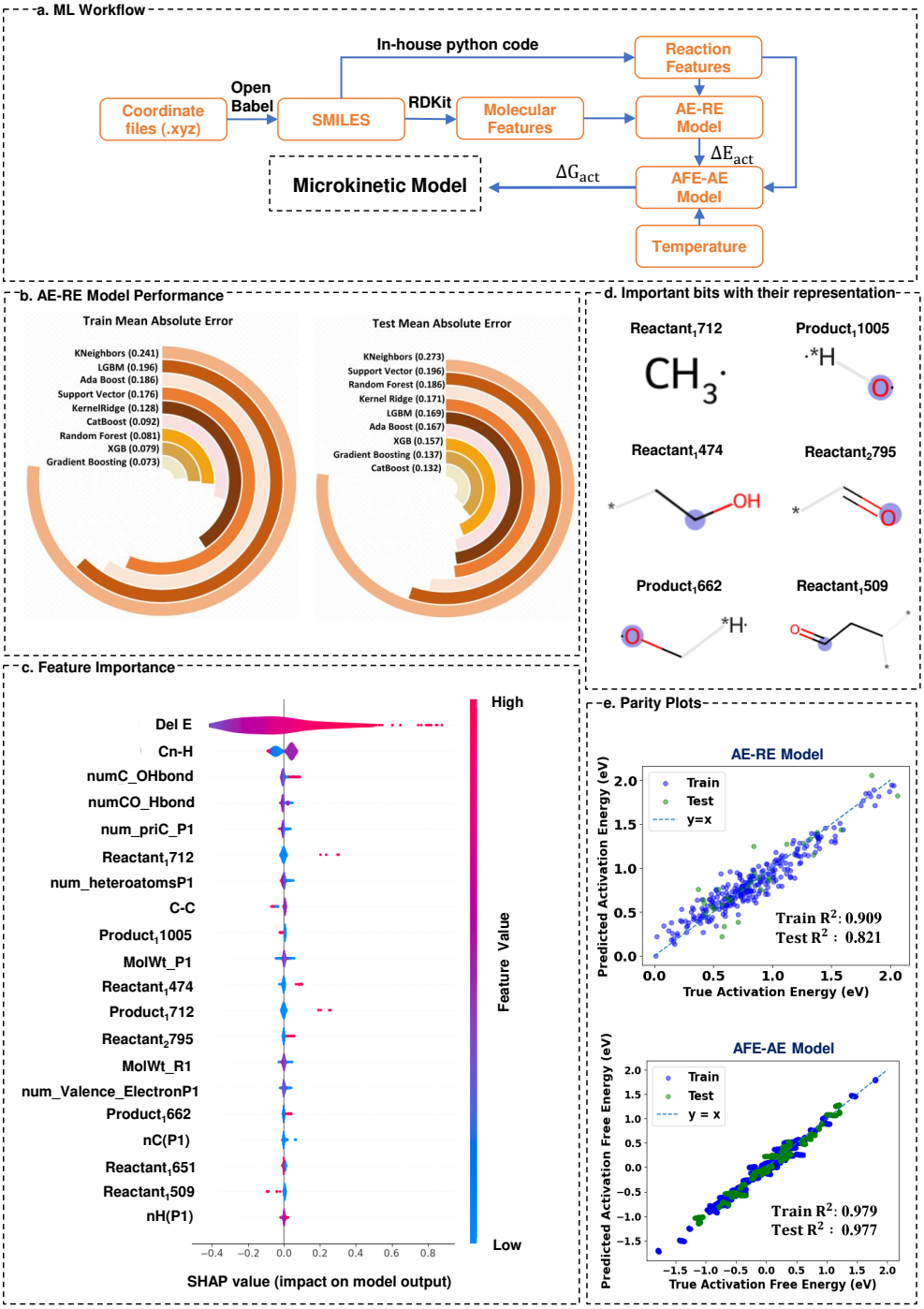


Figure 3: Machine learning models for the predictions of activation energy and activation free energy. (a) overall machine learning workflow starting from .xyz files to provide input for the MKM Model. (b) AE-RE model performance on train and test sets. (c) Feature importance analysis for the AE-RE model calculated using SHAP. (d) Morgan fingerprint bits of high importance in the AE-RE model and their representations. (e) Parity plots for both models, namely AE-RE and AFE-AE ML models.

test sets). This model was five-fold cross-validated to evaluate its robustness (Section S3).

## Automated reaction enumeration and elementary reaction identification

Our exhaustive reaction enumeration framework is realized by an integer linear programming framework, which determines all possible reactions that may occur on the catalyst surface. It requires the representation of adsorbed species in SMILES (simplified molecular-input line-entry system) form and their corresponding DFT energies and uses physical constraints such as mass balances, participation of two species in a reaction, non-existence of a species among both reactants and products, and at most two species among reactants and products. A detailed formulation can be found in the Methods (Section S4) and Section S5. Upon supplying all the SMILES strings of 105 adsorbates and their DFT energies, a total of 104723 reactions (including both elementary and non-elementary reactions) was obtained.

Identifying elementary reactions, i.e., those with a single transition state, is a crucial step before MKM. We leveraged molecular similarity metrics to decide on the elementarity of a reaction in an automated manner by determining which species among the reactants and products to compare to obtain the number of bonds formed/broken, mimicking human intelligence. The elementary reaction identification framework is implemented via a Python code assisted by the RDKit module and checks that the number of bonds (C-H, C-O, and C-C) being broken plus those being formed does not exceed two for an elementary reaction. Figure 4a depicts the workflow of the elementary reaction identification framework and the major rules associated with it (see also Methods in Section S4, and Section S6). We validated our framework with a set of 388 randomly selected and manually labeled reactions, making sure that it contained all types of intermediates ($C_1$, $C_2$, $C_3$, and $C_4$). We found that our framework correctly classifies each reaction, achieving 100 % accuracy on the test set of reactions. Using this approach, the automatically enumerated reactions were narrowed down to 9237 elementary reactions (excluding the 152 manually enumerated reactions), consisting

of 8088 hydrogenation/ dehydrogenation, 117 C-C coupling/ decoupling, 980 oxygenation deoxygenation/ dehydroxylation, and 52 isomerization reactions. The comparison between the number of reactions constructed manually and generated via the automated framework is depicted in Figure 4b, demonstrating the importance of our approach to avoid human bias towards certain classes of reactions.

## Comparing MKM of manually curated (152) and automated (9k) reactions

We first compared the microkinetic results of the 152-reaction network using DFT- and ML-based barriers, demonstrating an excellent match and providing confidence in the ML model (Section S7). We then proceeded to compare two MKM models: the first based on our database of 152 manually curated reactions with DFT-derived free energetics and the second based on 9389 automatically generated elementary reactions with ML-derived free energetics including the above 152 reactions. We employed gas-phase corrections[28–31] for CO to obtain the correct adsorption energy on Cu(111)[32] (Section S8), whereas $CO_2$ required no correction. A depiction of the production rates from both MKM models at varying temperatures and a fixed pressure of 50 bar is given in Figure 4c. The manual MKM, although state-of-the-art as compared to previous studies due to its extensive nature, predicted formic acid as one of the major products, as opposed to methanol, which is seen in experiments (Figure 5; see also refs.[4,33,34]). Moreover, methanol's production rate was $O(10^{-6})$ mole $m^{-3}$ $s^{-1}$ (Figure 4c). In contrast, for the automated MKM model based on the extensive set of reactions, the production rates and product distribution closely followed experimental trends (Figure 5). The automated implementation increased production rates by more than an order of magnitude and led to the formation of more methanol than formic acid, in agreement with experiments. Further, the methanol production rate decreased with increasing temperature, which is also seen in experiments. Note that the objective of this study is not to quantitatively match experimental results via fitting parameters but to match

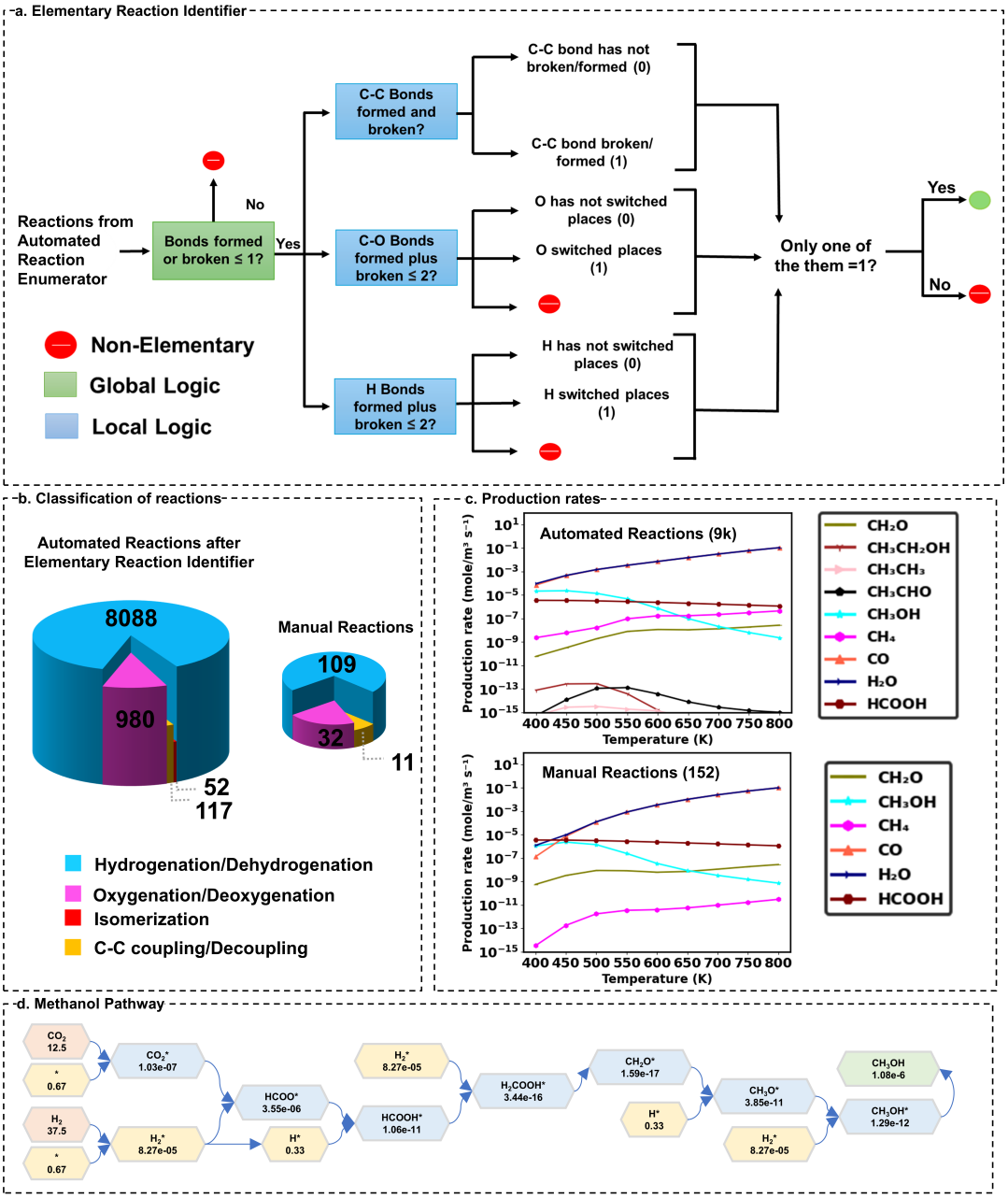


Figure 4: Contrast between the elementary reactions enumerated manually and automatically, and the automated and manual MKM predictions. (a) Elementary reaction identification workflow consists of one global logical check (in green) and three local logical checks (in blue). (b) Classification of manual and automated reactions into the categories hydrogenation/dehydrogenation, oxygenation/deoxygenation, C-C coupling/decoupling, and isomerization. (c) Production rate from MKM modeling of the automatically enumerated reactions and manual reactions with initial feed mixtures $CO_2$:$H_2$=1:3 at 50 bar pressure, over the temperature range of 400-800 K. (d) $CO_2$ hydrogenation pathway to methanol with the highest flux for automatically enumerated reactions at 500K and 50 bar. Pale orange, blue, cream, and green background blocks depict major reactants, surface intermediates, co-reactants (*/H*/$H_2$*), and products, respectively. Values for species marked with an asterisk (*) indicate fractional coverages, while the others represent partial pressures in bar.

them qualitatively and highlight the importance of accounting for all possible elementary reactions rather than just a few tens of reactions.

## Revealing the precise mechanistic pathways of the major products

Several theories abound on the formation mechanism of methanol via $CO_2$ hydrogenation, but none of them account for all possibilities and types of surface reactions. The inclusion of various types of reactions allowed us to clearly determine the actual course of the formation of various product species. For methanol production (Figure 4d), the mechanism starts with $CO_2$ adsorption followed by its hydrogenation to formate via surface hopping of the hydrogen atom of an adsorbed hydrogen molecule ($CO_2^* + H_2^* \rightleftharpoons HCOO^* + H^*$) rather than with an adsorbed hydrogen atom (H*). This is a surprising finding since the latter is often considered to be the main hydrogenation mechanism of $CO_2$*. To validate the favorability of the heretofore unexplored molecular hydrogenation pathway, we carried out explicit DFT calculations, which verified that hydrogenation by molecular hydrogen, as revealed by the ML-backed MKM, is indeed faster (Table 1).

Table 1: Hydrogenation reactions via molecular/atomic hydrogen and their free energy barriers at 500 K calculated using DFT.

| **Reaction** | **Hydrogenation type** | **$G_{act}$ (in eV)** |
|---|---|---|
| $CO_2^* + H_2^* \rightleftharpoons COOH^* + H^*$ / $CO_2^* + H^* \rightleftharpoons COOH^*$ | $H_2^*$ / $H^*$ | 0.61/1.43 |
| $CO_2^* + H_2^* \rightleftharpoons HCOO^* + H^*$ / $CO_2^* + H^* \rightleftharpoons HCOO^*$ | $H_2^*$ / $H^*$ | 0.18/0.64 |
| $HCOOH^* + H_2^* \rightleftharpoons H_2COOH^* + H^*$ / $HCOOH^* + H^* \rightleftharpoons H_2COOH^*$ | $H_2^*$ / $H^*$ | 0.54/0.91 |
| $CH_3O^* + H_2^* \rightleftharpoons CH_3OH^* + H^*$ / $CH_3O^* + H^* \rightleftharpoons CH_3OH^*$ | $H_2^*$ / $H^*$ | 0.44/1.06 |

Nevertheless, the hydrogenation of HCOO* to produce adsorbed formic acid (HCOOH*) does occur using an H*, following which, the formation of $H_2COOH^*$ again occurs with a hopping of hydrogen atom from $H_2^*$ (Table 1). Next, dehydroxylation of $H_2COOH^*$ yields $CH_2O^*$, which, with the assistance of adsorbed atomic hydrogen, forms $CH_3O^*$. Further hydrogenation of $CH_3O^*$ to produce $CH_3OH^*$ occurs with surface $H_2^*$ (Table 1) followed by the desorption of $CH_3OH^*$ to yield methanol. The production of CO (the other major

experimental product) follows the carboxyl pathway: adsorbed $CO_2$* hydrogenates, assisted with hopping of hydrogen atom from $H_2$*, to produce COOH*, which then dehydroxylates to CO* followed by desorption to gas-phase (Section S6).

## Experimental validation of the automated MKM predictions for $CO_2$ hydrogenation on copper

We performed an experimental investigation of $CO_2$ hydrogenation on $Cu/SiO_2$ catalysts, with the silica support purposefully selected to provide an inert carrier to finely disperse copper species (see Methods in Section S4, and Section S9). Briefly, the synthesized crystalline CuO nanoparticles were reduced to metallic copper, as confirmed by powder X-ray diffraction (Figure 5a), hydrogen temperature programmed reduction (Figure 5b), X-ray photoelectron spectroscopy (Figure 5c), and transmission electron microscopy (Figure 5d,e). Catalytic tests were performed using an Incoloy 800H tubular fixed bed reactor system equipped with online gas chromatography by feeding $H_2/CO_2 = 3$ at 50 bar pressure and in the temperature range of 483-523 K. Details on catalytic experiments, product analysis, and performance calculations are provided in Section S9. Figure 5h and 5i depict the $CO_2$ conversion and product selectivity, respectively. We observed that the $CO_2$ conversion rate increases with increasing temperature due to Arrhenius effects, mainly converting into CO and methanol, with the former being the major product (Section S9). The selectivity of methanol was ≈10% at 483 K, decreasing with an increase in temperature. This phenomenon can be exactly seen in the predictions of our MKM model based on automated reactions, which shows an increment and decrement in the production rates of CO and methanol, respectively, with increasing temperature (Figure 4c). The absence of methane from the experimental $CO_2$ hydrogenation product mixture can be associated with the kinetically demanding dehydroxylation reaction of methanol ($\Delta G_{act}$ = 1.69 eV) to produce methyl.

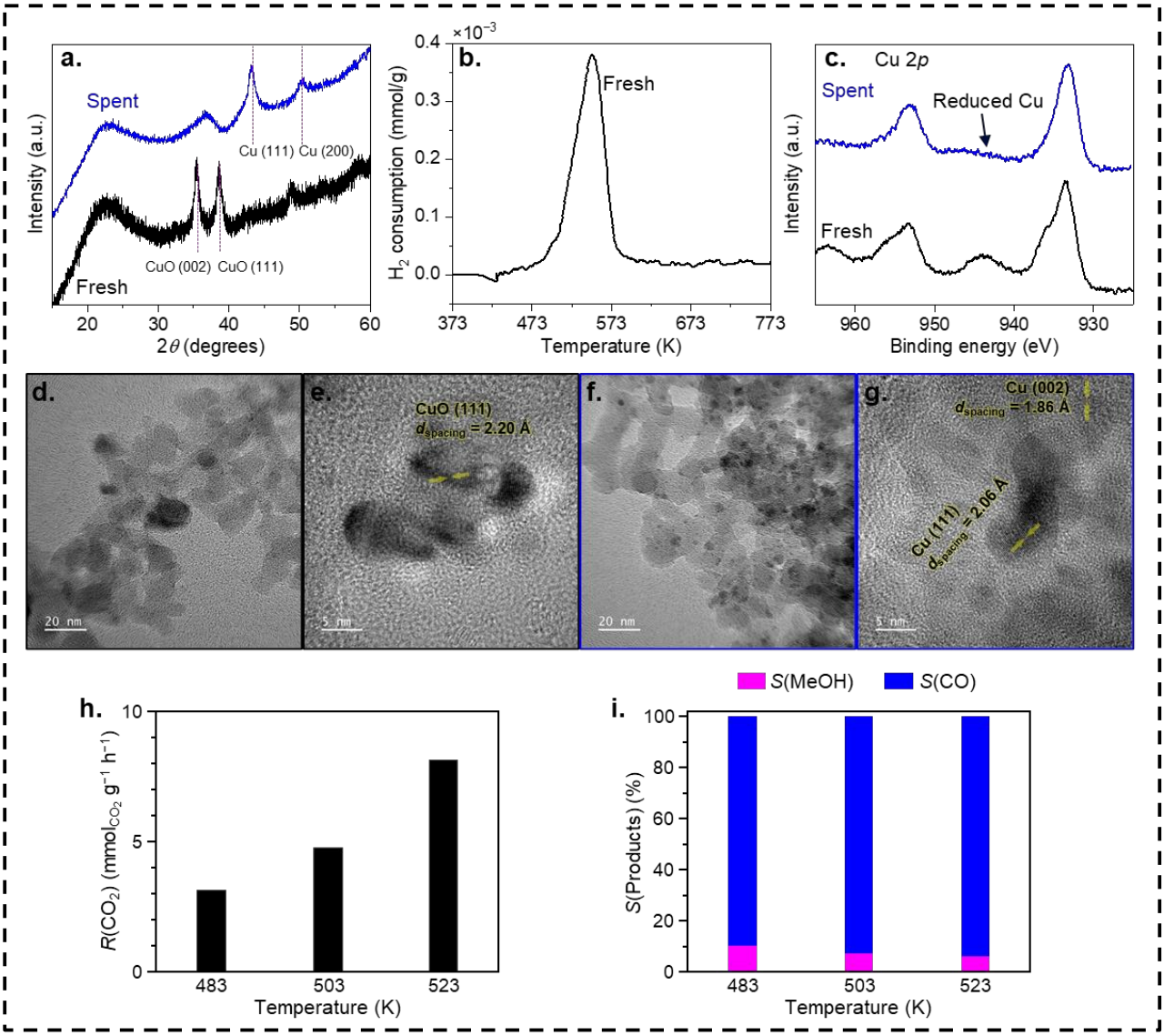


Figure 5: Characterization and catalytic data for $Cu/SiO_2$ catalyst in $CO_2$ hydrogenation reactions. (a) XRD patterns, (b) $H_2$-TPR, (c) XPS Cu 2p, (d-g) TEM images of fresh (black lines/borders) and spent (blue: after $CO_2$ hydrogenation) $Cu/SiO_2$ samples. (h,i) $CO_2$ conversion rate (R) and product selectivity (S) in $CO_2$ hydrogenation over $Cu/SiO_2$ at 50 bar, $H_2$:$CO_2$ = 3, GHSV = 7200 $cm^3$ $h^{-1}$ $g^{-1}$.

# Conclusions

To enable the automated exploration of massive reaction networks, we developed new frameworks for physics-inspired ML-based prediction of activation (free) energies, automated reaction enumeration, and human-intelligence-inspired elementary reaction identification. We circumvented the issue of considerable complexity in activation energy calculations using DFT by building a robust ML model on our dataset consisting of 152 reactions, the largest set of elementary catalytic $CO_2$ reduction reactions ever explored till now using DFT. The combination of automated reaction enumeration and elementary reaction identification generated an extensive set of 9389 surface reactions that were utilized in the MKM model. We subsequently compared two MKM models: one based on the 152 manually curated reactions and the other based on the 9389 automatically generated ones. With our newly devised approach combining DFT, automated reaction enumeration, elementary reaction identification, ML, and MKM, experimental trends of methanol production on Cu(111) were convincingly explained. In contrast, the smaller reaction network, even though significantly larger than previously considered networks, wrongly predicted more formic acid production than methanol and led to a much lower $CO_2$ conversion rate, juxtaposed to experimental results. We also unraveled the precise reaction mechanism leading to the formation of methanol and carbon monoxide, revealing the critical role of previously unexplored hydrogen-transfer and molecular hydrogenation reactions. The strategy developed here can also be readily applied to any thermochemical or electrochemical processes, including $CO_{(2)}$ reduction/hydrogenation, nitrogen reduction, and water splitting.[11,35–37] Future directions can involve generative artificial intelligence for enumerating possible reaction intermediates and improvements in neural network potentials for accurate activation barriers, which can be used in our framework. Overall, our study will significantly accelerate the study of complex chemical mechanisms, enabling better predictions and deep chemical insights, particularly into thermochemical $CO_2$ reduction, which is a crucial chemical conversion process for a cleaner future.

## Acknowledgement

AGR acknowledges financial support from the Hindustan Petroleum Corporation Limited. AA acknowledges funding from the A*STAR CDF (C210812023) and IAF-PP (A19E9a0103) grants. Computational resources were provided by the Supercomputer Education and Research Centre (SERC) at the Indian Institute of Science (IISc). The authors acknowledge Prof. Ivo Filot for his helpful advice regarding MKMCXX, Prof. Sudarshan Vijay for his valuable discussions on microkinetic modeling and his feedback on the manuscript, Prof. Narendra M Dixit and Prof. Ganapathy Ayappa for their valuable comments on the manuscript, and Raghavendra Rajagopalan for insightful discussions on ML and mechanism generation. AGR thanks the Infosys Foundation, Bengaluru, for an Infosys Young Investigator grant. AMV acknowledges the CV Raman Postdoctoral Fellowship from IISc. SC acknowledges the Ministry of Education, Government of India, for a PhD fellowship.

## Supporting Information Available

Details about $CO_2$ adsorption, the adsorption/desorption/formation energetics of intermediate species, discussion on the reaction network of $C_1$ to $C_4$ products, reaction thermodynamics/kinetics, ML parameters, and results, detailed description of the methods employed, MKM plots (TOFs and coverages), further computational details regarding the implementation of ML, automated reaction enumeration, elementary reaction identification, and MKM, and experimental details of the catalyst synthesis procedure, various characterization methods applied, and catalyst testing in the continuous-flow fixed-bed reactor.

## Author Contributions

AGR and AMV were involved in the ideation of the research. AMV performed the DFT simulations and MKM and led the project. SC was involved in developing the ML models and

performing MKM simulations. SP was involved in developing the ARE and ERI frameworks, ML feature generation, and MKM simulations. SN worked on different ML and microkinetic models and sped up the ARE framework. RS worked on feature generation for the ML models and founded a base for the ML models. GV and SK were involved in the ideation of the ML models and discussions on predicting the reaction outcomes. AD gave inputs on the ML models. AA conceptualized and directed the experimental work and prepared the $Cu/SiO_2$ catalyst. CGG and CQJN performed catalytic experiments. PYM carried out characterizations of the fresh and spent catalyst samples. AGR directed the overall research and developed the initial code for the ARE and ERI frameworks. AMV, SC, SP, and AGR wrote the paper, and AA wrote the description of the experimental results. All authors commented on the manuscript.

# References


(1) Margraf, J. T.; Jung, H.; Scheurer, C.; Reuter, K. Exploring catalytic reaction networks with machine learning. *Nature Catalysis* **2023**, *6*, 112–121.

(2) Ulissi, Z. W.; Medford, A. J.; Bligaard, T.; Nørskov, J. K. To address surface reaction network complexity using scaling relations machine learning and DFT calculations. *Nature Communications* **2017**, *8*, 14621.

(3) Grabow, L.; Mavrikakis, M. Mechanism of methanol synthesis on Cu through $CO_2$ and CO hydrogenation. *ACS Catalysis* **2011**, *1*, 365–384.

(4) Yang, Y.; Evans, J.; Rodriguez, J. A.; White, M. G.; Liu, P. Fundamental studies of methanol synthesis from $CO_2$ hydrogenation on Cu(111), Cu clusters, and Cu/ZnO(0001). *Physical Chemistry Chemical Physics* **2010**, *12*, 9909.

(5) Hutton, D. J.; Cordes, K. E.; Michel, C.; Göltl, F. Machine Learning-Based prediction

of activation energies for chemical reactions on metal surfaces. *Journal of Chemical Information and Modeling* **2023**, *63*, 6006–6013.

(6) Stocker, S.; Csanyi, G.; Reuter, K.; Margraf, J. T. Machine learning in chemical reaction space. *Nature communications* **2020**, *11*, 5505.

(7) Rangarajan, S.; Kaminski, T.; Van Wyk, E.; Bhan, A.; Daoutidis, P. Language-oriented rule-based reaction network generation and analysis: Algorithms of RING. *Computers Chemical Engineering* **2014**, *64*, 124–137.

(8) Govind Rajan, A.; Carter, E. A. Discovering Competing Electrocatalytic Mechanisms and Their Overpotentials: Automated Enumeration of Oxygen Evolution Pathways. *The Journal of Physical Chemistry C* **2020**, *124*, 24883–24898.

(9) Liu, M.; Dana, A. G.; Johnson, M. S.; Goldman, M. J.; Jocher, A.; Payne, A. M.; Grambow, C. A.; Han, K.; Yee, N. W.; Mazeau, E. J.; Blondal, K.; West, R. H.; Goldsmith, C. F.; Green, W. H. Reaction Mechanism Generator v3.0: Advances in Automatic Mechanism Generation. *Journal of Chemical Information and Modeling* **2021**, *61*, 2686–2696.

(10) Zhong, M.; Tran, K.; Min, Y.; Wang, C.; Wang, Z.; Dinh, C.-T.; De Luna, P.; Yu, Z.; Rasouli, A. S.; Brodersen, P., et al. Accelerated discovery of CO2 electrocatalysts using active machine learning. *Nature* **2020**, *581*, 178–183.

(11) Chen, X.; Chen, J.; Alghoraibi, N. M.; Henckel, D. A.; Zhang, R.; Nwabara, U. O.; Madsen, K. E.; Kenis, P. J. A.; Zimmerman, S. C.; Gewirth, A. A. Electrochemical $CO_2$-to-ethylene conversion on polyamine-incorporated Cu electrodes. *Nature Catalysis* **2020**, *4*, 20–27.

(12) Prašnikar, A.; Jurković, D. L.; Likozar, B. Reaction Path Analysis of $CO_2$ Reduction to Methanol through Multisite Microkinetic Modelling over $Cu/ZnO/Al_2O_3$ Catalysts. *Applied Catalysis B: Environmental* **2021**, *292*, 120190.

(13) Zhao, Y.-F.; Yang, Y.; Mims, C.; Peden, C. H.; Li, J.; Mei, D. Insight into methanol synthesis from $CO_2$ hydrogenation on Cu(111): Complex reaction network and the effects of $H_2O$. *Journal of Catalysis* **2011**, *281*, 199–211.

(14) Choi, Y. H.; Jang, Y. J.; Park, H.; Kim, W. Y.; Lee, Y. H.; Choi, S. H.; Lee, J. S. Carbon dioxide Fischer-Tropsch synthesis: A new path to carbon-neutral fuels. *Applied Catalysis B Environment and Energy* **2016**, *202*, 605–610.

(15) Nie, X.; Wang, H.; Janik, M. J.; Chen, Y.; Guo, X.; Song, C. Mechanistic Insight into C–C Coupling over Fe–Cu Bimetallic Catalysts in $CO_2$ Hydrogenation. *The Journal of Physical Chemistry C* **2017**, *121*, 13164–13174.

(16) Tackett, B. M.; Gomez, E.; Chen, J. G. Net reduction of $CO_2$ via its thermocatalytic and electrocatalytic transformation reactions in standard and hybrid processes. *Nature Catalysis* **2019**, *2*, 381–386.

(17) Gao, S.-T.; Xiang, S.-Q.; Shi, J.-L.; Zhang, W.; Zhao, L.-B. Theoretical understanding of the electrochemical reaction barrier: a kinetic study of $CO_2$ reduction reaction on copper electrodes. *Physical Chemistry Chemical Physics* **2020**, *22*, 9607–9615.

(18) Ting, L. R. L.; García-Muelas, R.; Martín, A. J.; Veenstra, F. L. P.; Chen, S. T.; Peng, Y.; Per, E. Y. X.; Pablo-García, S.; López, N.; Pérez-Ramírez, J.; Yeo, B. S. Electrochemical Reduction of Carbon Dioxide to 1-Butanol on Oxide-Derived Copper. *Angewandte Chemie International Edition* **2020**, *59*, 21072–21079.

(19) Zhao, Q.; Martirez, J. M. P.; Carter, E. A. Charting C–C coupling pathways in electrochemical $CO_2$ reduction on Cu(111) using embedded correlated wavefunction theory. *Proceedings of the National Academy of Sciences* **2022**, *119*, e2202931119.

(20) Duan, C.; Du, Y.; Jia, H.; Kulik, H. J. Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model. *Nature Computational Science* **2023**, *3*, 1045–1055.

(21) Choi, S.; Kim, Y.; Kim, J. W.; Kim, Z.; Kim, W. Y. Feasibility of activation energy prediction of Gas-Phase reactions by machine learning. *Chemistry - A European Journal* **2018**, *24*, 12354–12358.

(22) Lalith, N.; Singh, A.; Gauthier, J. *The importance of reaction energy in predicting chemical reaction barriers with machine learning models* **2023**, Preprint at https://doi.org/10.26434/chemrxiv–2023–t6zrj.

(23) Göeltl, F.; Mavrikakis, M. Generalized Brønsted-Evans-Polanyi Relationships for Reactions on Metal Surfaces from Machine Learning. *ChemCatChem* **2022**, *14*, e202201108.

(24) Winther, K. T.; Hoffmann, M. J.; Boes, J. R.; Mamun, O.; Bajdich, M.; Bligaard, T. Catalysis-Hub.org, an open electronic structure database for surface reactions. *Scientific Data* **2019**, *6*.

(25) Goldsmith, C. F.; West, R. H. Automatic Generation of Microkinetic Mechanisms for Heterogeneous Catalysis. *The Journal of Physical Chemistry C* **2017**, *121*, 9970–9981.

(26) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **2013**, *1*.

(27) Scott M. Lundberg Paul G. Allen School of Computer Science, S. W., University of Washington A unified approach to interpreting model predictions | Proceedings of the 31st International Conference on Neural Information Processing Systems. *Guide Proceedings* **2017**, 4768–4777.

(28) Korth, M.; Grimme, S. "Mindless" DFT benchmarking. *Journal of Chemical Theory and Computation* **2009**, *5*, 993–1003.

(29) Paier, J.; Hirschl, R.; Marsman, M.; Kresse, G. The Perdew–Burke–Ernzerhof exchange-correlation functional applied to the G2-1 test set using a plane-wave basis set. *The Journal of Chemical Physics* **2005**, *122*.

(30) Adamo, C.; Ernzerhof, M.; Scuseria, G. E. The meta-GGA functional: Thermochemistry with a kinetic energy density dependent exchange-correlation functional. *The Journal of Chemical Physics* **2000**, *112*, 2643–2649.

(31) Granda-Marulanda, L. P.; Rendón-Calle, A.; Builes, S.; Illas, F.; Koper, M. T. M.; Calle-Vallejo, F. A Semiempirical Method to Detect and Correct DFT-Based Gas-Phase Errors and Its Application in Electrocatalysis. *ACS Catalysis* **2020**, *10*, 6900–6907.

(32) Hensley, A. J. R.; Therrien, A. J.; Zhang, R.; Marcinkowski, M. D.; Lucci, F. R.; Sykes, E. C. H.; McEwen, J.-S. CO adsorption on the "29" CUXO/CU(111) surface: an integrated DFT, STM, and TPD study. *The Journal of Physical Chemistry C* **2016**, *120*, 25387–25394.

(33) Kunkes, E. L.; Studt, F.; Abild-Pedersen, F.; Schlögl, R.; Behrens, M. Hydrogenation of $CO_2$ to methanol and CO on $Cu/ZnO/Al_2O_3$: Is there a common intermediate or not? *Journal of Catalysis* **2015**, *328*, 43–48.

(34) Nakamura, I.; Fujitani, T.; Uchijima, T.; Nakamura, J. A model catalyst for methanol synthesis: Zn-deposited and Zn-free Cu surfaces. *Journal of Vacuum Science Technology A* **1996**, *14*, 1464–1468.

(35) Guzmán, H.; Russo, N.; Hernández, S. $CO_2$ valorisation towards alcohols by Cu-based electrocatalysts: challenges and perspectives. *Green Chemistry* **2021**, *23*, 1896–1920.

(36) Li, H.; Liu, T.; Wei, P.; Lin, L.; Gao, D.; Wang, G.; Bao, X. High-Rate $CO_2$ Electroreduction to $C_{2+}$ Products over a Copper-Copper Iodide Catalyst. *Angewandte Chemie* **2021**, *133*, 14450–14454.

(37) Nitopi, S.; Bertheussen, E.; Scott, S. B.; Liu, X.; Engstfeld, A. K.; Horch, S.; Seger, B.; Stephens, I. E. L.; Chan, K.; Hahn, C.; Nørskov, J. K.; Jaramillo, T. F.; Chorkendorff, I. Progress and Perspectives of Electrochemical $CO_2$ Reduction on Copper in Aqueous Electrolyte. *Chemical Reviews* **2019**, *119*, 7610–7672.