Exploring the intricacies of Glycerol Hydrodeoxygenation on Copper surface: A Comprehensive Investigation with the Aid of Machine Learning Forcefield.

Srishti Gupta[1], Ajin Rajan[2], Edvin Fako[3], Tiago J. F. Gonçalves[3], Imke B. Müller[3], Jithin John Varghese[2], Ansgar Schäfer[3], Sandip De[3].*

[1] BASF Chemicals India Pvt. Ltd., Plot No.12, TTC Area, Thane Belapur Road, Turbhe, Navi Mumbai (India), 400705.

[2] Department of Chemical Engineering, Indian Institute of Technology, Madras, Chennai (India), 600036.

[3] BASF SE, Group Research, Carl-Bosch-Straße 38, 67056 Ludwigshafen (Germany).

KEYWORDS: Glycerol, Machine learn force fields, Hydrodeoxygenation, Reaction network, Heterogeneous catalyst, conformational sampling, Propanediol, Machine learning.

ABSTRACT:

The utilization of biomass to feedstock chemicals often relies on transforming hydroxyl-containing molecules. One such example is glycerol which can undergo a selective hydrodeoxygenation reaction to produce propanediol, a valuable chemical precursor. Hence, glycerol's hydrodeoxygenation reaction combines immediate industrial application with foundation of fundamental research into the reaction class relevant for sustainable feedstock. Given the complex nature of large organic molecules, most modelling work in heterogeneous catalysis focusses on the reactivity of small (C1-2) organics exclusively. Glycerol, characterized by its C3-backbone, exhibits 75 distinct gas-phase conformers.[1] When considering its 11 reactive bonds (C-O, C-H and O-H), the modeling of glycerol's reactivity spans an extensive conformational and reactive space. High computational costs of Density Functional Theory simulations restrict exhaustive exploration of the factorial reaction space, leading to limited insights of the hydrodeoxygenation (HDO) mechanism and hindering rational catalyst design. Therefore, to date, there is no systematic study focusing on comprehensively sampling the energetics of surface conformers of glycerol and their reactivity.

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

In this study, we employ a message-passing graph neural network architecture (MACE) to develop a machine-learned force-field (MLFF) potential, utilizing active learning to investigate the impact of conformational complexity on the reaction network of glycerol HDO on a Cu(111) surface. Following five iterations, our trained MLFF model accurately predicts surface bound structures with a root-mean-square accuracy of 0.04 eV (< 0.6 meV/atom total energy), essential to accurately determine conformational minima of 24 meta-stable and 26 intermediate states along seven competitive pathways. Conformational sampling uncovers the intricate nature of the complex energy landscape, where conformers with multiple shallow minima lead to non-trivial trends in the transition state energies connecting them. Notably, the investigations predict lower activation barriers for O-H bond scissions of glycerol structures with α- and γ-backbone as compared to β-backbone. This is significant in case of scission of secondary O-H glycerol bonds where the activation barrier varies up to 0.44 eV depending upon the initial glycerol structure motif. Altogether, we identify dehydrogenation-dehydration-hydrogenation as dominant pathway resulting in PDO formation on the Cu(111) surface. The selectivity of glyceraldehyde towards C-H bond scission over C-OH bond scission explains higher selectivity of 1,2-PDO over 1,3-PDO.

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

1. Introduction

Glycerol (propane-1,2,3-triol) is the main by-product of transesterification of triglycerides from biomass feedstocks to form biodiesel.[2] Glycerol's surplus availability has created a need for effective utilization strategies. It's upcycling to propane-diol (PDO) by HDO reaction has been incentivized as a valuable and commercially attractive route for valorization.[3] 1,2-PDO, is used in unsaturated polyester resins, functional fluids, cosmetics and pharmaceuticals.[4] It's isomer 1,3-PDO, is used as a plasticizer, corrosion remover, in paint and copolymers, and as a monomer for polytrimethylene terephthalate (PTT).[5]

Recent efforts in catalyst design and development have focused on enhancing the activity, selectivity, and stability of catalysts for glycerol conversion to PDO.[6–11] The use of non-noble metal catalysts,[12] mixed metal oxide catalysts,[13] particularly Cu-based catalysts,[14] have shown promise in achieving higher 1,2-PDO yields. Additionally, Cu-based catalysts can be further modified by incorporating other metals or promoters, such as nickel,[15] chromium,[16] aluminum,[17] zinc,[18] and silica,[19] to improve glycerol conversion, increase selectivity to PDO, and enhance catalyst stability. Cu-based catalysts exhibit between 51-100% conversion of glycerol with selectivity up to 76-100% towards 1,2-PDO at 180-230 ℃ and 15-18 bar pressure ranges.[7,12–14] The conversion was observed to decrease at higher concentration of glycerol in the reactant mixture. The optimal reaction conditions for Cu-based catalysts, such as temperature and pressure, can vary depending on the specific catalyst composition and support. Therefore, finding the right conditions, support and promoters to maximize both glycerol conversion and selectivity to 1,2-PDO is an ongoing challenge.

Density Functional Theory (DFT) simulations have emerged as a vital tool for catalyst development, significantly enhancing our ability to understand and optimize catalytic processes.[20,21] DFT simulations predict key parameters such as adsorption energies, and activation barriers, thereby enabling the identification of optimal reaction pathways. Furthermore, DFT simulations facilitate the exploration of how various catalyst compositions, surface structures and environmental conditions influence catalytic performance.[22–24] Consequently, the insights gained from DFT studies can provide a strategic roadmap for optimizing reaction conditions and catalyst compositions,[25] ultimately guiding the development of advanced catalysts for efficient glycerol conversion to PDO.[26]

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

However, the usual process of expert knowledge-driven modelling workflows for mechanistic investigation through DFT simulations is laborious and time-consuming.[27] They are highly based on human-mediated enumeration of reaction possibilities and risks introducing human bias. Glycerol has 3^5=243 combinations of dihedral angles with a staggered configuration of the two C-C and three C-O bonds, out of which 126 are symmetry non-equivalent. 75 of these configurations correspond to stable conformers in the gas phase.[1] Callam et. al. (2001) indicated that intramolecular hydrogen bonding and entropic contributions are important for accurately determining the relative conformer energies. Selecting only a limited number of glycerol conformers can bias the conclusions drawn from an incomplete Boltzmann distribution of conformers in the gas phase, thereby skewing the results towards the specific ensemble considered, [1] which applies to reactions on solid state catalysts as well. On top of the conformational complexity, the proposed reaction network for glycerol conversion also suggests presence of 6 primary C_3 intermediates, from with formation of PDO is possible.[28]

Previous studies on the adsorption of glycerol on surfaces have not comprehensively evaluated all 75 stable configurations, highlighting a significant gap in the current literature. Often, the ab-initio investigation of glycerol HDO reactions is restricted to explain the results of experimental output with limited computational resources. Typical investigations rely on the energetically most favorable conformers for the discovery of minimum energy paths and reaction mechanisms.[12] However, it is by no means necessarily the case that the most favorable conformers of reactant and products are connected by the lowest lying transition state. Therefore, there is a severe lack of predictive computational studies which either provide in-depth insights or propose new avenues to further improve the catalytic processes for glycerol conversion to 1,2-PDO.

In response to the computational limitations associated with DFT simulations, machine learning provides a novel framework for enhancing the exploration of reaction networks and conformational diversity, thereby advancing our understanding of catalytic processes.[27,29] With the orders of magnitude faster machine learned force-field (MLFF) potentials, we could successfully delve into an extensive spectrum of glycerol conformers and reaction pathways, employing a blend of automated techniques and human-assisted methods for conformational and transition state exploration of glycerol conversion to 1,2-PDO. This study aims to simultaneously explore the complex conformational and reaction network landscape of the glycerol conversion to PDO on a pure Cu(111) surface, circumventing the constraints of

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

conventional DFT simulations by harnessing the benefits of a higher-order equivariant message-passing neural network architecture (MACE) MLFF.[30,31] Our ambition is to supplant the traditionally simplistic solution to the query of "what is the reaction mechanism?" with a more comprehensive, unbiased approach.

2. Methods

2.1. Active Learning Protocol

A common emerging strategy for developing MLFF involves an iterative approach, where the training set is expanded according to the application requirements of the force field.[32–36] The training set, consisting of relevant configurations, is used to train the MLFF by comparing properties with a reference method like DFT. The trained MLFF is then used to generate new configurations and predictions (often with some kind of uncertainty estimation), which are validated against the reference method. If the predictions are unsatisfactory, in the absence of any relevant limitations of the ML framework or issues with the reference data, the primary assumption is that the model prediction is based on unreliable extrapolation. This assumption is further scrutinized with structural similarity analysis. To “teach” the model the unseen region, those configurations are added to the training set and the training process is restarted or repeated until the requirements are met.

However, applying this strategy in heterogeneous catalysis for its complex reaction mechanisms is still challenging due to the lack of a reliable autonomous sampling strategy for reaction intermediates and transition states. We expand the previously developed method of active learning to scale up the traditional catalysis workflow,[32] where intermediates are proposed, verified by DFT relaxations, and transition states are searched based on these intermediates. We follow the steps: postulate the existence of numerous intermediate conformers and transition states, compute energies and forces with single-point DFT calculations to fit an MLFF and optimize the postulated configurations or reaction pathways with the MLFF. The configurations along the newly explored reaction pathways as well as altered configurations computed with new MLFF are verified by DFT. This iterative process continues until the emerging pathway converges to the desired accuracy threshold.

2.1.1. Density Functional Theory simulation parameters

Single-point periodic DFT calculations were performed for selected configurations using the Quantum Espresso software package,[37] and the Perdew-Burke-Enzerhof (PBE)[38] exchange and correlation functional. Standard solid-state pseudopotentials (SSSP)[39] were

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 ORCID: https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. License: CC BY-NC 4.0

employed to reduce the explicitly calculated orbitals to 1s of H; 2s and 2p of O and C, and 3d, 4s, and 4p of Cu atoms. The self-consistent field (SCF) cycle was considered to be converged when energy differences per electronic step were below 10^{-7} eV. The valence monoelectronic states were expanded in plain waves with a kinetic energy cutoff of 884 eV. Gaussian smearing with a width of 0.1 eV was applied. The density of k-points in reciprocal space was set to 0.25 $Å^{-1}$ in the xy plane.

2.1.2. ML model parameters

A higher-order equivariant message-passing neural network architecture, referred to as MACE[40,41] implemented in PyTorch[42] was employed to train a single MLFF that can describe all the reaction steps reported.[43,44] All models referred to in this work use two MACE layers, a spherical expansion of up to lmax = 3, and 4-body messages in each layer (correlation order 3). All models used a 64-channel dimension for tensor decomposition. We used a radial cutoff of 6 Å and expanded the interatomic distances in 10 Bessel functions multiplied by a smooth polynomial cutoff function to construct radial features. The radial features are in turn fed into a fully connected feed-forward neural network with three hidden layers of 64 hidden units, a shallow MLP for the second layer readout with 64 hidden units and SiLU non-linearities. We fit a model with the maximal message equivariance, L = 2. The irreducible representations of the messages have alternating parity (in e3nn notation 64x0e + 64x1o + 64x2e). The final model was trained on a single NVIDIA A100 (80GB) GPU over the last 4 iterations (fresh start from iteration number 2 and then always restarted with new training data for additional epochs) for additional 250,+50,+100,+300 epochs, respectively.

2.1.3. Initial training data generation and base model building

In this study we focus on the glycerol HDO reaction mechanism on the lowest energy facet of Cu – [111].[45,46] . The Cu(111) catalyst was represented by a slab consisting of 4x4x4 primitive units with a lattice constant of 3.61 Å. All slabs were separated by 20 Å of vacuum. For each slab, the bottom 2 layers were fixed. Starting with the previously reported[32] protocol of initial training set design, our approach includes a carefully constructed set of conformers of glycerol and other intermediates, in gas phase and on the Cu(111) surface. Additionally, we focus on maximizing diversity in the training data set configurations by applying statistical sampling approaches on both structural and property level. Considering the complexity in the glycerol HDO reaction scheme and high accuracy requirement, we initiated a large set (~1000) and performed single-point DFT calculations. This was done using a

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

combined approach of automation and expert knowledge-driven hypothesis generation for the entire reaction scheme as discussed in results section 3.2. While current best practices in active learning approaches suggest starting with a small number of initial training data and prioritizing higher number of iterations to reduce the cost of reference DFT data generation, in an industrial research environment, we prioritize having smaller number of iterations, which, with the present technological constraints, means the quickest way to reach the desired accuracy of the final MLFF model. This initial set included sufficient information to study the flexible backbone with multi-reactive groups of glycerol and other reaction intermediates in a highly branched reaction network. The base MLFF model in iteration 1 was trained on only 30% of randomly selected configurations and validated on the remaining 70%. This was sufficient to achieve moderate accuracy as noted in Table 1. The objective of this iteration was to get an estimate of the accuracy, identifying any issues with the reference data generation and getting ready for the required targeted data generation for next iterations. We used this model to curate, prune and generate additional data, especially relevant for transition pathway exploration to build a training and validation data set having 6589 and 1967 entries respectively for iteration 2.

2.1.4. Model finetuning over three iterations

Table 1: MACE MLFF active learning summary.

	epochs		RSME E (meV/atom)	RSME F (meV/Å)
Iteration 1	600	Training: 2972	2.8	20.4
		Validation: 6934	3.0	55.5
Iteration 2	250	Training: 6589	1.8	9.2
		Validation: 1967	0.9	21.0
Iteration 3	+50	Training: 9782	2.0	11.6
		Validation :1967	0.9	20.6
Iteration 4	+100	Training: 10373	1.7	10.2
		Validation: 1967	0.7	19.7
Iteration 5	+300	Training: 12224	1.7	9.1
		Validation: 1967	0.7	18.3

Although we counted the base model building as iteration 1, the actual active learning started from iteration 2 onwards, where a new model with the consolidated new training and validation set mentioned in the last section was used to start a model training from scratch for 200 epochs. The MLFF model was then applied to explore the reaction network and collect new train data employing the principle described before, followed by explicit DFT calculations resulting in new training data. Iteration 3-5 resumed the training from the last checkpoints of the earlier iteration, expanding the training sets to 9782 and 12224 respectively. The validation

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 ORCID: https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. License: CC BY-NC 4.0

set was kept fixed throughout these 3 iterations. The summary of the results is shown in Table 1.

2.2. Applied Corrections

2.2.1. **Dispersion correction:** Long-range dispersion corrections (van der Waals, vdW), utilized in correction schemes such as DFT-D3[47] functioning with a Becke-Johnson damping[48] was applied to MLFF calculations (PyTorch[49] implementation).

2.2.2. **Thermodynamic correction:** Thermal effects on the full reaction network investigated in this work were addressed by performing MLFF vibrational calculations on the optimized structures (reactant/product and transition-state converged structures). Gibbs energies were then calculated by computing the zero-point energy corrections, heat capacities at constant volume and entropy based on the resulted frequencies from the vibrational calculations, within the harmonic approximation at temperature of 473 K (details are available in the SI-1). By comparing Gibbs energies and DFT?ML energies, we observe that accounting for thermodynamic corrections has pronounced effects on adsorption and reaction energetics and will drastically change the rates of reactions for temperatures above room temperature (Figure SI-15). For example, the minimum energy glycerol configuration, with reference to the clean slab and gasphase species, has a Gibbs energy of adsorption of -0.57 eV (as compared to -0.94 eV of DFT adsorption energy). Similarly, the Gibbs energy of reaction for the primary C-OH scission reaction is ΔG_{rxn} = -0.02 eV as compared to ΔE_{rxn}= -0.18 eV, with zero-point corrections and entropic effects amounting to the difference of 0.16 eV. To cite a reactive example, the Gibb's energy of activation ($\Delta G^{\ddagger}$ = 1.04 eV) of the primary C-H scission is 0.15 eV lower than the activation energy ($\Delta E^{\ddagger}$ = 1.19 eV) without thermodynamic corrections. For this reason, the final energy profiles are reported in terms of Gibb's energies (ΔG) at 473K and 1 bar.

2.3. Transition-state calculations

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

For each elementary reaction step, 1500 conformers of the reactant were selected using farthest point sampling and nudged elastic bands (NEB) for each conformer were prepared and evaluated. Products for bond breaking reactions (for e.g., C-OH scission of glycerol (Figure 1)) were created in a way that the cleaved product (OH^*) is placed on the neighboring Cu surface site. The nearest neighboring surface sites (top, hollow or bridge) were selected for the placement of the dissociated fragments after scission.

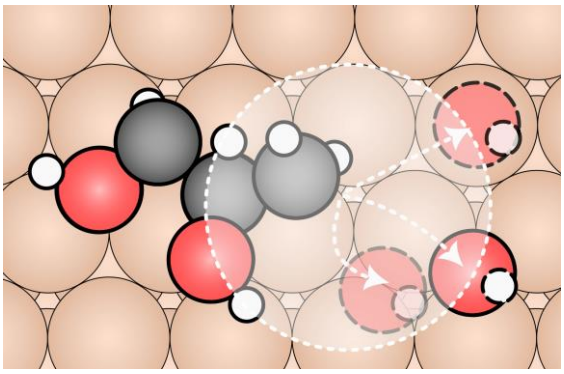


Figure 1: Postulated primary C-O scission of a glycerol molecule adsorbed on the Cu (111) surface. The white shaded circular area corresponds to the part of the surface that can potentially host the cleaved OH group. The dotted white arrows point towards potential top, bridge and hollow sites. (Colorcode: Cu – orange, C – gray, O – red and H - white)

Once the products created above were relaxed, the NEB pathway connecting a product conformer with the original reactant conformer was created with 20 intermediate images using the idpp[50] method. For the bond association reactions (example hydrogenation of hydroxyacetone to 1,2-PDO), the reaction is set up in the reverse order (the O-H scission of 1,2-PDO), and the reverse activation barriers and reaction energies are computed using MLFF to find the minimum energy reaction pathways. Transition state (TS) scans were performed using the dynamic nudged elastic band (DyNEB) [51–54] method as implemented in the atomic simulation environment (ASE)[55] with a maximum force threshold (fmax) of 0.05 eV/Å. TS structures attained using NEBs are further refined to a tighter convergence of fmax=0.02 eV/Å with the automated relaxed potential energy surface scans (ARPESS)[56] method. Obtained TS structures were confirmed by performing vibrational calculations with the MLFF, ensuring an imaginary frequency at the reaction vibrational mode. Connection points of a given transition state were reconfirmed using the intrinsic reaction coordinate (IRC)[57] method.

3. Results and Discussion

3.1. Exploration of glycerol conformers and transition state

3.1.1. Adsorption of glycerol conformers on the Cu(111) surface

In this study, we enumerate glycerol structures on Cu(111) surface by analyzing their adsorption energies. Callam et. al. (2001)[1] identified 126 distinct monomer configurations of glycerol, of which 75 were determined to be stable conformers. However, previous studies[3,9,58–60] on adsorption of glycerol only considered a subset of these structures as

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 ORCID: https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. License: CC BY-NC 4.0

starting points and, therefore, a large part of the potential energy landscape remains undiscovered.

To address this gap, we employed MACE-MLFF with minima hopping54 search method to systematically search for minimum structures, initiated using 70 stable glycerol conformers in gas phase, obtaining a collection of over 16000 relaxed configurations resulting from surface-induced alterations and the structural intricacies. The structural representations of adsorbed glycerol structures and initial gas-phase monomers were encoded using smooth overlap of atomic positions (SOAPs) descriptors,[62] which compares the structures' local atomic environments (rcut =3.0 Å). The high-dimensional SOAP vectors are further reduced to two-dimensional representations using the uniform manifold approximation and projection (UMAP)56 method for easier visualization, as seen in Figure 2. The two-dimensional representation shows all the configurations as points where "similar" configurations are projected close to each other while far away points represent configurations that are structurally different. The distinct clusters forming out of this projection represent groups with major structural similarity motifs among adsorbed glycerol configurations. We further project the pure gas phase glycerol conformers on top of this map to highlight the overlapping motifs with adsorbed conformers.

We find adsorbed glycerol structures showing complex structural motifs which includes the essence of gas phase conformer motifs, but also add additional richness due to structural changes arising from surface adsorbate interactions as seen in Figure 2(a). We observed that the distinction between the glycerol conformer groups arises primarily from the orientation of the C_3O_3 backbone dihedral bond angles, as seen in Figure SI-2, where there is a strong correlation between dihedrals and clusters in Figure 2. The backbone dihedrals on each side of glycerol structure can be categorized into 3 types (α, β, γ), as shown in Figure 2(d). This leads to 6 distinct categories (or families) of glycerol structures (αα, αβ(/βα), αγ(/γα), ββ, βγ(/γβ), γγ), that can explain the clusters in UMAP (Figure 2(b)). This observation indicates that glycerol backbone dihedrals remain the most dominant feature in distinguishing glycerol motifs in adsorbed glycerol structures similar to gas phase structures. However, note that, even though the dihedrals can be used to segregate glycerol structures into distinct categories, they cannot be used alone to determine the adsorption energy of the glycerol structures. The adsorption energy profiles also show low energy structures (< 0.1 eV above the minimum energy structure) in all conformer families defined by backbone, except βγ (Figure SI-4). The most strongly adsorbed glycerol conformer is part of the αγ family with adsorption energy (ΔE = -0.94 eV)

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

followed closely by members of the αα (ΔE = -0.93 eV), γγ and ββ families (ΔE = -0.9 eV). This illustrates the presence of many shallow wells around the global minimum of the potential energy landscape of glycerol adsorption on the Cu(111) surface. This result aligns well with the experimental[64] and MD[65] simulated percentage probability distribution of glycerol conformers.

It is challenging to rationalize the energetic hierarchy of the glycerol structures and consequently deduce the most stable configurations. Apart for backbone dihedrals, other features such HOCC glycerol dihedrals, C-C-C and H-O-C angles, intramolecular H-bonds, average distance from Cu(111) surface, etc. vary and give rise to smaller clusters among larger families (αα, αβ, αγ, ββ, βγ, γγ), as seen in Figure 2(b). For example, the largest αα cluster with no gas phase projection overlaying on top of this cluster has a slightly larger C-C-C angle (116˚) as compared to the relaxed gas-phase C-C-C angle (~110˚). The dihedrals of C-O bonds and intra-molecular H-bonds also produce smaller but distinct clusters among the larger conformational clusters. The chemiscope[66] of these features and glycerol structures are also provided as a supplementary information for readers to visualize the distinctions.

Overall, the energetic ordering is dictated by the balance between surface contact and hydrogen bonding. The average distance of glycerol atoms to the Cu(111) surface can be partially used to determine the strength of adsorption: the lower the average distance of glycerol conformer the more favorable is the adsorption energy (Figure 2c). However, we also see that structures with shorter average distance (~ 3Å) from the Cu(111) surface can have an adsorption energy between -0.94 eV to -0.4 eV. Other factors such as intramolecular H-bonding also play a role in determining the adsorption behavior which are not easily quantifiable. The lack of simple descriptors for such problems highlights the importance of MLFF and our active learning strategy which were successful in addressing this complexity.

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

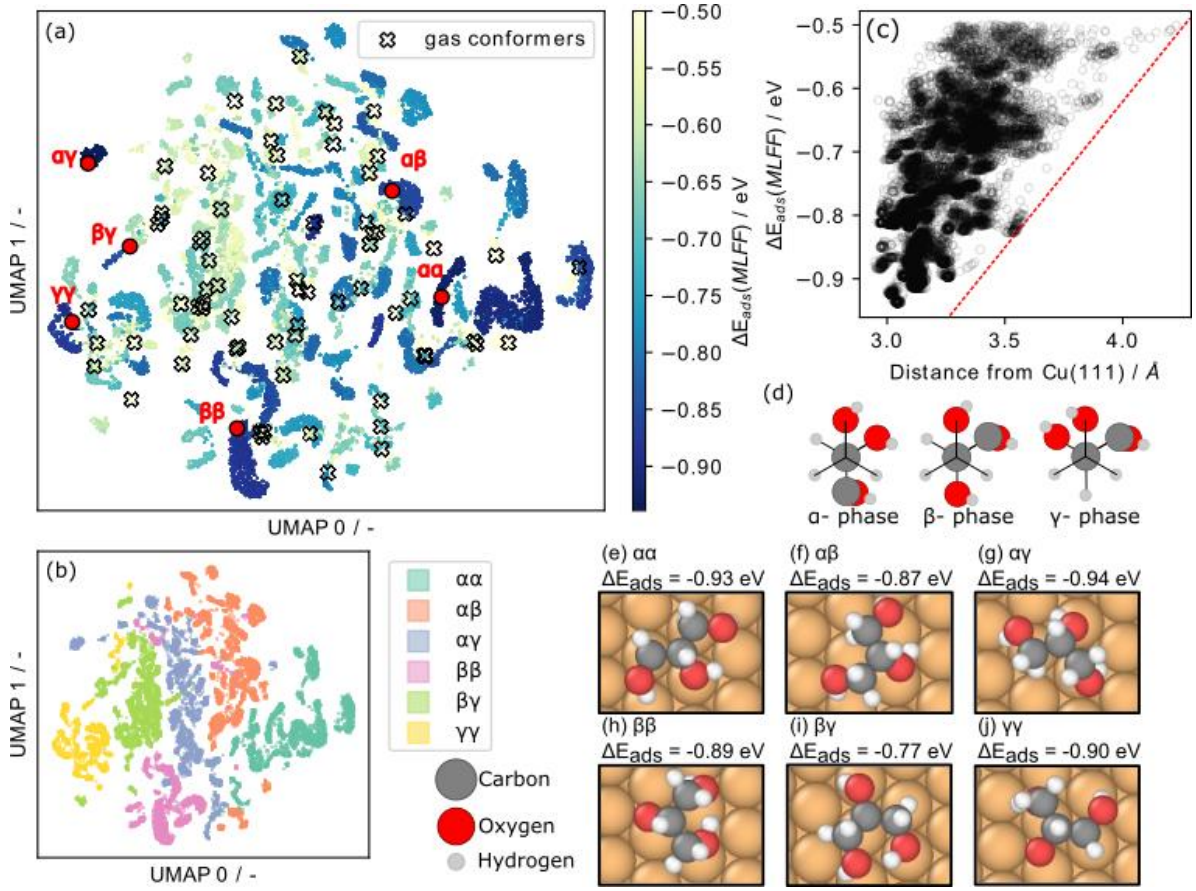


Figure 2: (a) Adsorption energy of explored 16862 relaxed glycerol structures on the Cu(111) surface represented as a function of their structure similarity. The map has been constructed computing SOAP descriptors followed by application of a Uniform Manifold approximation and Projection (UMAP) model to arrive at the 2-dimensional representation. The crosses indicate the 70 gas phase glycerol structures and red dots indicate global minimum glycerol structure on Cu(111) in each conformer. (b) Distinction of glycerol conformers on Cu(111) surface as seen on UMAP. (c) Adsorption energy of glycerol as a function of average distance of each atom on glycerol molecule from top surface of Cu(111). (d) a sketch representation of α, β and γ phases determining six glycerol conformers (αα, ββ, γγ, αβ, αγ and βγ). (e)-(j) The top view of most exothermic glycerol conformers (αα, ββ, γγ, αβ, αγ and βγ) and their adsorption energies calculated using MACE MLFF on the Cu(111) surface.

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

3.1.2. Glycerol bond scission on Cu(111) surface

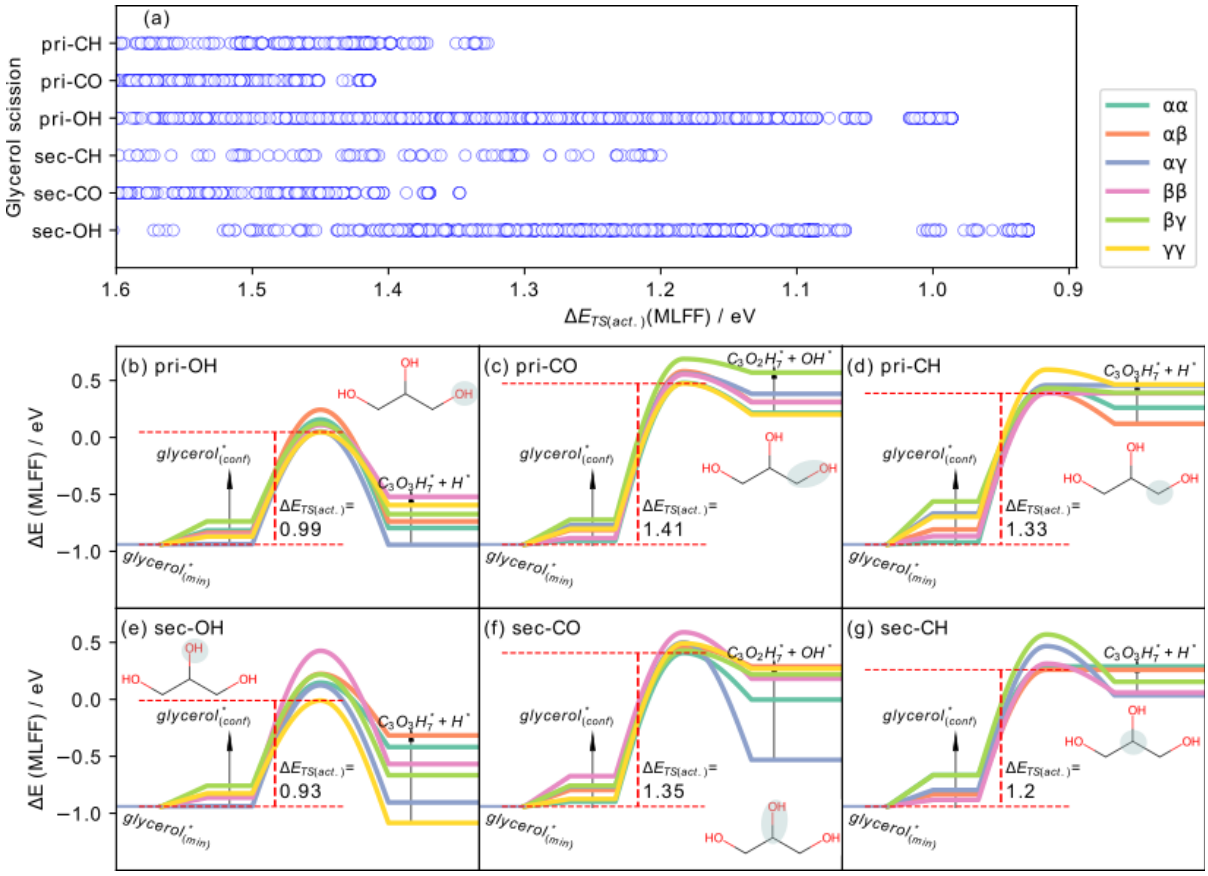


Figure 3: (a) MACE MLFF calculated activation energy barrier for glycerol bond scissions on Cu(111) surface. Calculated minimum energy pathways from αα, ββ, γγ, αβ, αγ and βγ glycerol conformers for primary (b) O-H, (c) C-OH and (d) C-H scission and secondary (e) O-H, (f) C-OH and (g) C-H scissions. The transition state energies ($\Delta E_{TS\{abs.\}}$) are computed in reference to minimum energy glycerol conformer (ΔE_{ads} = -0.94 eV).

Glycerol has six non-equivalent potential reactive bonds: primary and secondary C-O bonds, C-H bonds, and O-H bonds, that must be accounted for when constructing a reaction network. Here, we tackle the reaction network exploration for glycerol hydrodeoxygenation with both large conformational complexity and minimum energy path searches. From the ~16000 glycerol conformations, we select 1500 adsorbed glycerol structures as unique structures using the farthest point sampling[62] (Figure SI-1) to construct initial paths for calculating the minimum energy pathways for scission of the six glycerol bonds. The initial minimum energy paths were constructed and optimized using NEBs, followed by refining using ARPESS (Section 2.3) taking vibrational analysis into account. Again, the glycerol structures (reactant image) of minimum pathways obtained using intrinsic reaction coordinate (IRC) simulations

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 ORCID: https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. License: CC BY-NC 4.0

were categorized into six conformers (αα, αβ, αγ, ββ, βγ, γγ) based on C_3O_3 motifs (Figure 3, SI 5-10).

Figure 3(a) presents all the activation barriers calculated for the six glycerol bond scissions with activation barriers less than 1.8 eV with respect to the lowest energy glycerol structure on the Cu(111) surface. To ensure a comprehensive scan and to reduce the risk of missing any crucial details, we incorporated approximately 1500 glycerol configurations, as described in the previous section, into our transition state scans. This approach aims to identify the most favorable activation barrier. Given the automated nature of our workflow, we occasionally encountered chemically similar configurations due to minor variations, such as slight translations or rotations of the glycerol molecules on the Cu surface. These similarities were easily identified through a combination of unsupervised machine learning and expert human inspection at various stages. Rather than eliminating these redundancies with strict similarity checks from the outset, we allowed for some overlap, considering the complex reaction network of this molecule. This decision was particularly influenced by the computational efficiency of optimizing the NEB calculations using the MLFF method, with each calculation taking approximately 45 minutes on a single CPU.

Figure 3b-g shows the calculated minimum energy pathways for the scission of glycerol bonds with the lowest activation barriers, relative to the lowest-energy glycerol configuration ($\Delta E_{TS\{abs.\}}$). Secondary O-H scission has the lowest activation barrier ($\Delta E_{TS\{abs.\}}$ = 0.93 eV) followed by primary O-H scission ($\Delta E_{TS\{abs.\}}$ = 0.99 eV). Primary C-OH, secondary C-OH and primary C-H scission have higher activation barriers ($\Delta E_{TS\{abs.\}}$ = 1.41 eV, 1.35 eV and 1.33 eV, respectively), and hence are highly unlikely to occur on a Cu(111) surface. The breaking of the secondary C-H bond displayed both a strongly endothermic reaction energy (ΔE_{rxn} = 1.0 eV) and a high activation barrier ($\Delta E_{TS\{abs.\}}$ = 1.2 eV). Therefore, secondary C-H scission is likely to occur only with high reversibility in backward reaction. The activation energy barriers calculated in this study (by MACE-MLFF and verified using DFT-SCF calculations) are lower than those reported in the literature, although with similar trends. For instance, Zhang et al. (2019) reported an activation barrier of 1.12 eV for the breaking of the secondary O-H bond on a Cu(111) surface, which is reasonably close considering the differences in reactant conformations. Similarly, Liu and Greeley (2013) reported a 0.66 eV difference in activation barriers between the breaking of primary O-H ($\Delta E_{TS\{abs.\}}$ = 0.8 eV) and primary C-OH ($\Delta E_{TS\{abs.\}}$ = 1.46 eV) bonds using scaling relations. In contrast, by explicitly optimizing the TS structures and thoroughly exploring the saddle points for each reaction, we report a

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

significantly lower difference of 0.34 eV and 0.42 eV between O-H:C-H and O-H:C-OH bond scissions, respectively. Therefore, while the trends of bond scissions are not altered due to enhanced sampling of minimum energy pathways and transition states, the absolute values and energy differences are affected by considering different glycerol structural effects. This will be critical when computing reaction rate constants for kinetic modelling to estimate the activity and selectivity of reactions.

We also observe that the minimum energy pathways to determine transition states are not always linked to the minimum energy glycerol conformer on the Cu(111) surface. Additionally, the initial glycerol conformer structure has a slightly higher effect on secondary O-H bond scission as compared to primary O-H scission (SI Figure 5 and 8). For secondary O-H scission, the difference in activation barriers originating from γγ and ββ glycerol conformer structures is 0.44 eV. While primary O-H scission has only 0.19 eV difference between the lowest activation barriers of different conformers. Therefore, this suggests that catalytic surfaces that favor a high number of glycerol structures with β dihedral, might be more selective to primary O-H scission as compared to secondary O-H scission.

3.2. Glycerol HDO reaction mechanisms to PDO

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

Glycerol deoxygenation can occur through several pathways, as shown in Figure 4: (1) a two-step direct HDO process (P1 and P7 pathway),[67] (2) dehydration to enol and tautomerization to either hydroxyacetone or 3-hydroxypropionaldehyde, followed by hydrogenation to PDO (P2 and P6 pathway)[6], and (3) dehydrogenation to either glyceraldehyde or dihydroxyacetone, followed by dehydration and hydrogenation (as illustrated in P3, P4 and P5)[68]. These pathways are generally accepted in the scientific community, as the presence of glyceraldehyde and hydroxyacetone has been observed during experimental studies of glycerol HDO on metal catalysts.61

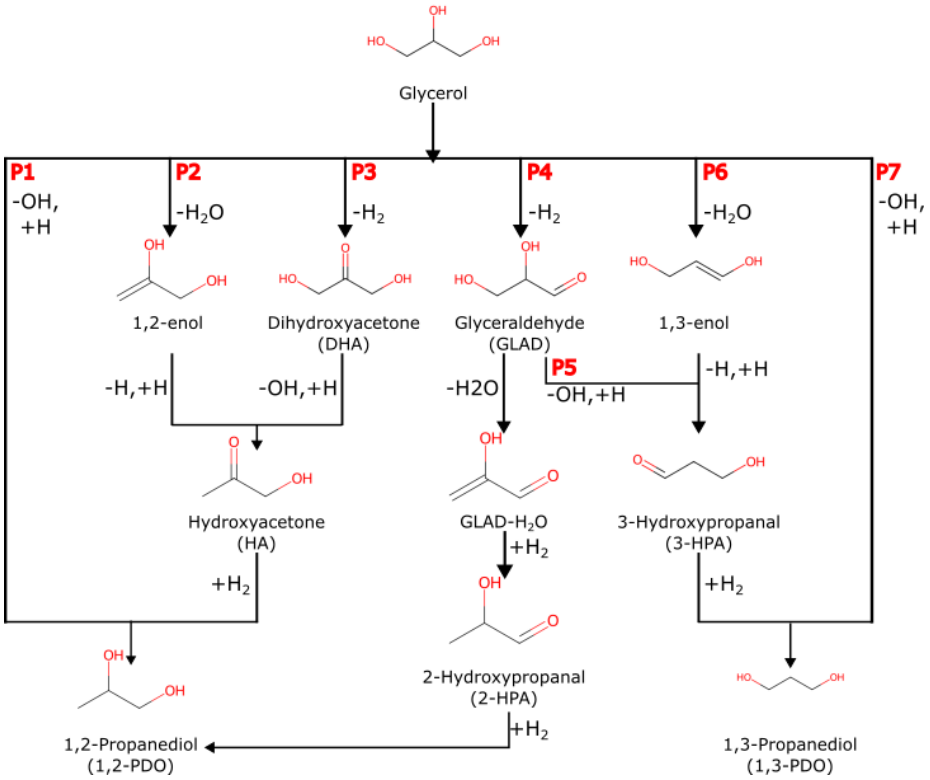


Figure 4**:** Proposed reaction mechanisms for glycerol deoxygenation to propane-diol (PDO).

The reaction network of glycerol HDO on a heterogeneous catalytic surface compares the formation of 4 primary intermediates (1,2-enol, 1,3-enol, glyceraldehyde and dihydroxyacetone), along with 1,2-PDO and 1,3-PDO formation through the direct method (Figure 4). Here, we extend our conformational sampling and minimum energy pathway search for intermediates of other reactions involved in glycerol hydrodeoxygenation reaction to 1,2-PDO as well as 1,3-PDO on a Cu(111) surface, similar to the search described in section 3.1. Again, for each intermediate ~1500 adsorbed structures were selected using the farthest point sampling from conformers obtained using minima hopping exploration. The minimum energy

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

paths using all 1500 structures were constructed and optimized using NEB followed by ARPESS refinement (Section 2.7).

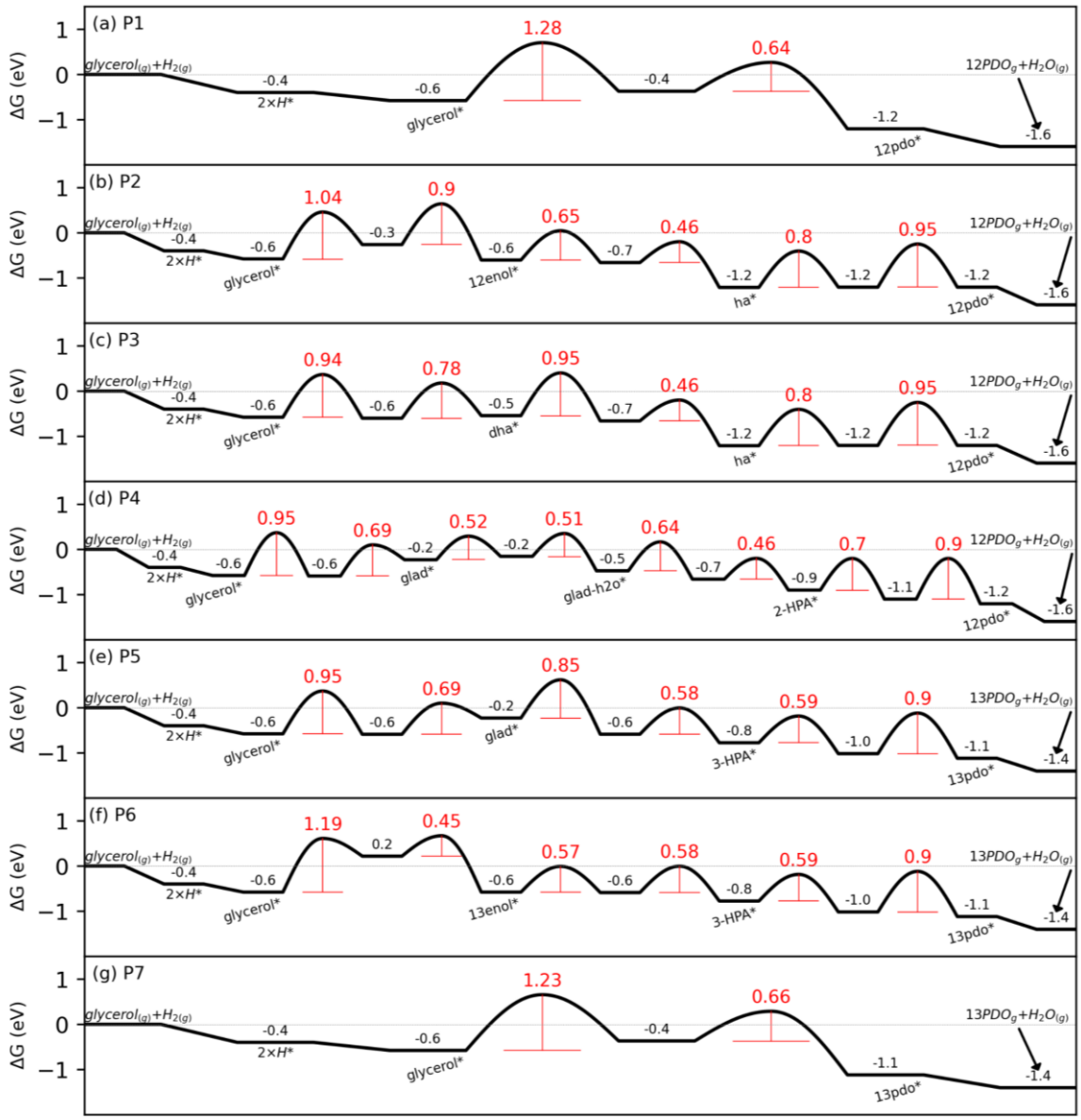


Figure 5: The P1-P7 reaction pathway for glycerol HDO to 1,2-PDO(a-d) and 1,3-PDO(e-g) on Cu(111) surface calculated using MACE MLFF. The black text indicates the relative energy of the intermediate with respect to the initial state (glycerol and H_2 in gas phase). The red text indicates the Gibb's energy of activation for all reaction steps.

In Figure 5, we provide the Gibb's energy profiles for the lowest-energy transition states and intermediates calculated at 473 K as described in Section 2.2.

As seen in section 3.1.2, the scission of the primary C-OH bond ($\Delta G_{act.}$ = 1.34 eV) and secondary C-OH bond ($\Delta G_{act.}$ = 1.33 eV) exhibits the highest Gibb's energy of activation among the bond breaking reactions. The scission of the secondary C-H and primary C-H bonds,

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 ORCID: https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. License: CC BY-NC 4.0

which represent alternative pathways for glycerol dehydration (P2 and P6), displayed both a high Gibb's energy for reaction (ΔG_{rxn} = 0.67 eV) and a high Gibb's energy for activation ($\Delta G_{act.}$ = 1.14 and 1.28 eV, respectively), unlike the primary and secondary O-H bond scission with barriers of 0.94 eV and 0.96 eV, respectively. Consequently, we conclude that only P3-P5 pathways are viable routes for the conversion of glycerol HDO to 1,2-PDO and 1,3-PDO on the Cu (111) surface and hence these are discussed further.

By following the P3 pathway, initiated by the secondary O-H scission ($\Delta G_{act.}$ = 0.94 eV), the Gibb's energy for activation for the scission of the secondary C-H bond and subsequent formation of dihydroxyacetone on the Cu(111) surface is 0.78 eV ($\Delta G_{rxn.}$ = 0.1 eV). Dihydroxyacetone is determined to be the most likely product of glycerol dehydrogenation on the Cu(111) surface, as reported in the literature.[70] Interestingly, while C-OH bond cleavage is unlikely for glycerol due to high Gibb's energy for activation, it is found to be more likely to occur for carbonylic compounds. The Gibb's energy for activation for the primary C-OH cleavage from dihydroxyacetone to form the precursor to hydroxyacetone is only 0.95 eV. Subsequent hydrogenation of hydroxyacetone led to the formation of 1,2-PDO. Note that all elementary steps further in this reaction pathway have Gibb's energy for activation lower than 1 eV. It is important to note that this pathway does not allow for the formation of 1,3-PDO as dihydroxyacetone only contains primary hydroxyl groups that can be cleaved. The initial dehydrogenation step protects the secondary oxygen from potential cleavage and lowers the likelihood of 1,3-PDO formation.

Following primary O-H bond scission ($\Delta G_{act.}$ = 0.94 eV, P4 and P5 pathway), the Gibb's energy for activation for the primary C-H bond scission to form glyceraldehyde is 0.69 eV ($\Delta G_{rxn.}$ = 0.4 eV). Investigation of the dehydration of glyceraldehyde for PDO conversion reveals that there are two possible bond scissions for the glyceraldehyde molecule on the surface. These include secondary C-H bond scission and secondary C-OH bond scission. The Gibb's energy for activation for the scission of the secondary C-H and C-OH bonds are 0.52 eV and 0.85 eV, respectively, relative to the minimum energy configuration of glyceraldehyde on the Cu(111) surface. Thus, a favorable route to 1,2-PDO via glyceraldehyde is also likely to be through the low Gibb's energy for activation secondary C-H split. The scission of the primary C-OH bond from the product of this elementary step only requires 0.51 eV. However, despite the low Gibb's energy for activation for all the elementary steps in this mechanism to form 1,2-PDO, the formation of glyceraldehyde from glycerol is endothermic ($\Delta G_{rxn.}$ = 0.4 eV). This is likely to cause a reverse reaction flux via this mechanism.

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

Although the experimental data clearly shows that a Cu catalyst is selective towards 1,2-PDO, we also investigated the potential pathways for converting glycerol to 1,3-PDO to understand why it is a minor product. Following P5 pathway, the only possible route, we observed that the scission of the secondary C-OH bond in glyceraldehyde requires 0.33 eV higher Gibb's energy for activation compared to the scission of the secondary C-H bond. This indicates a preference for the formation of 1,2-PDO from glyceraldehyde. Therefore, the Cu(111) surface shows higher selectivity for 1,2-PDO over 1,3-PDO, which may be further influenced by other factors, particularly the presence of oxidic species during experiments. [71]

4. Outlook on Machine learned potential performance and speed.

The integration of machine learning (ML) into catalyst development has become a transformative approach, leveraging the high speed and accuracy of ab-initio calculations. However, current machine learning studies primarily focus on identifying novel catalysts, which rely on understanding critical reactions between key material features, such as scaling relationships with formation energies and Brønsted-Evans-Polanyi (BEP) relationships that correlate with catalytic activity.[25,72–78] The complexity of reaction networks, however, especially with longer-chain carbon molecules such as glycerol extends beyond these relationships, as conformational searches are often not fully considered.

This study explores a rigorous sampling for glycerol and intermediate structures and minimum energy pathways to comment on the role of conformational space of larger molecules like glycerol on reaction energetics. This meant unrestricted simulations above 15000 minimum energy configurations per reactant and 1000 minimum energy paths per reaction. Here, we have a total of 22 intermediates and 26 transition states. Considering the speed of current DFT simulations, this would not have been possible using only DFT. Even with more restrictive constraints, we still would have to consider around 100 intermediate structures (as shown in Section 3.1.1) and minimum energy pathways (section 3.1.2) with high levels of human interference at each step. This would also take years with the most advanced high-performance computers to complete. With MLFF potentials, a single relaxation took ~10 minutes on single CPU, i.e, 0.002% of the compute time taken by DFT relaxation simulations. Therefore, with 5 active learning iterations requiring recalculations, ~15000 DFT SCF simulations for development of training set and ML model training, computational costs with MLFF are a fraction (<0.01%) of the computational cost required for DFT-only calculations. For that

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

reason, as we approach more sophisticated reaction networks, we demonstrate a major advantage of adopting this active learning approach over relying solely on DFT simulations.

In section 3.2, we highlight that the crucial steps that drive the selectivity of the deoxygenation reaction towards 1,2-PDO over 1,3-PDO on the Cu (111) surface is determined by small changes in energetic landscape ($\Delta\Delta E < 0.1$ eV or 1.35 meV/atom), for which a high level of prediction accuracy was necessary. The accuracy of the MLFF model was required to (a) distinguish the energetics of conformer sampling, and (b) accurately compare the reaction energies and activation barriers for competing reaction pathways. The universal machine learning models trained on open QM databases (RMSE > 0.085 eV/atom) cannot be considered to produce reliable accuracies for sensitive chemistry required for conformational sampling. Matbench discovery,[79] a benchmark for comparing universal MLFF also describes these models as "triaging tools for effectively allocating compute budget in high-throughput DFT". In addition, by conducting the preliminary tests on the recently published mace-mp0 universal force field model,[40] it is evident that although an overall visual trend may appear promising, it is not suitable for this particular application (Figure SI-12 and Figure SI-13). However, it is important to note that this constitutes an out-of-domain extrapolation for the model, as it was solely trained on bulk lattice data from the materials project. The parity plot between our MACE MLFF calculated (ΔE_{MLFF}) vs. DFT calculated (ΔE_{DFT}) adsorption energies for glycerol conformers on the Cu(111) surface with reference to Cu(111) slabs and gas-phase glycerol molecule verifies that lowest energy conformers are achieved using MLFF. The highest error in conformer sampling is 0.08 eV (Figure SI-11). Similarly, the transition states of glycerol scissions have root mean square errors (RMSE) of less than 0.07 eV. Overall, our trained MLFF model accurately discerns 22 intermediates, and 26 transition states, reconfirmed with DFT, with an average RMSE of 0.048 eV (< 0.6 meV/atom total energy) embedded within an intricate network of seven competitive pathways. These errors reveal that the energy difference between competitive reactions should be more than 0.1 eV for reliable conclusions using MLFF, which has been taken into consideration here.

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

5. Conclusion:

The goal of this study was to investigate the detailed HDO reaction mechanism of glycerol to PDO on a Cu(111) surface including a comprehensive conformational sampling on glycerol adsorption and other intermediates. While the conformational sampling of gas phase glycerol molecules, predicting 75 conformers, is well-documented, its exploration in heterogeneous catalytic systems remains limited due to the high computational costs and constraints of DFT. Here, an active learning-based MACE Machine Learned Force Field is utilized to explore a complex network of 7 competing pathways starting from >16000 possible structures of glycerol on the Cu(111) surface.

Out of 16000 glycerol adsorbed structures, we find many shallow low energy potential wells comprised of 3382 adsorbed glycerol structures with similar energetics ($-0.94 < \Delta E_{ads} < -0.84$) and structurally unique conformations are present on the Cu(111) surface. The structural dissimilarity arises primarily from dihedral angles along the C_3O_3 backbone. Other factors effecting the dissimilarity are average distance from the Cu(111) surface, intramolecular H-bonds, constraints on C-C-C bond angles due to adsorption, among many smaller variations in bond lengths and angles. The conformational sampling is also essential to determine the magnitude of absolute differences between activation barriers of glycerol bond scission (C-H:C-OH:O-H). We also observed the β-phase dihedrals having higher activation barriers for glycerol O-H bond scissions as compared to α- and γ-phase dihedrals. This particularly affects the secondary O-H bond scission, where the β in either side of the glycerol structure (αβ, ββ and βγ) leads to a higher activation barrier. These observations re-enforce the importance of considering conformational sampling for describing such reaction mechanisms accurately at quantum mechanical level – a prerequisite for exploring new catalyst development. We found that glycerol is likely to form the dihydroxyacetone intermediate before undergoing C-O cleavage leading to the experimentally reported intermediate, acetol (hydroxyacetone). On the contrary, glyceraldehyde, another intermediate of glycerol dehydrogenation, shows better activity for C-H bond breakage, which is followed by primary C-O bond cleavage. Further hydrogenation of acetol and 3-HPA in P3 and P4 pathways, respectively, lead to 1,2-PDO formation. This study qualitatively explains the preference of 1,2 PDO over 1,3 PDO based on the explored pathways on the pure Cu(111) surface.

Finally, our approach here describes a need for advancing the active learning based MLFF technology to supplement the DFT studies for comprehensive investigations of complex

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

chemistries of larger molecules on heterogeneous catalysts with high accuracies in a fraction of time. Here, only ~16,000 DFT SCF calculations were required to estimate the energetics of an almost 400 times larger sampling space. The RMSE of less than 0.6 meV/atom for intermediates and transition states suggests that MLFF potentials can be used as a first pass for predictive estimation. The high computational cost gains and low energetic prediction errors should be encouraging enough for the catalyst community to embrace these novel approaches as integral part of their strategy. The final training dataset and ML models will be available upon publication, and we encourage other researchers to build on this.

6. Acknowledgements

Ajin Rajan was supported by the Indo-German Science & Technology Centre (IGSTC) 2023 PhD Industrial Exposure Fellowship (PIEF) and performed the work during his stay in BASF SE. The authors would like to thank Moritz Otto Haus, Stephan Andreas Schunk, Sebastian Weber, Piyush Ingale from BASF SE, Germany and hte GmbH, Heidelberg for insightful discussions. Frequent exchange of technical details on MLFF architecture with prof. Gábor Csányi, David Kovacs, Lars Schaaf, Harry Moore and Ilyes Batatia from the MACE developer group in University of Cambridge, UK are gratefully acknowledged. Jithin John Varghese acknowledges support from the Indian Institute of Technology Madras.

7. Supporting Information:

The Supporting information includes Gibbs energy calculation equations, comment on computational cost comparison, example of Quantum Espresso calculation and MACE training protocol, accuracy of other machine learning potentials present in literature, activation barriers of hydrogenation reactions.

8. Author contributions

Authors E.F., I.M. and S.D. jointly conceived the research. S.D designed and developed the active-learning protocols and trained the MLFF models. S.G., A.R., E.F. and T.G. generated the training data, tested and applied the MLFF model to do the calculations. S.G. analyzed the results and drew conclusions. S.D supervised the project, and all authors actively took part in discussions and drafting the manuscript.

9. Competing interests

All authors declare no financial or non-financial competing interests.

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

10. References

[1] C.S. Callam, S.J. Singer, T.L. Lowary, C.M. Hadad, Computational Analysis of the Potential Energy Surfaces of Glycerol in the Gas and Aqueous Phases: Effects of Level of Theory, Basis Set, and Solvation on Strongly Intramolecularly Hydrogen-Bonded Systems, J Am Chem Soc 123 (2001) 11743–11754. https://doi.org/10.1021/ja011785r.

[2] M.R. Monteiro, C.L. Kugelmeier, R.S. Pinheiro, M.O. Batalha, A. da Silva César, Glycerol from biodiesel production: Technological paths for sustainability, Renewable and Sustainable Energy Reviews 88 (2018) 109–122. https://doi.org/https://doi.org/10.1016/j.rser.2018.02.019.

[3] W. Wan, S.C. Ammal, Z. Lin, K.-E. You, A. Heyden, J.G. Chen, Controlling reaction pathways of selective C–O bond cleavage of glycerol, Nat Commun 9 (2018) 4612. https://doi.org/10.1038/s41467-018-07047-7.

[4] D. Judas, A. Fradet, E. Marechal, Mechanism of the thickening reaction of polyester resins: Study on models, Journal of Polymer Science: Polymer Chemistry Edition 22 (1984) 3309–3318. https://doi.org/https://doi.org/10.1002/pol.1984.170221146.

[5] J. Zhang, Study of poly(trimethylene terephthalate) as an engineering thermoplastics material, J Appl Polym Sci 91 (2004) 1657–1666. https://doi.org/https://doi.org/10.1002/app.13322.

[6] X. Zhang, G. Cui, H. Feng, L. Chen, H. Wang, B. Wang, X. Zhang, L. Zheng, S. Hong, M. Wei, Platinum–copper single atom alloy catalysts with high performance towards glycerol hydrogenolysis, Nat Commun 10 (2019) 1–12. https://doi.org/10.1038/s41467-019-13685-2.

[7] H. Zhao, L. Zheng, X. Li, P. Chen, Z. Hou, Hydrogenolysis of glycerol to 1,2-propanediol over Cu-based catalysts: A short review, Catal Today 355 (2020) 84–95. https://doi.org/https://doi.org/10.1016/j.cattod.2019.03.011.

[8] W. Mondach, S. Chanklang, P. Somchuea, T. Witoon, M. Chareonpanich, K. Faungnawakij, H. Sohn, A. Seubsai, Highly efficient TiO2-supported Co–Cu catalysts for conversion of glycerol to 1,2-propanediol, Sci Rep 11 (2021) 23042. https://doi.org/10.1038/s41598-021-02416-7.

[9] R. Singh, P. Biswas, P.K. Jha, Study of the Glycerol Hydrogenolysis Reaction on Cu, Cu–Zn, and Cu–ZnO Clusters, ACS Omega 7 (2022) 33629–33636. https://doi.org/10.1021/acsomega.2c05342.

[10] R. Mane, A. Potdar, Y. Jeon, C. Rode, Calcination Temperature Impacting the Structure and Activity of CuAl Catalyst in Aqueous Glycerol Hydrogenolysis to 1,2-Propanediol, Top Catal (2024). https://doi.org/10.1007/s11244-024-02032-5.

[11] N.K. Mishra, P. Kumar, V.C. Srivastava, U.L. Štangar, Synthesis of Cu-based catalysts for hydrogenolysis of glycerol to 1,2-propanediol with in-situ generated hydrogen, J Environ Chem Eng 9 (2021) 105263. https://doi.org/https://doi.org/10.1016/j.jece.2021.105263.

[12] R. Mane, Y. Jeon, C. Rode, A review on non-noble metal catalysts for glycerol hydrodeoxygenation to 1,2-propanediol with and without external hydrogen, Green Chem. 24 (2022) 6751–6781. https://doi.org/10.1039/D2GC01879A.

[13] A. Lete, R. Raso, L. García, J. Ruiz, J. Arauzo, Synthesis of ketones from glycerol and 1,2-propanediol using copper and nickel catalysts: Unraveling the impact of reaction phase and active metal, Fuel 371 (2024) 132001. https://doi.org/https://doi.org/10.1016/j.fuel.2024.132001.

[14] D.K. Pandey, P. Biswas, Review of the Development of Heterogeneous Catalysts for Liquid and Vapor Phase Hydrogenolysis of Glycerol to Propylene Glycol (1,2-Propanediol): State-of-the-Art

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

and Outlook, Energy & Fuels 37 (2023) 6879–6906. https://doi.org/10.1021/acs.energyfuels.2c03806.

[15] S.M. Pudi, P. Biswas, S. Kumar, B. Sarkar, Selective hydrogenolysis of glycerol to 1, 2-propanediol over bimetallic Cu-Ni Catalysts supported on γ-Al2O3, J Braz Chem Soc 26 (2015) 1551–1564. https://doi.org/10.5935/0103-5053.20150123.

[16] R. V Sharma, P. Kumar, A.K. Dalai, Selective hydrogenolysis of glycerol to propylene glycol by using Cu:Zn:Cr:Zr mixed metal oxides catalyst, Appl Catal A Gen 477 (2014) 147–156. https://doi.org/https://doi.org/10.1016/j.apcata.2014.03.007.

[17] Z. Wu, Y. Mao, M. Song, X. Yin, M. Zhang, Cu/boehmite: A highly active catalyst for hydrogenolysis of glycerol to 1,2-propanediol, Catal Commun 32 (2013) 52–57. https://doi.org/https://doi.org/10.1016/j.catcom.2012.12.006.

[18] H. Tan, M.N. Hedhill, Y. Wang, J. Zhang, K. Li, S. Sioud, Z.A. Al-Talla, M.H. Amad, T. Zhan, O.E. Tall, Y. Han, One-pot synthesis Of Cu/ZnO/ZnAl2O4 catalysts and their catalytic performance in glycerol hydrogenolysis, Catal. Sci. Technol. 3 (2013) 3360–3370. https://doi.org/10.1039/C3CY00661A.

[19] E.S. Vasiliadou, A.A. Lemonidou, Investigating the performance and deactivation behaviour of silica-supported copper catalysts in glycerol hydrogenolysis, Appl Catal A Gen 396 (2011) 177–185. https://doi.org/https://doi.org/10.1016/j.apcata.2011.02.014.

[20] Z.-J. Zhao, S. Liu, S. Zha, D. Cheng, F. Studt, G. Henkelman, J. Gong, Publisher Correction: Theory-guided design of catalytic materials using scaling relationships and reactivity descriptors, Nat Rev Mater 7 (2022) 671. https://doi.org/10.1038/s41578-020-0192-2.

[21] Y. Ma, J. Xiong, P. Zhang, Y. Li, S. Zhang, Z. Wang, L. Xu, H. Guo, K. Chen, Y. Wei, Recent Progress on Density Functional Theory Calculations for Catalytic Control of Air Pollution, ACS ES&T Engineering 4 (2024) 47–65. https://doi.org/10.1021/acsestengg.3c00208.

[22] C. Chu, D. Huang, S. Gupta, S. Weon, J. Niu, E. Stavitski, C. Muhich, J.-H. Kim, Neighboring Pd single atoms surpass isolated single atoms for selective hydrodehalogenation catalysis, Nat Commun 12 (2021) 5179. https://doi.org/10.1038/s41467-021-25526-2.

[23] X. Wu, M. Nazemi, S. Gupta, A. Chismar, K. Hong, H. Jacobs, W. Zhang, K. Rigby, T. Hedtke, Q. Wang, E. Stavitski, M.S. Wong, C. Muhich, J.-H. Kim, Contrasting Capability of Single Atom Palladium for Thermocatalytic versus Electrocatalytic Nitrate Reduction Reaction, ACS Catal 13 (2023) 6804–6812. https://doi.org/10.1021/acscatal.3c01285.

[24] G.H. Gu, J. Noh, S. Kim, S. Back, Z. Ulissi, Y. Jung, Practical Deep-Learning Representation for Fast Heterogeneous Catalyst Screening, J Phys Chem Lett 11 (2020) 3185–3191. https://doi.org/10.1021/acs.jpclett.0c00634.

[25] M. Khatamirad, E. Fako, C. Boscagli, M. Müller, F. Ebert, R. Naumann d'Alnoncourt, A. Schaefer, S.A. Schunk, I. Jevtovikj, F. Rosowski, S. De, A data-driven high-throughput workflow applied to promoted In-oxide catalysts for CO2 hydrogenation to methanol, Catal. Sci. Technol. 13 (2023) 2656–2661. https://doi.org/10.1039/D3CY00148B.

[26] D. Weijing, Z. Weihong, Z. Xiaodong, Z. Baofeng, C. Lei, S. Laizhi, Y. Shuangxia, G. Haibin, C. Guanyi, Z. Liang, S. Ge, The application of DFT in catalysis and adsorption reaction system, Energy Procedia 152 (2018) 997–1002. https://doi.org/https://doi.org/10.1016/j.egypro.2018.09.106.

[27] Shambhawi, O. Mohan, T.S. Choksi, A.A. Lapkin, The design and optimization of heterogeneous catalysts using computational methods, Catal. Sci. Technol. 14 (2024) 515–532. https://doi.org/10.1039/D3CY01160G.

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

[28] P.-L. Kang, Y.-F. Shi, C. Shang, Z.-P. Liu, Artificial intelligence pathway search to resolve catalytic glycerol hydrogenolysis selectivity, Chem. Sci. 13 (2022) 8148–8160. https://doi.org/10.1039/D2SC02107B.

[29] A. Rajan, A.P. Pushkar, B.C. Dharmalingam, J.J. Varghese, iScience Iterative multiscale and multi-physics computations for operando catalyst nanostructure elucidation and kinetic modeling, IScience 26 (2023) 107029. https://doi.org/10.1016/j.isci.

[30] V. Kapil, D.P. Kovács, G. Csányi, A. Michaelides, First-principles spectroscopy of aqueous interfaces using machine-learned electronic and quantum nuclear effects, Faraday Discuss. 249 (2024) 50–68. https://doi.org/10.1039/D3FD00113J.

[31] T. Zarrouk, R. Ibragimova, A.P. Bartók, M.A. Caro, Experiment-driven atomistic materials modeling: A case study combining XPS and ML potentials to infer the structure of oxygen-rich amorphous carbon, ArXiv Preprint ArXiv:2402.03219 (2024).

[32] L.L. Schaaf, E. Fako, S. De, A. Schäfer, G. Csányi, Accurate energy barriers for catalytic reaction pathways: an automatic training protocol for machine learning force fields, NPJ Comput Mater 9 (2023) 180. https://doi.org/10.1038/s41524-023-01124-2.

[33] L. Zhang, D.-Y. Lin, H. Wang, R. Car, W. E, Active learning of uniformly accurate interatomic potentials for materials simulation, Phys Rev Mater 3 (2019) 23804. https://doi.org/10.1103/PhysRevMaterials.3.023804.

[34] J.S. Smith, B. Nebgen, N. Lubbers, O. Isayev, A.E. Roitberg, Less is more: Sampling chemical space with active learning, J Chem Phys 148 (2018) 241733. https://doi.org/10.1063/1.5023802.

[35] N. Bernstein, G. Csányi, V.L. Deringer, De novo exploration and self-guided learning of potential-energy surfaces, NPJ Comput Mater 5 (2019) 99. https://doi.org/10.1038/s41524-019-0236-6.

[36] Q. Lin, L. Zhang, Y. Zhang, B. Jiang, Searching Configurations in Uncertainty Space: Active Learning of High-Dimensional Neural Network Reactive Potentials, J Chem Theory Comput 17 (2021) 2691–2701. https://doi.org/10.1021/acs.jctc.1c00166.

[37] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M.B. Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A.D. Corso, S. de Gironcoli, P. Delugas, R.A. DiStasio, A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer, U. Gerstmann, F. Giustino, T. Gorni, J. Jia, M. Kawamura, H.-Y. Ko, A. Kokalj, E. Küçükbenli, M. Lazzeri, M. Marsili, N. Marzari, F. Mauri, N.L. Nguyen, H.-V. Nguyen, A. Otero-de-la-Roza, L. Paulatto, S. Poncé, D. Rocca, R. Sabatini, B. Santra, M. Schlipf, A.P. Seitsonen, A. Smogunov, I. Timrov, T. Thonhauser, P. Umari, N. Vast, X. Wu, S. Baroni, Advanced capabilities for materials modelling with Quantum ESPRESSO, Journal of Physics: Condensed Matter 29 (2017) 465901. https://doi.org/10.1088/1361-648X/aa8f79.

[38] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized Gradient Approximation Made Simple, Phys. Rev. Lett. 77 (1996) 3865–3868. https://doi.org/10.1103/PhysRevLett.77.3865.

[39] G. Prandini, A. Marrazzo, I.E. Castelli, N. Mounet, N. Marzari, Precision and efficiency in solid-state pseudopotential calculations, NPJ Comput Mater 4 (2018) 72. https://doi.org/10.1038/s41524-018-0127-2.

[40] I. Batatia, P. Benner, Y. Chiang, A.M. Elena, D.P. Kovács, J. Riebesell, X.R. Advincula, M. Asta, W.J. Baldwin, N. Bernstein, others, A foundation model for atomistic materials chemistry, ArXiv Preprint ArXiv:2401.00096 (2023) https://doi.org/10.48550/arXiv.2206.07697.

[41] I. Batatia, D.P. Kovacs, G. Simm, C. Ortner, G. Csányi, MACE: Higher order equivariant message passing neural networks for fast and accurate force fields, Adv Neural Inf Process Syst 35 (2022) 11423–11436. https://doi.org/10.48550/arXiv.2206.07697

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

[42] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, others, Pytorch: An imperative style, high-performance deep learning library, Adv Neural Inf Process Syst 32 (2019). https://doi.org/10.48550/arXiv.1912.01703.

[43] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J.P. Mailoa, M. Kornbluth, N. Molinari, T.E. Smidt, B. Kozinsky, E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials, Nat Commun 13 (2022) 2453. https://doi.org/10.1038/s41467-022-29939-5.

[44] K. Schütt, P.-J. Kindermans, H.E. Sauceda Felix, S. Chmiela, A. Tkatchenko, K.-R. Müller, Schnet: A continuous-filter convolutional neural network for modeling quantum interactions, Adv Neural Inf Process Syst 30 (2017). https://doi.org/10.48550/arXiv.1706.08566.

[45] L. Kabalan, I. Kowalec, C.R.A. Catlow, A.J. Logsdail, A computational study of the properties of low- and high-index Pd, Cu and Zn surfaces, Phys. Chem. Chem. Phys. 23 (2021) 14649–14661. https://doi.org/10.1039/D1CP01602D.

[46] R.L. Calabro, F.J. Burpo, S.F. Bartolucci, J.A. Maurer, Seed-Mediated Growth of Oxidation-Resistant Copper Nanoparticles, The Journal of Physical Chemistry C 127 (2023) 15307–15315. https://doi.org/10.1021/acs.jpcc.3c02369.

[47] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu, J Chem Phys 132 (2010). https://doi.org/10.1063/1.3382344.

[48] S. Grimme, S. Ehrlich, L. Goerigk, Effect of the damping function in dispersion corrected density functional theory, J Comput Chem 32 (2011) 1456–1465. https://doi.org/10.1002/jcc.21759.

[49] S. Takamoto, C. Shinagawa, D. Motoki, K. Nakago, W. Li, I. Kurata, T. Watanabe, Y. Yayama, H. Iriguchi, Y. Asano, T. Onodera, T. Ishii, T. Kudo, H. Ono, R. Sawada, R. Ishitani, M. Ong, T. Yamaguchi, T. Kataoka, A. Hayashi, N. Charoenphakdee, T. Ibuka, Towards universal neural network potential for material discovery applicable to arbitrary combination of 45 elements, Nat Commun 13 (2022) 2991. https://doi.org/10.1038/s41467-022-30687-9.

[50] S. Makri, C. Ortner, J.R. Kermode, A preconditioning scheme for minimum energy path finding methods, J Chem Phys 150 (2019). https://doi.org/10.1063/1.5064465.

[51] P. Lindgren, G. Kastlunger, A.A. Peterson, Scaled and Dynamic Optimizations of Nudged Elastic Bands, J Chem Theory Comput 15 (2019) 5787–5793. https://doi.org/10.1021/acs.jctc.9b00633.

[52] G. Henkelman, H. Jónsson, Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points, J Chem Phys 113 (2000) 9978–9985. https://doi.org/10.1063/1.1323224.

[53] G. and J.H. Henkelman Graeme and Jóhannesson, Methods for Finding Saddle Points and Minimum Energy Paths, in: S.D. Schwartz (Ed.), Theoretical Methods in Condensed Phase Chemistry, Springer Netherlands, Dordrecht, 2002: pp. 269–302. https://doi.org/10.1007/0-306-46949-9_10.

[54] H. Jónsson, G. Mills, K.W. Jacobsen, Nudged elastic band method for finding minimum energy paths of transitions, in: Classical and Quantum Dynamics in Condensed Phase Simulations, World Scientific, 1998: pp. 385–404. https://doi.org/10.1142/9789812839664_0016.

[55] A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I.E. Castelli, R. Christensen, M. Dułak, J. Friis, M.N. Groves, B. Hammer, C. Hargus, E.D. Hermes, P.C. Jennings, P. Bjerre Jensen, J. Kermode, J.R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K.S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, K.W. Jacobsen, The

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

atomic simulation environment—a Python library for working with atoms, Journal of Physics: Condensed Matter 29 (2017) 273002. https://doi.org/10.1088/1361-648X/aa680e.

[56] P.N. Plessow, Efficient Transition State Optimization of Periodic Structures through Automated Relaxed Potential Energy Surface Scans, J Chem Theory Comput 14 (2018) 981–990. https://doi.org/10.1021/acs.jctc.7b01070.

[57] S. Maeda, Y. Harabuchi, Y. Ono, T. Taketsugu, K. Morokuma, Intrinsic reaction coordinate: Calculation, bifurcation, and automated search, Int J Quantum Chem 115 (2015) 258–269. https://doi.org/https://doi.org/10.1002/qua.24757.

[58] B. Liu, Decomposition Pathways of Glycerol via C À H , O À H , and C À C Bond Scission on Pt (111): A Density Functional Theory Study, (2011) 19702–19709. https://doi.org/10.1021/jp202923w.

[59] K.-E. You, S.C. Ammal, Z. Lin, W. Wan, J.G. Chen, A. Heyden, Understanding the effect of Mo2C support on the activity of Cu for the hydrodeoxygenation of glycerol, J Catal 388 (2020) 141–153. https://doi.org/https://doi.org/10.1016/j.jcat.2020.05.007.

[60] K.-E. You, S.C. Ammal, Z. Lin, A. Heyden, Understanding Selective Hydrodeoxygenation of 1,2- and 1,3-Propanediols on Cu/Mo2C via Multiscale Modeling, ACS Catal 12 (2022) 4581–4596. https://doi.org/10.1021/acscatal.2c00261.

[61] A.A. Peterson, Global Optimization of Adsorbate–Surface Structures While Preserving Molecular Identity, Top Catal 57 (2014) 40–53. https://doi.org/10.1007/s11244-013-0161-8.

[62] S. De, A.P. Bartók, G. Csányi, M. Ceriotti, Comparing molecules and solids across structural and alchemical space, Phys. Chem. Chem. Phys. 18 (2016) 13754–13769. https://doi.org/10.1039/C6CP00415F.

[63] L. McInnes, J. Healy, J. Melville, Umap: Uniform manifold approximation and projection for dimension reduction, ArXiv Preprint ArXiv:1802.03426 (2018). https://doi.org/10.48550/arXiv.1802.03426.

[64] Y. Nishida, R. Aono, H. Dohi, W. Ding, H. Uzawa, 1H-NMR Karplus Analysis of Molecular Conformations of Glycerol under Different Solvent Conditions: A Consistent Rotational Isomerism in the Backbone Governed by Glycerol/Water Interactions, Int J Mol Sci 24 (2023). https://doi.org/10.3390/ijms24032766.

[65] H. van Koningsveld, The crystal structure of glycerol and its conformation, Recueil Des Travaux Chimiques Des Pays-Bas 87 (1968) 243–254. https://doi.org/https://doi.org/10.1002/recl.19680870303.

[66] G. Fraux, R. Cersonsky, M. Ceriotti, Chemiscope: interactive structure-property explorer for materials and molecules, 5 (2020). https://doi.org/10.21105/joss.02117.

[67] A. Rajan, J.J. Varghese, Towards selective glycerol hydrodeoxygenation to 1,3-propanediol with effective Pt-WOx catalyst design: Insights from first principles, J Catal 423 (2023) 94–104. https://doi.org/https://doi.org/10.1016/j.jcat.2023.04.019.

[68] Y. Wang, J. Zhou, X. Guo, Catalytic hydrogenolysis of glycerol to propanediols: a review, RSC Adv. 5 (2015) 74611–74628. https://doi.org/10.1039/C5RA11957J.

[69] M. Besson, P. Gallezot, C. Pinel, Conversion of Biomass into Chemicals over Metal Catalysts, Chem Rev 114 (2014) 1827–1870. https://doi.org/10.1021/cr4002269.

[70] V.-L. Yfanti, D. Ipsakis, A.A. Lemonidou, Kinetic study of liquid phase glycerol hydrodeoxygenation under inert conditions over a Cu-based catalyst, React. Chem. Eng. 3 (2018) 559–571. https://doi.org/10.1039/C8RE00061A.

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

[71] M. Harisekhar, V. Pavan Kumar, S. Shanthi Priya, K.V.R. Chary, Vapour phase hydrogenolysis of glycerol to propanediols over Cu/SBA-15 catalysts, Journal of Chemical Technology & Biotechnology 90 (2015) 1906–1917. https://doi.org/https://doi.org/10.1002/jctb.4728.

[72] Z. Han, S. Xia, Z. Chen, Y. Guo, Z. Li, Q. Huang, X.-J. Liu, W.-W. Xu, Facilitated the discovery of new γ/γ′ Co-based superalloys by combining first-principles and machine learning, NPJ Comput Mater 10 (2024) 259. https://doi.org/10.1038/s41524-024-01455-8.

[73] L.-F. Zhu, F. Körmann, Q. Chen, M. Selleby, J. Neugebauer, B. Grabowski, Accelerating ab initio melting property calculations with machine learning: application to the high entropy alloy TaVCrW, NPJ Comput Mater 10 (2024) 274. https://doi.org/10.1038/s41524-024-01464-7.

[74] V. Zinkovich, V. Sotskov, A. Shapeev, E. Podryabinkin, Exhaustive search for novel multicomponent alloys with brute force and machine learning, NPJ Comput Mater 10 (2024) 269. https://doi.org/10.1038/s41524-024-01452-x.

[75] L. Pereyaslavets, G. Kamath, O. Butin, A. Illarionov, M. Olevanov, I. Kurnikov, S. Sakipov, I. Leontyev, E. Voronina, T. Gannon, G. Nawrocki, M. Darkhovskiy, I. Ivahnenko, A. Kostikov, J. Scaranto, M.G. Kurnikova, S. Banik, H. Chan, M.G. Sternberg, S.K.R.S. Sankaranarayanan, B. Crawford, J. Potoff, M. Levitt, R.D. Kornberg, B. Fain, Accurate determination of solvation free energies of neutral organic compounds from first principles, Nat Commun 13 (2022) 414. https://doi.org/10.1038/s41467-022-28041-0.

[76] D. Lemm, G.F. von Rudorff, O.A. von Lilienfeld, Machine learning based energy-free structure predictions of molecules, transition states, and solids, Nat Commun 12 (2021) 4468. https://doi.org/10.1038/s41467-021-24525-7.

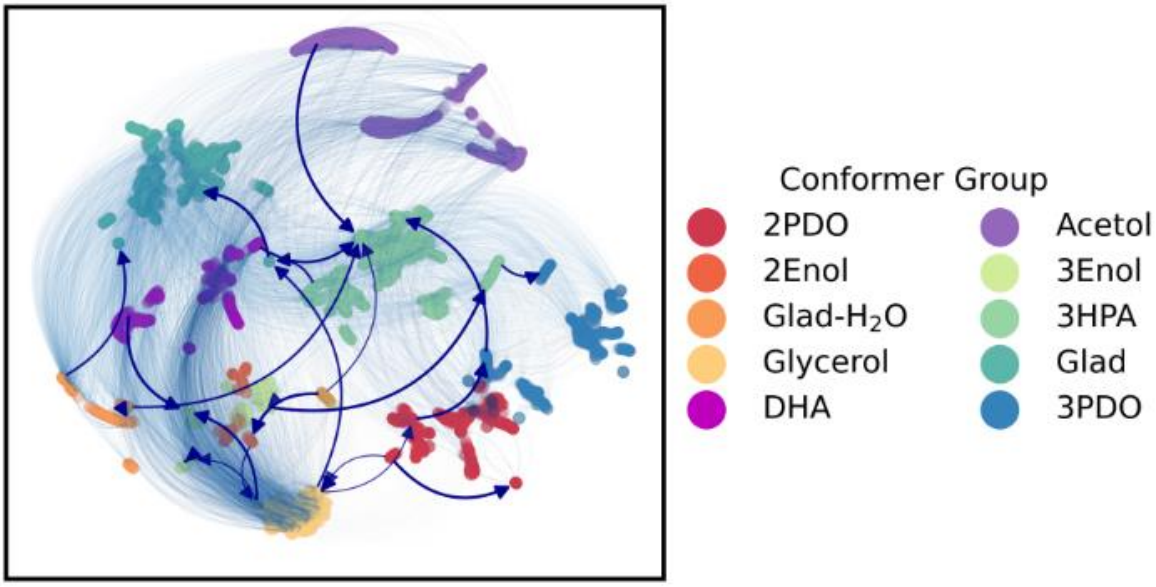
[77] A. Aldossary, J.A. Campos-Gonzalez-Angulo, S. Pablo-García, S.X. Leong, E.M. Rajaonson, L. Thiede, G. Tom, A. Wang, D. Avagliano, A. Aspuru-Guzik, In Silico Chemical Experiments in the Age of AI: From Quantum Chemistry to Machine Learning and Back, Advanced Materials 36 (2024) 2402369. https://doi.org/https://doi.org/10.1002/adma.202402369.

[78] Y. Ma, X. Zhang, L. Zhu, X. Feng, J.A.H. Kowah, J. Jiang, L. Wang, L. Jiang, X. Liu, Machine Learning and Quantum Calculation for Predicting Yield in Cu-Catalyzed P–H Reactions, Molecules 28 (2023). https://doi.org/10.3390/molecules28165995.

[79] J. Riebesell, R.E.A. Goodall, P. Benner, C. Yuan, B. Deng, G. Ceder, M. Asta, A.A. Lee, A. Jain, K.A. Persson, Matbench Discovery – A framework to evaluate machine learning crystal stability predictions, in: 2023. https://api.semanticscholar.org/CorpusID:261276994.

https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

For Table of Contents Only



https://doi.org/10.26434/chemrxiv-2024-34mq8-v2 **ORCID:** https://orcid.org/0000-0001-8434-3497 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0