Exploring the Intricacies of Glycerol Hydrodeoxygenation to Propanediol on Cu Catalysts: A Comprehensive Investigation with the Aid of Machine Learning Forcefield

Srishti Gupta[a], Ajin Rajan[b], Edvin Fako[c], Tiago J. Goncalves[c], Imke B. Mueller[c], Jithin John Varghese[b], Sandip De[c],*.

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

[b] A. Rajan, Dr. J. J. Varghese
Department of Chemical Engineering,
Indian Institute of Technology, Madras
600036, Chennai (India)

[c] Dr. E. Fako, Dr. T. J. Goncalves, Dr. I. B. Mueller, Dr. S. De
BASF SE, Group Research
Carl-Bosch-Straße 38
67056, Ludwigshafen (Germany)
* Email: sandip.de@basf.com

Abstract

Utilization of biomass to feedstock chemical relies on transforming hydroxyl containing molecules, as the hydroxyl group is found on the backbone of bio-molecules. For example, glycerol can undergo a hydro-deoxygenation reaction to produce propanediol, a valuable chemical precursor. This reaction captures the complexity and challenges of modelling surface-reactivity of flexible organic molecules in heterogenous catalysis, where surface intermediates can have many configurations. High computational costs of Density Functional Theory(DFT) restrict exhaustive exploration of the factorial reaction space, leading to having limited insights of the hydrodeoxygenation mechanism and hindering rational catalyst design. We deploy a Machine-Learned-Force-Field (MLFF) driven approach to elucidate the complex reaction network involved in the hydro-deoxygenation of glycerol on Cu(111). We present the bond cleavage activity ranking order for glycerol, and other intermediates and identify reaction pathways resulting in 1,2-PDO formation while highlighting indications of its higher selectivity over 1,3-PDO. This investigation delivers a comprehensive exploration of the transformation process from glycerol to propanediol, addressing the existing knowledge deficit through an advanced active-learning based MLFF approach. Notably, following a mere four iterations, our trained MLFF model accurately discerns 26 transition-states reconfirmed with DFT with root-mean-squared-error of 0.056 eV (0.74 meV/atom total-energy) embedded within network of seven competitive pathways.

1. Introduction

Glycerol (propane-1,2,3-triol) is the main by-product of transesterification of triglycerides from biomass feedstocks to form biodiesel[1] and from soap production.[2] Due to its abundance and low cost, glycerol's upcycling to other value-added products such as propane-diol (PDO) has increasingly become an incentivized reaction for large-scale commercial process.[3] For that reason, hydrodeoxygenation (HDO) reaction to convert glycerol into PDO is an important research topic. The two isomeric products, 1,2-Propanediol (1,2-PDO) and 1,3-Propanediol (1,3-PDO) are both industrially relevant. 1,2-PDO is used as a raw material in polyester resin manufacturing.[4] Additionally, it serves as a precursor for D-lactic acid, a valuable chiral building block utilized in the asymmetric synthesis of D-amino acids[5], as well as in several other applications. 1,3-

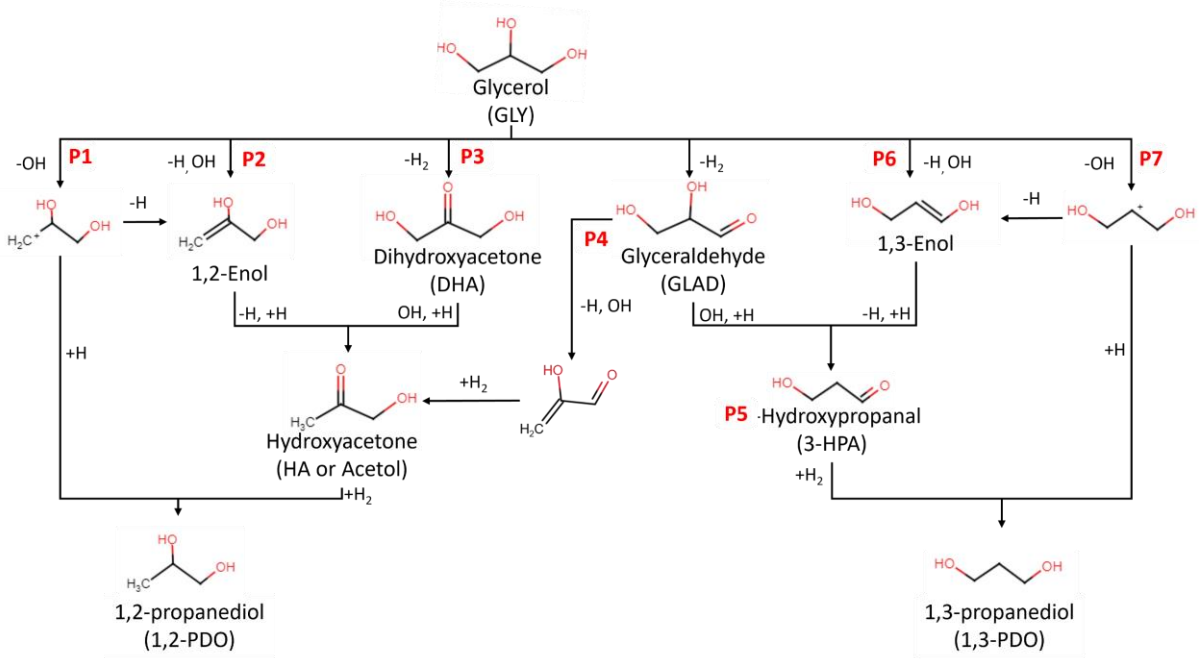

Figure 1: Proposed reaction mechanisms for glycerol deoxygenation to propane-diol (PDO).

PDO, on the other hand, is a primary ingredient in the manufacturing process of poly trimethylene terephthalate,[5] a polymer exhibiting properties of polyethylene terephthalate.[6] Currently, 1,2-PDO is synthesized by hydrating propylene oxide, whereas 1,3-PDO is synthesized either by hydroformylation and subsequent hydrogenation of ethylene oxide or by fermentation of glucose. The former requiring harsh conditions or expensive catalysts,[7] the latter requiring glucose feed quality tolerated by the bacteria.

Researchers have explored a wide range of metallic catalysts, such as Cu,[8] Ru,[9] Pd,[10] Rh,[10] and Pt,[11] on various supports for glycerol HDO reaction. Among all, Cu-based catalysts have gained higher interest due to their low cost, good performance and stability in gas phase reaction condition. In 2011, Bienholz and Claus,[12] demonstrated 87% activity of glycerol conversion to 68% acetol and 18% 1,2-PDO over Cu/SiO_2 surface. In 2009, Akiyama et. al.,[13] improved the efficiency of glycerol HDO process to achieve 100% activity and 78% selectivity to 1,2-PDO on Cu/Al_2O_3 by altering the H_2 flow rate and increasing the reaction temperature to 200 °C. In 2015, Harisekhar et. al.,[14] found that reducing Cu loading from 20 wt% to 5 wt% on Mesoporous silica increased the conversion of glycerol to 1-2PDO by 63%. Other studies have also focused on altering Cu-based catalysts, and reaction conditions to improve the selectivity of glycerol HDO to 1,2-PDO.[15]

Despite many experimental investigations delving into the development of a highly efficient and selective commercial Cu-based catalyst, rational design is hampered by the lack of progress made in understanding the HDO reaction mechanism of glycerol to 1,2-PDO. Glycerol with its flexible carbon backbone, has many energy-equivalent reactant conformers which can initialize the deoxygenation process.

The usual process of expert knowledge driven modelling workflows, highly based on human-mediated enumeration of various reaction and reactant possibilities, is both laborious, time-consuming, and has the risk of introducing human bias. Consider glycerol, which has 11 potential reactive bonds that must be accounted for when constructing a reaction network: two

primary C-O bonds, one secondary C-O bond, four primary C-H bonds, one secondary C-H bond, two primary O-H bonds, and one secondary O-H bond. Each of these bonds can initiate the deoxygenation process in the glycerol molecule (see Figure 1). Systematic high throughput exploration, though highly desired, has been fundamentally limited due to the high cost of *ab-initio* simulations namely Density functional theory (DFT) for each step, with number of steps quickly increasing due to combinatorial effects.

Glycerol deoxygenation can occur through several pathways, as shown in Figure 1: (1) a two-step direct HDO process (P1 and P7 pathway),[16] (2) dehydration to enol and tautomerization to either hydroxyacetone or 3-hydroxypropionaldehyde, followed by hydrogenation to PDO (P2 and P6 pathway)[17], and (3) dehydrogenation to either glyceraldehyde or dihydroxyacetone, followed by dehydration and hydrogenation (as illustrated in P3, P4 and P5 of Figuer 1)[18]. 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.[19]

There have been multiple research studies that have used ab-initio calculations to delve into the reaction mechanism of glycerol HDO, leading to differing outcomes. For instance, Jianfeng et al.,[20] used DFT simulations and reported the direct formation of 1,2-PDO from glycerol through primary OH cleavage followed by hydrogenation on the Cu(111) surface. Conversely, Xi et al.,[17] reported a lower energy barrier for primary C-H and O-H dissociation, suggesting a pathway involving glyceraldehyde for 1,2-PDO formation from glycerol on Cu(111)) as well. In another study, Weiming et al.,[3] reported that acetol formation occurred directly through multiple elementary steps on a 2D Cu cluster, supported on Mo_2C (0001).

These examples highlight a significant limitation of traditional approaches, where ab-initio calculations are the only workable tool: the severe under-sampling of the heterogeneously catalyzed reaction intermediates. This limitation is primarily due to the associated high costs and long execution times. Unfortunately, unlike the field of homogenous catalysis involving only (non-periodic) organic molecules where semi empirical methods[21] like Extended tight-binding (XTB)[22] can perform reasonably well, no generally reliable alternatives exist for heterogenous catalysis. The absence of studies comparing all competing reaction pathways for glycerol HDO on a Cu surface limits our present understanding of the key reaction steps that impact the activity of the catalytic surface significantly. This study aims to address this challenge by developing a Machine Learning Forcefield (MLFF)[23,24] with MACE: a higher-order equivariant message-passing neural network architecture combined with further development built on top of previously reported active learning protocols that iteratively improve the prediction accuracy of reaction intermediates and pathways, learning from reference DFT based calculations.

In the present work, our objective is to address the constraints of conventional DFT simulations by harnessing the benefits of MACE MLFF to analyze glycerol HDO to PDO on the Cu (111) surface. Our methodology in this context is to illustrate how current technology can be applied in an industrial research scenario, where achieving the ultimate scientific objective within tight timeline boundaries, without sacrificing the reliability of predictions is of highest concern. Therefore, the aim was not to build a minimal training set or spend resources to optimize the MLFF to squeeze out the last ounce of performance. Consequently, we adopt a cautious

approach, permitting adequate redundancy in the generation of DFT training data and opt for frequent explicit DFT validation.

With the orders of magnitude faster MLFF model, we could successfully delve into an extensive spectrum of glycerol conformers and reaction pathways, employing a blend of automated techniques (where conformations of molecules and intermediary surrogates were initialized with rdkit and active sites enumerated using catkit python packages)[25,26] and human-assisted methods for conformational and transition state exploration. 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

A common strategy for fitting MLFF involves an iterative approach, where the training set is expanded according to the application requirements of the force field.[27–29] The training set, consisting of relevant configurations, is used to train the MLFF by comparing properties with a reference method like DFT. The fitted MLFF is then used to generate new configurations and predictions, 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. To “teach” the model the unseen region, those configurations are added to the training set and the training process is restarted or repeated until 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,[30] 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 intermediates, compute energy and forces with single point DFT calculations- fit a MLFF- Optimize the postulated configurations or reaction pathways with MLFF - verify by DFT. With limited human intervention, mechanistic implications are considered to generate new relevant configurations that are added to the training set. This iterative process continues until the emerging pathway converges to the desired accuracy threshold.

2.2. Initial training data generation and base model building

We specifically focused on a simple surface (Cu (111)). Starting with the previously reported [30] protocol of initial training-set design, our approach includes a carefully constructed set of relevant molecules, surface and bulk atomic structures. Additionally, we focus on maximizing diversity and ensuring an appropriate statistical distribution of force, energy, and stress values. Considering higher factor of complexity in the present reaction schemes, we took additional steps. We initiated a large set (~1000) of reaction intermediates to account for the intricacy involved in these reactions. This was done using a combined approach of automation and expert knowledge-driven hypothesis generation for the entire reaction scheme presented in figure 1. 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 a smaller number of iterations, which with present technological constraints, means the quickest way to reach the desired accuracy of the final MLFF model. We computed single-point

DFT data for 9906 configurations. This initial set should already include sufficient information to study the flexible backbone with multi-reactive groups of glycerol in a highly branched reaction network. The base 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 a sense of expected accuracy, identifying any issues with the reference data generation and getting ready for the required targeted data generation for next iterations. We use this model to curate, prune and generate small amounts of 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.

2.3. Density Functional Theory simulations

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

Single-point periodic DFT calculations were performed for selected configurations using the Quantum Espresso software package,[31] and the Perdew-Burke-Ernzerhof (PBE)[32] exchange and correlation functional. Standard Solid-State Pseudopotentials (SSSP) [33] were 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 is converged when energy differences per electronic step are below 10^{-7} eV. The valence monoelectronic states were expanded as 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. The Cu catalyst was represented by (111) terminated slab consisting of 4x4x4 primitive units with a lattice constant of 3.61 Å. All slabs were separated by at least 20 Å of vacuum. Each slab has 4 layers in total, from which the bottom 2 were fixed.

2.4. ML training set-up

A higher-order equivariant message-passing neural network architecture, referred to as MACE[34,35] implemented in PyTorch[36] is employed to train a single Machine Learning Force Field (MLFF) that can describe all the reaction steps reported.[37,38] 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 use a 64-channel dimension for tensor decomposition. We use a radial cutoff of 6 Å and expand the interatomic distances into 10 Bessel functions multiplied by a smooth polynomial cutoff function to construct radial features, 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 is trained on a single NVIDIA A100 (80GB) GPU over the last 3 iterations (fresh start from iteration number 2 and then always restart with new train data for additional epochs) for a total 400 epochs (~22 GPU hours).

2.5. Model finetuning over three iterations

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 is then applied to explore the reaction network and collect new train data following the principle described before, followed by explicit DFT calculations resulting in new training data. Iteration 3 and 4 resume the training from the last checkpoints of the earlier iteration, expanding the training sets to 9782 and 10373 respectively. The validation set was kept fixed throughout these 3 iterations. The summary of the results is indicated in Table 1.

2.6. Applied Corrections

2.6.1. Dispersion correction

Long-range dispersion interactions (van der Waals, vdW), are crucial for describing the weak, long-range interactions between atoms. The additive property of vdW interactions is utilized in correction schemes such as DFT-D3[39] functioning as an interatomic potential which uses tabulated values of atomic polarizabilities to describe two-body (or three-body) dispersion interactions. This correction can be applied to DFT-computed configurations and to the MLFF, trained to PBE energies and forces. The same additive D3 dispersion correction (PyTorch[40] implementation) with a Becke-Johnson damping[41] is hence applied to DFT and MLFF calculations in the same manner and on the same footing.

2.6.2. 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 free 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, the harmonic approximation at temperatures of 473K (details are available in the SI-1). To evaluate the importance of such effects on adsorption, reaction and activation barriers, we first look at the secondary CH-scission reaction (Figure 5b, P2 pathway, first scission) from surface-bound glycerol (Gly*) to $C_3O_3H_7$*. The energies, referenced to the clean slab and gas-phase species, of Gly* (before the barrier) change from ΔE=-1.55 eV to ΔG=-0.39 eV and reactant $C_3O_3H_7$* (after the barrier) change from ΔE=-0.88 eV to ΔG=0.35 eV, yielding differences in reaction energies of ΔE_{rxn}=0.67 eV against ΔG_{rxn}=0.74 eV. The intrinsic reaction barrier of the aforementioned example changes from $\Delta E^{\ddagger}$ =1.24 eV to $\Delta G^{\ddagger}$ =1.16 eV (difference of only 0.08 eV). Similarly, the energies of the secondary OH-scission reaction (Figure 5c, P3 pathway, first scission) are $\Delta E_{C_3O_3H_7*}$=-1.72 eV against $\Delta G_{C_3O_3H_7*}$=-0.4 eV, resulting in ΔE_{rxn}=-0.17 eV against ΔG_{rxn}=-0.01 eV, whereas the intrinsic reaction barrier changes from $\Delta E^{\ddagger}$ =1.07 eV to $\Delta G^{\ddagger}$ =1.00 eV. By comparing both secondary scissions, we observe that accounting for thermodynamic corrections has no pronounced effects on intrinsic energy barriers or reaction energy differences between intermediate species. This means that conclusions drawn concerning preferred reaction pathways or which barrier is rate limiting, will not depend on discussing ΔE or ΔG. However, accounting for such corrections has been proven to be important to report more realistic reaction energies referenced to the clean slabs and gas-phase species, for higher temperatures in particular, than reporting energies discarding entropic effects altogether. For this reason, energy profiles will be reported in terms of Gibbs free energies (ΔG) throughout this work except for cases

where comparison of ΔE is the only valid choice.

2.7. Transition state calculations

For all the relaxed elementary step reactant conformers, Nudged Elastic Band (NEB) simulations were set up between the reactants and products. Products for bond breaking reactions (for e.g., OH scission of glycerol (Figure 2)) were created in a way that the broken product (OH) is placed on the neighboring Cu surface site. The neighboring surface sites as position of the broken product were selected from sites (top, hollow or bridge) at maximum distance of 2 Å from the original site.

Once the final intermediate was relaxed, the NEB pathways were created between the reactant and intermediate with 20 intermediate images using the idpp[42] 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 scans were performed using the Dynamic Nudge Elastic Band (DyNEB) [43–46] method as implemented in the atomic simulation environment (ASE)[47] with a maximum force threshold (fmax) of 0.05 eV/Å. At least 5 refined TS structures with minimum activation barriers are further refined to a tighter convergence of fmax=0.02 eV/Å with Automated Relaxed Potential Energy Surface Scans (ARPESS)[48] method. Obtained TS structures were confirmed by performing vibrational calculations with the MLFF, ensuring an imaginary frequency at the reaction vibrational mode (SI-1). Connection points of a given transition state were reconfirmed using the Intrinsic Reaction Coordinate (IRC)[49] method.

3. Results and Discussion

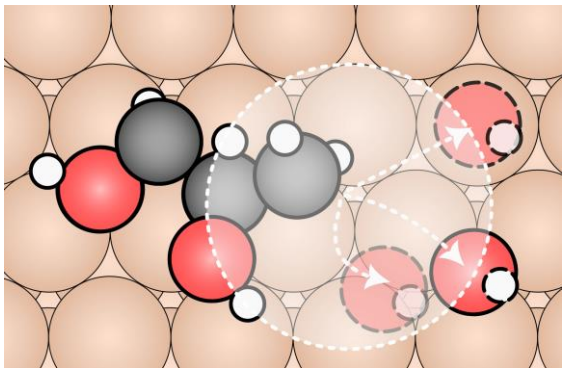

Figure 2: Postulated primary C-O scission of a glycerol molecule adsorbed on the Cu (111) surface. The white shaded circular area corresponds to the part of the surface that can potentially host the cleaved OH group. The dotted white arrows point towards potential top, bridge and hollow sites. (Colorcode: Cu – orange, C – gray, O – red and H - white)

In this study, we extensively investigated the isomeric structures of glycerol and the minimum energy path for glycerol primary C-OH bond scission on the Cu (111) surface in section 3.1. By comparing the reaction free energies (ΔG_{rxn}) and activation free energy barriers (ΔG_{act}) of the sampled surface intermediates for competing reactions, we identify the most likely mechanism for glycerol deoxygenation to 1,2-PDO (section 3.2). We highlight the crucial steps that drive the selectivity of the deoxygenation reaction towards 1,2-PDO over 1,3-PDO on the Cu (111) surface (section 3.3). The fact that the selectivity between these two products is determined by small changes in the energetic landscape, a high-level of prediction accuracy of approximately 0.05 eV in reaction energy (equivalent to <0.8 meV/atom total energy accuracy) was necessary. To accurately predict these complex reactions, we utilized the state-of-the-art MACE architecture, to train a MLFF which underwent iterative refinement. Only four iterations, following an active learning protocol (where reaction mechanism exploration was done in parallel with

dataset engineering, section 2.2) were enough to ensure the desired accuracy.

3.1. Exploration of glycerol conformers and transition state

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

Glycerol, a highly adaptable 3-Carbon molecule with a significant level of structural flexibility, is known to have 76 low-energy conformers within 0.4 eV in the gas phase. However, in relation to surface-induced alterations, the structural intricacies are considerably more complex, making reliable sampling arguably the most critical task. We utilized a workflow that utilizes a combination of domain knowledge, semi-automated adsorption site identification, enumeration of theoretical interaction possibilities, structural diversity ensuring sampling via use of unsupervised machine learning, molecular dynamics, geometric relaxation, and preliminary MLFF runs to derive 264 low-lying yet structurally diverse conformers on the Cu(111) surface from 330 initial glycerol configuration. We optimized these 264 glycerol conformers on the surface using MLFF until the maximum forces on atoms were decreased to less than 0.02 eV/Å. The distribution of the adsorption free energy is depicted in Figure 3c.

A supervised machine-learning based analysis derived representation as shown in Figure 3a, further ensured the structural diversity of the starting conformer pool. This method is based on encoding local atomic environments (rcut =3.0 Å) via SOAP [50–54] combined with Uniform Manifold approximation and Projection (UMAP).[55,56] We identified 25 distinct groups of conformers, underlining various surface and hydrogen bonding motifs and encapsulating a substantial spectrum of glycerol structures (Figure 3e). Due to the flexible nature of H atoms, many conformers exhibit similar C_3O_3 backbone motifs with minor alterations in H placement relative to the surface and other OH groups. These analogous backbone structures are visible as clustered groups in Figure 3a. Understanding the energetic hierarchy of the conformers and consequently deducing the most stable configurations is challenging (as seen in Figure SI-1), but certain trends can be inferred from the results. For instance, the six configurations within a 0.16 eV range of their relative electronic energy to the minimum configuration are all closely attached to the surface, either by extending across the surface or by forming two Cu-O bonds. The top five conformers display H-bonds from the binding OH to a different OH-group, and throughout the generated set of conformers, H-bonds from the binding oxygen to different OH-groups are slightly preferred over H-bonds to the binding oxygen or excluding it. Overall, the energetic ordering is dictated by the balance between surface contact and hydrogen bonding. As anticipated, more hydrogen bonds are advantageous, but surface contact competes with this basic principle, leading to different conformational motifs being closely matched in energy.

The MLFF-calculated, and verified with DFT (Figure 3d), lowest-energy conformer of glycerol on the Cu (111) surface (Figure 3e) has an adsorption free energy ΔG_{ads} of -0.40 eV (ΔE_{ads}=-0.91 eV). The glycerol conformer is adsorbed in a horizontal orientation with the secondary OH group nearest to the surface, at a distance of 2.19 Å from the Cu slab. A hydrogen bond is formed from the secondary OH to a primary OH group, with a second hydrogen bond from the remaining primary OH to the surface-

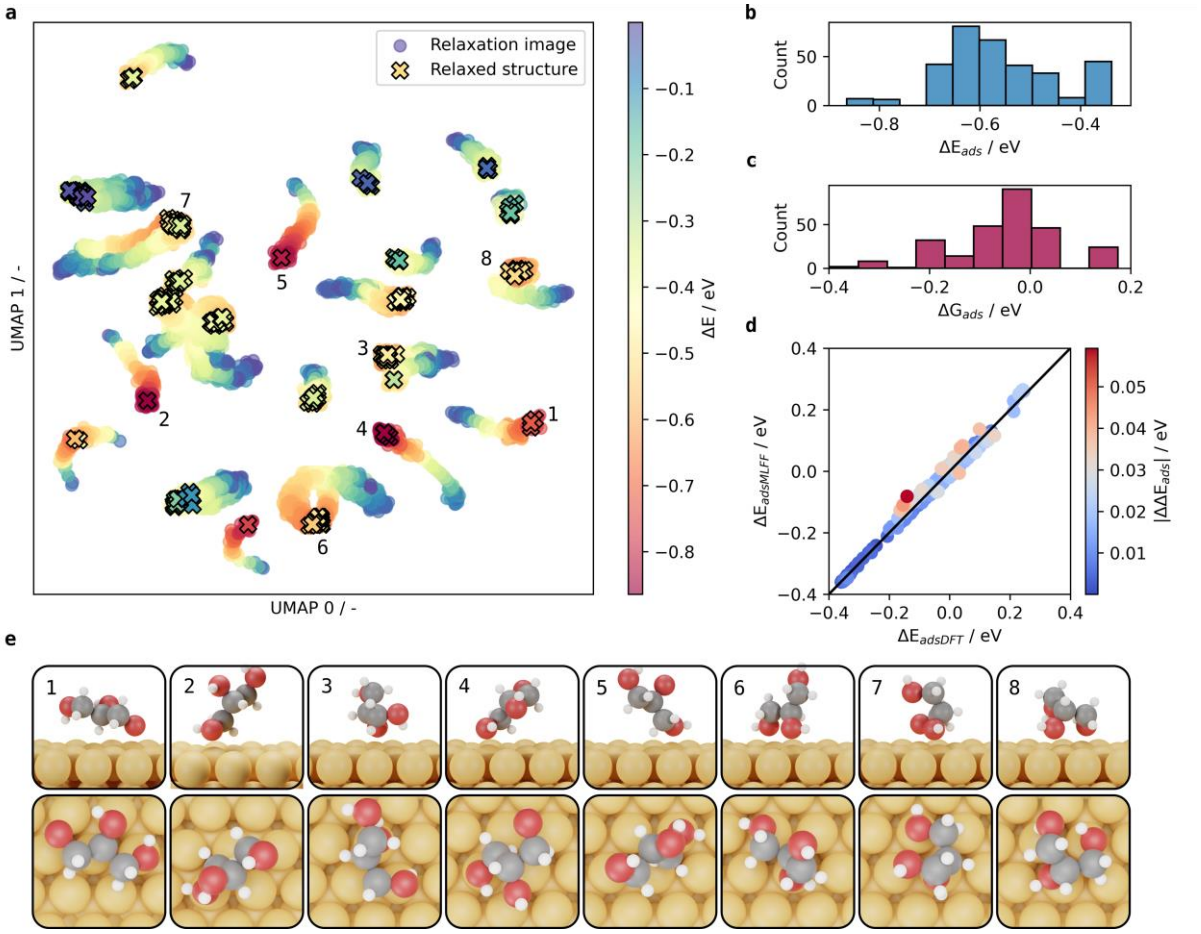

Figure 3: (a) Adsorption energy of explored 77153 configurations from 330 glycerol relaxation paths on the Cu(111) surface represented as a function of their structure similarity. The map has been constructed by first selecting 1000 structurally most diverse representative configurations using Farthest point sampling algorithm on computed SOAP descriptors followed by fitting a Uniform Manifold approximation and Projection (UMAP) model first to optimize the base representation and then projecting rest of the data on the map to arrive at the final 2-dimensional representation. The crosses indicate the 264 MACE MLFF relaxed glycerol configurations on the Cu(111) surface. The population density of (b) ΔE_{ads} and (c) ΔG_{ads} calculated for 264 relaxed glycerol structures on the surface. (d) The parity plot between MACE MLFF calculated ($\Delta E_{adsMLFF}$) vs. DFT calculated (ΔE_{adsDFT}) energies for glycerol conformers on the Cu(111) surface. The ΔE_{ads} calculated here using reference DFT calculated glycerol gas molecule and pristine Cu(111) surface energy. The ΔE_{MLFF} is calculated without vdw effects for comparison to DFT energies. (e) Side and top view of diverse glycerol conformers relaxed using MACE MLFF on the Cu(111) surface. The number notation corresponds to the crossed numbers specified in (a).

attached secondary OH. The primary OH groups are 3.08 Å and 3.52 Å away from the Cu (111) slab respectively. The next most stable conformer (with ΔE_{ads} = -0.86 eV) attaches to the surface via a primary hydroxyl group and consists of a hydrogen bond from the secondary hydroxyl to a non-attached primary hydroxyl.

Our research findings align well with previous studies. Specifically, a 2019 study by Zhang et al.,[17] found similar initial configurations of glycerol on the Cu (111)

surface and recorded an adsorption energy of -0.89 eV, which is comparable to our second-best conformer. Additionally, a 2023 report by Wang et al.,[57] identified the lowest-energy configuration of glycerol on the Cu (111) surface that closely matches our fourth-best conformer. They recorded an adsorption energy of ΔE_{ads} = -1.12 eV and a Cu-O(sec) distance of 2.18 Å. Finally, Liu and Greeley's[58] work, which did not include any dispersion corrections, found results comparable to our third-best conformer. As inconsistencies in energetic ordering within such close energetic span are expected to arise due to the different choice of DFT parameters and functionals, it further stresses the importance of keeping all the conformers for subsequent reaction studies.

3.1.2. Primary C-OH glycerol bond scission on Cu(111) surface

Glycerol has two primary C-OH bonds, both of which have an equal likelihood of undergoing C-O bond cleavage on the Cu (111) surface. In this section, we focus on the cleavage of the primary C-OH bond in glycerol as an example to demonstrate our step-by-step exploration of the activation barrier landscape using MLFF. This initial C-OH bond cleavage is a crucial step in glycerol deoxygenation, occurring through both the direct hydrogenolysis pathway (P1) and the dehydration-hydrogenation pathway (P2, see Figure 1). In subsequent sections we will only compare the lowest activation free energy barrier of the pathways for glycerol deoxygenation on Cu (111) with PDO.

The accuracy of the actual activation barrier is heavily influenced by the selection of glycerol conformers and the chosen C-OH scission Nudged Elastic Band (NEB) paths, as evidenced by previous literature. For example, Jianfeng et al.,[20] reported a 1.85 eV barrier for the primary C-O cleavage of glycerol on the Cu (111) surface, while Weiming et al.,[3] reported a 1.38 eV barrier for the primary C-O cleavage of glycerol on a 2D Cu cluster supported on the β-Mo_2C(0001) surface. Aside from differences in the catalyst model, the former study did not stabilize the C* from the product state on the surface, which could significantly affect the barrier and reaction energy. Liu and Greely (2013)[58] calculated a 1.46 eV activation barrier for primary C-OH scission on the Cu (111) surface using scaling relations, but their calculations did not consider van der Waals (vdW) effects, which influence the prediction of adsorption and desorption energies of hydrocarbon molecules on catalyst surfaces. All these observations further motivate us to systematically explore potential reaction pathways.

We employ a high throughput approach to identify and optimize transition states utilizing the MLFF, as detailed in section 2.2. To prioritize a thorough scan and minimize the risk of omitting any vital information, we included all 264 glycerol configurations described in the previous section in our transition-state scans, aiming to capture the most favorable activation barrier. Our focus was on the cleavage of either primary C-O bond and the displacement of the broken OH intermediate to the nearest hollow, top, or bridge sites within a radius of 2 Å from the original C-OH position (see Figure 2). We obtained 2683 product states in which the primary C-OH bond was broken on the Cu (111) surface. We eliminated product states with no C-O bond cleavage and those with reaction energies greater than 2 eV from the initial glycerol conformer, as a reaction energy of 2 eV was considered unreasonably high for further progression. This left us with 978 product states for primary C-OH bond cleavage, which were used to construct the NEB pathways. Each pathway consisted of 20 intermediate images that were optimized to fmax < 0.1 eV/Å. NEB paths with activation barriers exceeding 2.2 eV were discarded. Ultimately, we optimized 339 NEB paths for primary C-OH bond cleavage to fmax <

0.05 eV, as shown in Figure 4a with grey dashed lines. The diversity of bond scission along 339 NEB paths, also represented in Figure 4(b), changes the C-O bond length, which ranges between 2.29 – 3.5 Å for transition states.

The corresponding $\Delta E_{TS\{rel.\}}$ and $\Delta E_{TS\{act.\}}$ values for this conformer are 1.54 eV and 1.58 eV, respectively. However, the other C-OH bond cleavage for the same glycerol conformer exhibits a $\Delta E_{TS\{act.\}}$ value of 1.60 eV (figure 4f). The pathway for C-OH bond

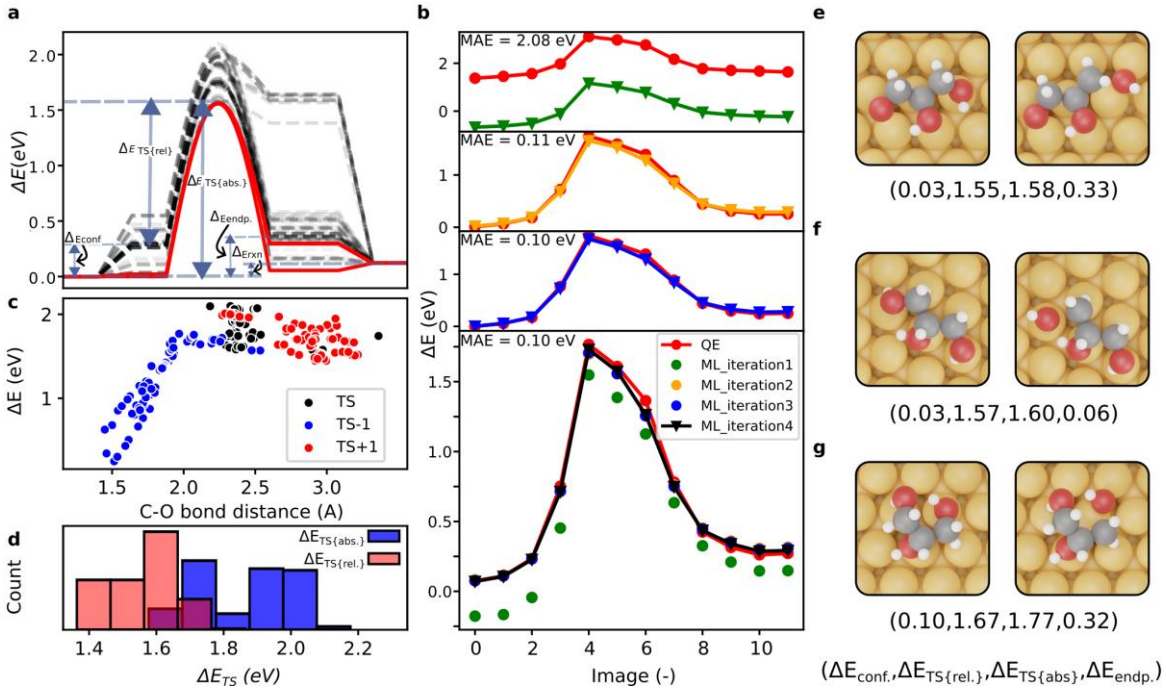

Figure 4: (a) MACE MLFF calculated activation barriers for primary C-O scission of glycerol optimized using NEB(grey) method till fmax <0.05 eV/Å. The 5 lowest TS are optimized using ARPESS(red) method till fmax < 0.02 eV/Å method, and their endpoints are recalculated using IRC method. (b) C-O scission NEB path optimized over 4 iterations of MLFF. The sub-figures (top to bottom) were relaxed with ML_iteration 1,2,3,4, respectively, and the corresponding DFT energies are indicated in red dots. The reference for all points was the DFT energy of glycerol conformer relaxed with ML_iteration 4 (image 0). (c) The C-O bond distance at transition state (TS), one image before (TS-1) and after (TS+1) the transition state vs. corresponding MLFF energies. (d) The distribution of relative activation barrier and absolute activation barrier for primary C-O scission. The glycerol conformer (left) and transition state (right) of minimum energy path (e) and other paths (f) and (g).

The five configurations with the lowest absolute activation energy barriers ($\Delta E_{TS\{abs.\}}$) for primary C-OH bond cleavage were further refined using the Automated Relaxed Potential Energy Surface Scans (ARPESS) method, as depicted in figure 4a. The minimum energy path (MEP) with the minimum MLFF energy barrier begins with a glycerol conformer with $\Delta E_{conf.}$ = 0.03 eV compared to the global minima reference (figure 4e).

cleavage also influences the endpoint energies ($\Delta E_{endp.}$) resulting from the C-O bond cleavage up to 1.5 eV, depending on intermolecular hydrogen bonding or the orientation of the glycerol intermediate on the Cu (111) surface (figure 4a).

The MLFF model went through 4 iterations with the first one being a very rough starting point and true active learning starting from second iteration where a model is re-trained from scratch with

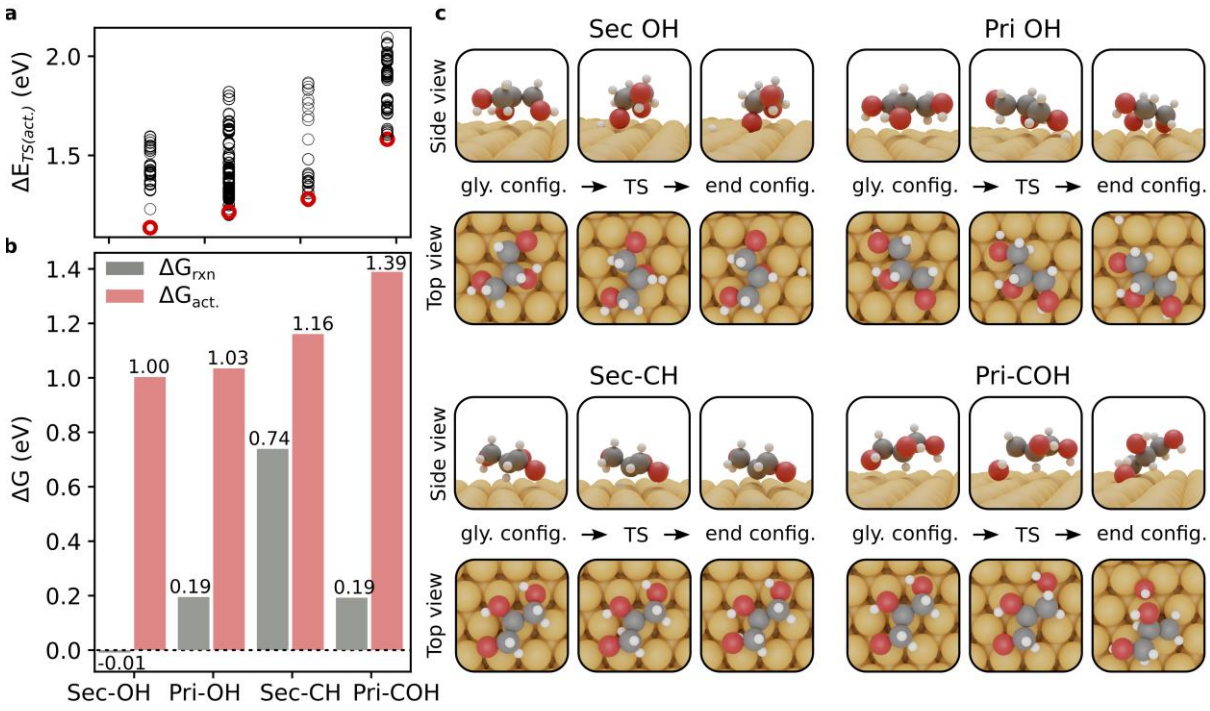

Figure 5: (a) MACE MLFF calculated activation energy barrier. The red circles indicate the lowest activation energy barriers. (b) Activation free energy and reaction free energies for glycerol primary C-OH, primary C-H, secondary C-H and secondary O-H scission. (c) Top and side view of the starting glycerol configuration, transition state and end configuration of glycerol intermediate on Cu(111) surface for scission reactions calculated using ARPESS and the IRC method.

additional data generated in iteration (figure 4b). Subsequent iterations restart the training process from the existing model with additional new data as we progress towards our desired accuracy, as highlighted in the figure. To validate our findings based on MLFF simulations, we did explicit DFT SCF calculations on key steps and considered geometrical parameters, specifically the primary C-O bond length of glycerol at and near the transition state. Figure 4b illustrates this chemically relevant quantity in the structure. Along the smallest activation energy path (figure 4d), the C-O bond distance changes from 1.44 Å (glycerol configuration) to 2.39 Å at the saddle point.

Due to the automated nature of our workflow, we observed the presence of chemically very similar configurations in a few cases, resulting from subtle variations such as minor translations or rotations of the glycerol molecules on the Cu surface or within the OH bond. These similarities could be easily identified through a combination of unsupervised ML and expert human inspection at various stages. Instead of eliminating them with strict similarity checks, we decided to allow for some redundancy considering the complex reaction network of this molecule. This decision was particularly influenced by the computationally efficient nature of optimizing Nudged Elastic Band (NEB) calculations using the MLFF method, with each calculation taking approximately 30 minutes on a single CPU.

3.2. Reaction HDO mechanisms for glycerol to 1,2-PDO

The conversion of glycerol via hydrogenation and deoxygenation (HDO) to 1,2-propanediol (1,2-PDO) has been proposed based on experimental observations, suggesting the involvement of P1-P4 reaction mechanisms (Figure 1).

Here, we conducted a comprehensive investigation of each reaction pathway to determine the most probable route for glycerol HDO on a Cu (111) catalyst surface. The calculation of reaction free energies and activation free energy barriers for intermediates and products followed the methodology described in section 2.2 and discussed in detail in section 3.1.

Figure 5a presents the calculated activation barriers ($\Delta E_{TS\{abs.\}}$) for the breaking of primary C-OH, secondary C-H, primary O-H, and secondary O-H bonds in configurations with the lowest absolute energy activation barriers ($\Delta E_{TS\{abs.\}}$) for glycerol bond breaking and the calculated activation free energy barriers ($\Delta G_{act.}$) and reaction free energies ($\Delta G_{rxn.}$) of the resulting products are shown in Figure 5b. Among these, the breaking of the secondary O-H bond exhibited the lowest activation free energy barrier ($\Delta G_{act.}$ = 1.00 eV), closely followed by the primary O-H bond breaking ($\Delta G_{act.}$ = 1.03 eV), relative to the energetically most favorable glycerol configuration adsorbed on the Cu (111) surface. Regarding the initial steps, the activation energy for O-H bond breaking was generally lower than that for C-H bond breaking, which in turn was lower than that for C-OH bond breaking. These trends in activation free energy barriers align with previous findings in the literature. For instance, Zhang et al. (2019)[17] 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)[58] reported a 0.66 eV difference in activation barriers between the breaking of primary O-H (0.8 eV) and primary C-OH (1.46 eV) bonds using scaling relations. In contrast, by explicitly optimizing the transition-state structures and thoroughly exploring the saddle points for each reaction, we report a significantly lower difference of 0.16 eV and 0.39 eV between O-H:C-H and O-H:C-OH bond scissions, respectively (Figure 5b). We determined that the breaking of the primary C-OH bond ($\Delta G_{act.}$ = 1.39 eV) exhibits the highest activation free energy among the bond breaking reactions. On the other hand, the breaking of the secondary C-H bond, which represents an alternative pathway for glycerol dehydration (P2, figure 1), displayed both a high reaction free energy (ΔG_{rxn} = 0.74 eV) and a high activation barrier ($\Delta G_{act.}$ = 1.16 eV), unlike the secondary O-H bond scission with a reaction free energy of -0.01 eV and a barrier of 1.00 eV. Consequently, we conclude that the P1 and P2 pathways are not viable routes for the conversion of glycerol HDO to 1,2-PDO on the Cu (111) surface (Figure 6).

By following the P3 pathway (figure 6c), the activation free energy barrier for the scission of the secondary C-H bond and subsequent formation of dihydroxyacetone on the Cu(111) surface is 0.93 eV (with ΔG_{rxn} = -0.19 – (-0.40) = +0.21 eV). Dihydroxyacetone is determined to be the most likely product of glycerol dehydrogenation on the Cu(111) surface. Interestingly, while C-OH bond cleavage is unlikely for glycerol due to high activation free energy barriers, it is found to be more likely to occur for carbonylic compounds. The activation free energy barrier for the primary C-OH cleavage from dihydroxyacetone to form the precursor to hydroxyacetone is only 0.91 eV (with ΔG_{rxn} = -0.32 eV). Subsequent hydrogenation of hydroxyacetone and further hydrogenation of the intermediate will lead to the formation of 1,2-PDO. It is important to note that this pathway does not allow for the formation of 1,3-PDO (discussed in next section) as dihydroxyacetone only contains primary hydroxyl groups that cannot be cleaved. The initial dehydrogenation step protects the secondary oxygen from potential cleavage.

In the P4 pathway, after the primary O-H bond scission, the activation free energy barrier for the primary C-H bond scission to form glyceraldehyde is 0.75 eV.

However, the reaction free energy for the transformation of glycerol to glyceraldehyde is endothermic by 0.48 eV. Therefore, we expect a reaction equilibrium occurring between glycerol and glyceraldehyde, which will limit the activation of further pathway after formation of glyceraldehyde.

on the surface. These include primary O-H bond scission, primary C-OH bond scission, secondary C-H bond scission, secondary C-OH bond scission, and secondary O-H bond scission. The activation free energy barriers for the scission of the secondary C-H and C-OH bonds are 0.58 eV and 0.88 eV,

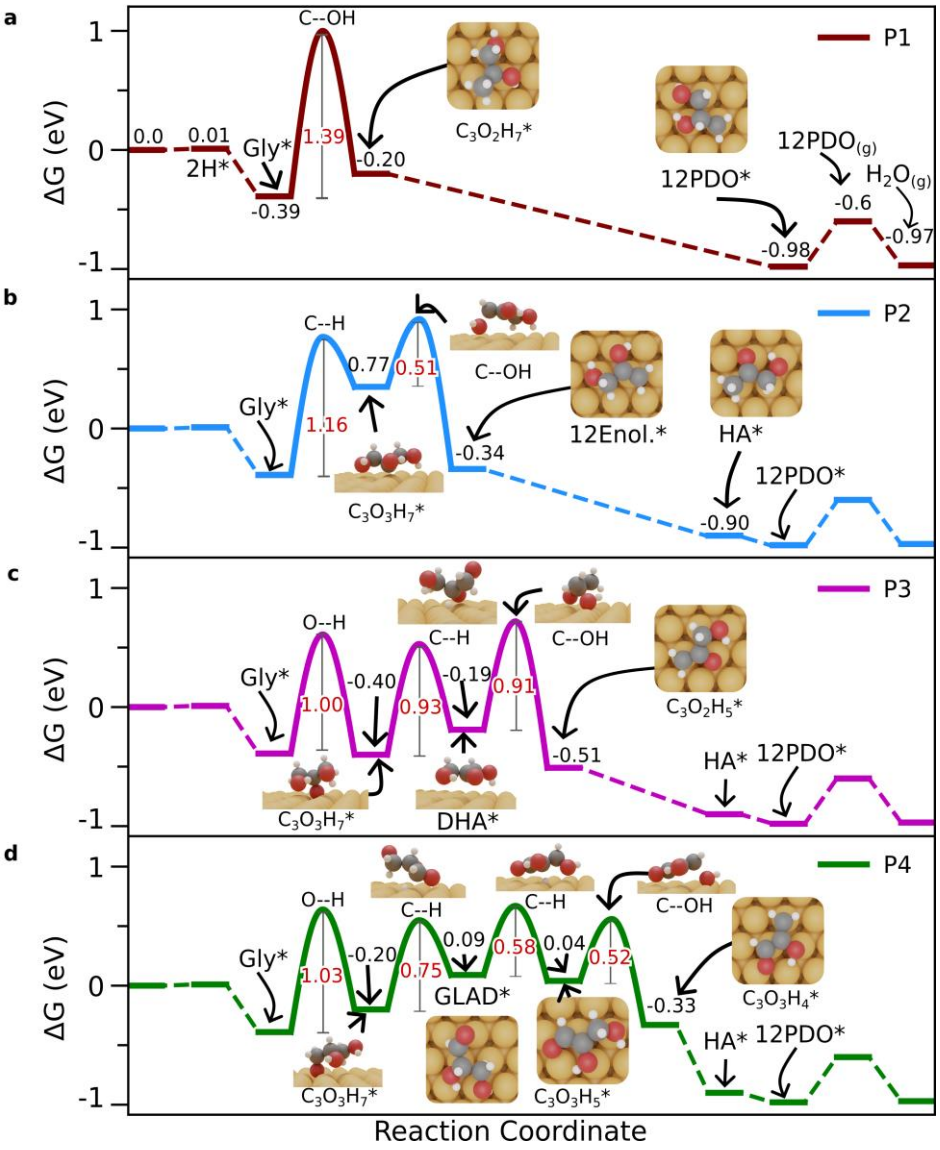

Figure 6: The P1(a), P2(b), P3(c) and P4(d) reaction pathway for glycerol HDO on Cu(111) surface calculated using MACE MLFF. The black text indicates the relative energy of the intermediate with respect to the initial state (glycerol and H_2 in gas phase). The red text indicates the activation free energy barrier for dehydration and dehydrogenation steps. The inserted images represent the minimum free energy intermediate and saddle point configurations.

Investigation into the dehydration of glyceraldehyde for PDO conversion reveals that there are five possible bond scissions for the glyceraldehyde molecule

respectively, relative to the minimum energy configuration of glyceraldehyde on the Cu(111) surface. The activation free energies for the other three bond scissions

are higher, specifically primary O-H, primary C-OH, and secondary O-H being, 0.99, 1.53, and 0.94, respectively (figure 7c). A favorable route to 1,2-PDO via glyceraldehyde is achieved through the low-activation free energy barrier secondary C-H split. The scission of the primary C-OH bond from the product of this elementary step only requires 0.52 eV (Figure 6d).

found that the selectivity of 1,2-PDO or 1,3-PDO from the glyceraldehyde intermediate (in pathway P4 vs P5) is determined by the relative activation barrier differences between the scission of the secondary C-H and secondary C-OH bonds of the glyceraldehyde molecule on Cu(111), which were identified as the most probable elementary steps in section 3.2 and seen in Figure 7.

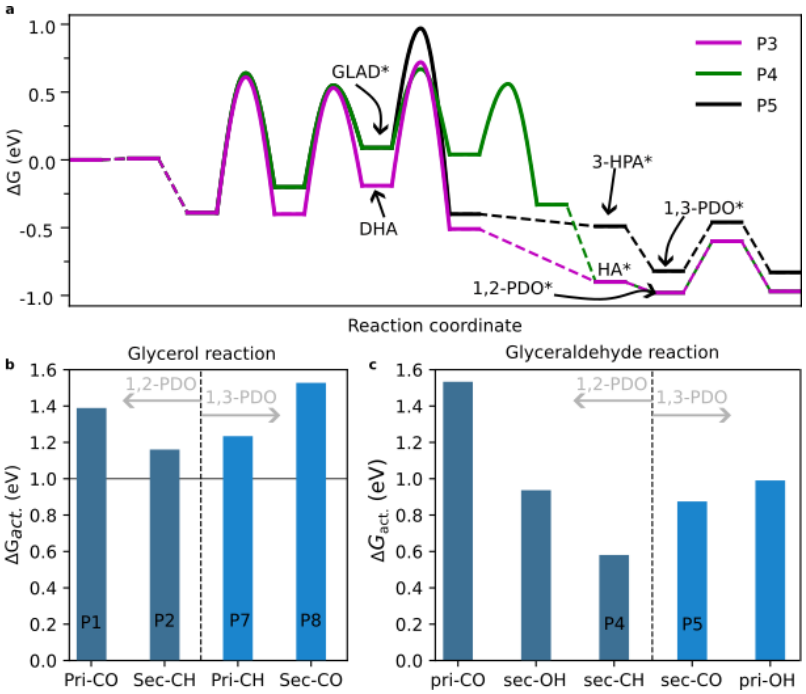

Figure 7: (a)MACE MLFF calculated free energies of reaction pathways from glycerol HDO to 1,2-PDO(P3 and P4) and 1,3-PDO(P5) on Cu(111) surface. (b) The activation free energy barrier of glycerol primary and secondary C-OH and C-H scission affecting the selectivity of HDO towards PDO isomer. (c) The activation free energy barrier of glyceraldehyde bond scissions which dictate the selectivity towards PDO isomers.

The hydrogenation steps following the dehydration and dehydrogenation steps encountered in P1-P4 pathways steps have activation barriers of less than 1.00 eV (figure SI-4), and hence are not the activity or selectivity determining steps.

3.3. Glycerol deoxygenation to 1,3-PDO.

Although the experimental data clearly shows that Cu catalyst is selective towards 1,2-PDO, we also investigated the potential pathways for converting glycerol to 1,3-PDO for the sake of completeness. We

We observed that the scission of the secondary C-OH bond in glyceraldehyde requires 0.3 eV higher activation free energy compared to the scission of the secondary C-H bond, indicating a preference for the formation of 1,2-PDO from glyceraldehyde. Additionally, considering the higher chance of reverse reaction for glyceraldehyde to glycerol intermediate, we see indications that Cu in its reduced form contributes to the observed selectivity of 1,2-PDO formation over 1,3-PDO formation. However, there is experimental evidence suggesting that

other factors, particularly oxidic species present, can further explain the high selectivity of 1,2-PDO on the Cu(111) surface compared to 1,3-PDO through the dihydroxyacetone intermediate.[14]

In the direct PDO formation pathways (P1 vs P7), the selectivity of 1,2-PDO versus 1,3-PDO is dictated by the scission of the primary C-OH and secondary C-OH bonds of glycerol on the Cu(111) surface (Figure 7b). The activation free energy barrier for the scission of the secondary C-OH glycerol bond (1.50 eV) is higher than the activation barrier for the scission of the primary C-OH bond (1.36 eV) of glycerol on Cu(111) surface. Therefore, even through direct HDO, 1,2-PDO is preferred over 1,3-PDO formation on Cu(111). Finally, the activation free energy barrier for primary C-H scission is 1.23 eV, an alternative to secondary C-OH scission for P6 pathway (Figure 7b). This is 0.20 eV higher than primary O-H scission. Therefore, the P5 is the only likely route to form 1,3-PDO from glycerol.

4. Summary and Conclusion:

This study presents a comprehensive analysis of the glycerol to PDO conversion reaction. Despite its significant industrial relevance, literature reviews show a significant lack in a complete overview of mechanistic understanding until now. The study utilizes an active learning based MLFF approach to explore a complex network of 7 competing pathways starting from identifying all 264 possible conformers of glycerol on the Cu(111) surface. Within only 4 iterations, the trained single MLFF model could accurately explore the 26 transition states of the reaction network investigated. The key TS structures and energies were identified and successfully confirmed by DFT calculations (Table SI-1).

The findings reveal important insights into the difference in reactivity trends and bond cleavage preferences between glycerol and glyceraldehyde. The O-H bond cleavage is identified as the initiation step in glycerol, leading to glyceraldehyde or dihydroxyacetone intermediates before further C-O bond cleavage. In contrast, glyceraldehyde shows a higher reactivity towards C-H bond activation followed by adjacent C-OH bond activation.

The study also highlights the dehydrogenation and hydrogenation products of glycerol on the Cu(111) surface, with dihydroxyacetone identified as the most likely dehydrogenation product. The other intermediate, glyceraldehyde has a lower activation barrier for hydrogenation of secondary C as compared to dehydrogenation of secondary C-H and is hence expected to reverse back to glycerol.

While arguably the studied model involving clean Cu(111) surface with gas phase reaction network exploration is not sufficient to explain the very high selectivity of 1,2 PDO on commercial Cu catalysts, we could already qualitatively explain the preference of 1,2 PDO over 1,3 PDO based on the explored pathways. Future studies will build on the present model to include the effect of oxide surface and effect of water solvent with different other Cu facets which understandably are critical for making quantitative claim of selectivity.

In addition, we conducted tests to evaluate the usability of the recently published mace-mp0 universal force field[59] in this specific application. 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. Although the model technically functions and exhibits an overall visual trend that may appear promising, it is evident that it is not suitable for this particular application (Figure SI-2 and Figure SI-3). Consequently, we are left with the option of relying on active learning approaches for now, which have proven to be more effective in meeting our requirements. Our

approach here has been to show how the present technology can find application in industrial research environment, where reaching the final scientific goal without compromising on the reliability of the prediction as quickly as possible are of paramount interest. Hence the objective of keeping a minimal training set was secondary and we allowed for sufficient redundancy in training data generation and opted for frequent explicit validation rather than relying merely on model uncertainty estimates. A retrospective analysis of the training set shows we have at least 20% redundant information from structural point of view (Figure SI-5), elimination of which would likely not affect the model performance. The total number of energy and force calculations calls in this study are of the order of ~421,600 (×500 single point steps). It was only necessary to use ~14,000 for DFT calculations so effectively only ~0.006% of total endeavor besides the significant efficiency gain in terms of time or computational cost: that 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.

5. 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 Ansgar Schaefer, Moritz Otto Haus, Stephan Andreas Schunk, Sebastian Weber, Piyush Ingale from BASF SE, Germany and the, 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.

6. Keywords

Glycerol, machine learn force fields, Hydrodeoxygenation, reaction network, heterogeneous catalyst, Propanediol., Graph neural network, machine learning.

7. References

[1] M. R. Monteiro, C. L. Kugelmeier, R. S. Pinheiro, M. O. Batalha, A. da Silva César, *Renewable and Sustainable Energy Reviews* **2018**, *88*, 109–122.

[2] F. D. Snell, *J Chem Educ* **1942**, *19*, 172.

[3] W. Wan, S. C. Ammal, Z. Lin, K.-E. You, A. Heyden, J. G. Chen, *Nat Commun* **2018**, *9*, 4612.

[4] D. Judas, A. Fradet, E. Marechal, *Journal of Polymer Science: Polymer Chemistry Edition* **1984**, *22*, 3309–3318.

[5] X. G. Li, G. Song, M. R. Huang, *ACS Sustain Chem Eng* **2017**, *5*, 2181–2195.

[6] J. Zhang, *J Appl Polym Sci* **2004**, *91*, 1657–1666.

[7] G. A. Kraus, *Clean (Weinh)* **2008**, *36*, 648–651.

[8] H. Zhao, L. Zheng, X. Li, P. Chen, Z. Hou, *Catal Today* **2020**, *355*, 84–95.

[9] L. Ma, D. He, *Top Catal* **2009**, *52*, 834–844.

[10] Y. Nakagawa, K. Tomishige, *Catal. Sci. Technol.* **2011**, *1*, 179–190.

[11] Z. Luo, Z. Zhu, R. Xiao, D. Chu, *Chem Asian J* **2022**, DOI 10.1002/asia.202201046.

[12] A. Bienholz, H. Hofmann, P. Claus, *Appl Catal A Gen* **2011**, *391*, 153–157.

[13] M. Akiyama, S. Sato, R. Takahashi, K. Inui, M. Yokota, *Appl Catal A Gen* **2009**, *371*, 60–66.

[14] M. Harisekhar, V. Pavan Kumar, S. Shanthi Priya, K. V. R. Chary, *Journal of Chemical Technology & Biotechnology* **2015**, *90*, 1906–1917.

[15] D. Sun, Y. Yamada, S. Sato, W. Ueda, *Appl Catal B* **2016**, *193*, 75–92.

[16] A. Rajan, J. J. Varghese, *J Catal* **2023**, *423*, 94–104.

[17] X. Zhang, G. Cui, H. Feng, L. Chen, H. Wang, B. Wang, X. Zhang, L. Zheng, S. Hong, M. Wei, *Nat Commun* **2019**, *10*, 1–12.

[18] Y. Wang, J. Zhou, X. Guo, *RSC Adv.* **2015**, *5*, 74611–74628.

[19] M. Besson, P. Gallezot, C. Pinel, *Chem Rev* **2014**, *114*, 1827–1870.

[20] J. Shan, H. Liu, K. Lu, S. Zhu, J. Li, J. Wang, W. Fan, *J Catal* **2020**, *383*, 13–23.

[21] A. Rajan, A. P. Pushkar, B. C. Dharmalingam, J. J. Varghese, *iScience* **2023**, *26*, 107029.

[22] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, S. Grimme, *WIREs Computational Molecular Science* **2021**, *11*, e1493.

[23] D. Chen, C. Shang, Z.-P. Liu, *NPJ Comput Mater* **2023**, *9*, 2.

[24] J. T. Margraf, H. Jung, C. Scheurer, K. Reuter, *Nat Catal* **2023**, *6*, 112–121.

[25] A. P. Bento, A. Hersey, E. Félix, G. Landrum, A. Gaulton, F. Atkinson, L. J. Bellis, M. De Veij, A. R. Leach, *J Cheminform* **2020**, *12*, 51.

[26] J. R. Boes, T. Bligaard, in *2018 AIChE Annual Meeting*, **2018**.

[27] W. Xu, E. Diesen, T. He, K. Reuter, J. T. Margraf, **2023**.

[28] V. Kapil, D. P. Kovács, G. Csányi, A. Michaelides, *Faraday Discuss.* **2024**, *249*, 50–68.

[29] T. Zarrouk, R. Ibragimova, A. P. Bartók, M. A. Caro, *arXiv preprint arXiv:2402.03219* **2024**.

[30] L. L. Schaaf, E. Fako, S. De, A. Schäfer, G. Csányi, *NPJ Comput Mater* **2023**, *9*, 180.

[31] 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, *Journal of Physics: Condensed Matter* **2017**, *29*, 465901.

[32] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[33] G. Prandini, A. Marrazzo, I. E. Castelli, N. Mounet, N. Marzari, *NPJ Comput Mater* **2018**, *4*, 72.

[34] 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, *arXiv preprint arXiv:2401.00096* **2023**.

[35] I. Batatia, D. P. Kovacs, G. Simm, C. Ortner, G. Csányi, *Adv Neural Inf Process Syst* **2022**, *35*, 11423–11436.

[36] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L.

Antiga, others, *Adv Neural Inf Process Syst* **2019**, *32*.

[37] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, B. Kozinsky, *Nat Commun* **2022**, *13*, 2453.

[38] K. Schütt, P.-J. Kindermans, H. E. Sauceda Felix, S. Chmiela, A. Tkatchenko, K.-R. Müller, *Adv Neural Inf Process Syst* **2017**, *30*.

[39] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J Chem Phys* **2010**, *132*.

[40] 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, *Nat Commun* **2022**, *13*, 2991.

[41] S. Grimme, S. Ehrlich, L. Goerigk, *J Comput Chem* **2011**, *32*, 1456–1465.

[42] S. Makri, C. Ortner, J. R. Kermode, *J Chem Phys* **2019**, *150*.

[43] P. Lindgren, G. Kastlunger, A. A. Peterson, *J Chem Theory Comput* **2019**, *15*, 5787–5793.

[44] G. Henkelman, H. Jónsson, *J Chem Phys* **2000**, *113*, 9978–9985.

[45] G. and J. H. Henkelman Graeme and Jóhannesson, in *Theoretical Methods in Condensed Phase Chemistry* (Ed.: S.D. Schwartz), Springer Netherlands, Dordrecht, **2002**, pp. 269–302.

[46] H. Jónsson, G. Mills, K. W. Jacobsen, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, World Scientific, **1998**, pp. 385–404.

[47] 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, *Journal of Physics: Condensed Matter* **2017**, *29*, 273002.

[48] P. N. Plessow, *J Chem Theory Comput* **2018**, *14*, 981–990.

[49] S. Maeda, Y. Harabuchi, Y. Ono, T. Taketsugu, K. Morokuma, *Int J Quantum Chem* **2015**, *115*, 258–269.

[50] A. P. Bartók, R. Kondor, G. Csányi, *Phys Rev B* **2013**, *87*, 184115.

[51] 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, *Catal. Sci. Technol.* **2023**, *13*, 2656–2661.

[52] L. Himanen, M. O. J. Jäger, E. V Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke, A. S. Foster, *Comput Phys Commun* **2020**, *247*, 106949.

[53] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, M. Ceriotti, *Sci Adv* **2017**, *3*, e1701816.

[54] S. De, A. P. Bartók, G. Csányi, M. Ceriotti, *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754–13769.

[55] R. K. Cersonsky, S. De, in *Quantum Chemistry in the Age of Machine Learning* (Ed.: P.O. Dral), Elsevier, **2023**, pp. 153–181.

[56] L. McInnes, J. Healy, J. Melville, *arXiv preprint arXiv:1802.03426* **2018**.

[57] F. Drake, P. Jayamaha, L. Wang, **2023**, doi:10.26434/chemrxiv-2023-3g3g81

[58] B. Liu, J. Greeley, *Phys. Chem. Chem. Phys.* **2013**, *15*, 6475–6485.

[59] I. Batatia, P. Benner, et. al., *arXiv* preprint, arXiv:2401.00096, **2024**