Supplementary Information

Artificial Intelligence Pathway Search to Resolve Catalytic Glycerol Hydrogenolysis Selectivity

Pei-Lin Kang[1], Yun-Fei Shi[1], Cheng Shang[1,2], Zhi-Pan Liu[1,2,3]*

[1]Collaborative Innovation Center of Chemistry for Energy Material, Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Key Laboratory of Computational Physical Science, Department of Chemistry, Fudan University, Shanghai 200433, China

[2]Shanghai Qi Zhi Institution, Shanghai 200030, China

[3]Key Laboratory of Synthetic and Self-Assembly Chemistry for Organic Functional Molecules, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China

*email: zpliu@fudan.edu.cn

Table of Contents

1. Summary of experimental data of glycerol hydrogenolysis on Cu-based catalyst

Table S1 Representative experimental data of glycerol hydrogenolysis on Cu-based catalysts

Catalyst	Conversion (%)	1,2-PDO	HA	EG	1,3-PDO
Cu/Al_2O_3[1]	100	96.1	0.8	2.2	-
Raney Cu[2]	100	95.0	0.5	1.9	0
Cu/SiO2[3]	100	98.3	-	-	-
Cu/MMO[4,7]	38.9	90.9	-	~5	-
$Cu\text{-}H_4SiW_{12}O_{40}/SiO_2$ [5]	83.4	22.2	2.9	-	32.1
$Cu\text{-}WO_x\text{-}TiO_2$[6]	12.7	27.6	-	-	32.3

1. Ref. 1: Reaction conditions: Gradient temperatures, 200–130 ℃, 1atm
2. Ref. 2: Reaction conditions: 205 ℃, 14 atm
3. Ref. 3: Reaction conditions: 200 ℃, 50 atm
4. Ref. 4: Reaction conditions: 200 ℃, liquid phase glycerol and 2.0 MPa of H_2 pressure
5. Ref. 5: Reaction conditions: 210 ℃, 5.4 atm
6. Ref. 6: Reaction conditions: 180 ℃, 35 atm, liquid phase glycerol (10 % wt)
7. Ref. 4: MMO: mixed metal oxides (CuMgAl–MMO)

2. SSW-NN method for dataset generation and G-NN potential training

2.1 Architecture of G-NN potential and the double-net architecture

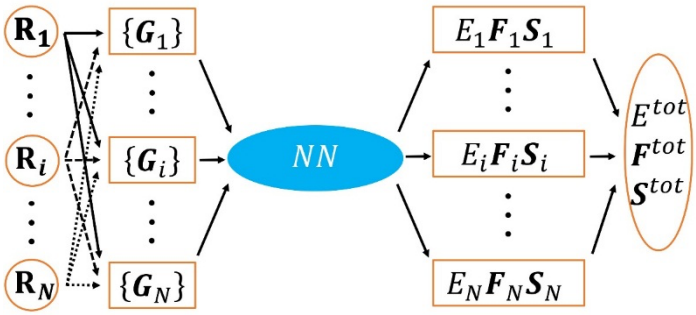


Figure S1. Scheme for the HDNN architecture. The subscripts (1, *i* and *N*) are atom indices, representing the total number of atoms in a structure. The inputs to NN are a set of structural descriptors {**G**} constructed from the Cartesian coordinates {**R**} of the structure, and the outputs are the atomic properties { E_i, $\boldsymbol{F}_i$, $\boldsymbol{S}_i$ }, i.e. energies, forces and stresses. The overall properties E^{tot}, $\boldsymbol{F}^{tot}$, and $\boldsymbol{S}^{tot}$ are calculated from the individual atomic contribution.

Our LASP code (website: www.lasphub.com) utilized the high dimensional neural network (HDNN) scheme to construct the global NN (G-NN) potential[7–10], as shown in Figure S1. The input nodes to NN are a set of power-type structural descriptors $\{\boldsymbol{G}_i\}$ for a structure, as proposed in our previous works[9,10]. The total energy E^{tot} is decomposed as a linear combination of atomic energy E^i from the output of NN

$$E^{tot} = \sum_i E_i \qquad \text{(S1)}$$

The atomic force can be analytically derived according to Eq. S2, where the force component $F_{k,\alpha}$ (α=x, y or z) acting on atom k is the derivative of total energy with respect to the coordinate $R_{k,\alpha}$. In combination with Eq. 1, the force component could be further related to the derivatives of the atomic energy with respect to j^{th} structural descriptors of atom i, $G_{j,i}$

$$F_{k,\alpha} = -\frac{\partial E^{tot}}{\partial R_{k,\alpha}} = -\sum_{i,j} \frac{\partial E_i}{\partial G_{j,i}} \frac{\partial G_{j,i}}{\partial R_{k,\alpha}} \qquad \text{(S2)}$$

Similarly, the element $\sigma_{\alpha\beta}$ of static stress tensor matrix can be analytically derived as

$$\sigma_{\alpha\beta} = -\frac{1}{V}\sum_{i,j,d} \frac{(\boldsymbol{r}_d)_\alpha (\boldsymbol{r}_d)_\beta}{r_d} \frac{\partial E_i}{\partial G_{j,i}} \frac{\partial G_{j,i}}{\partial r_d} \qquad \text{(S3)}$$

where $\boldsymbol{r}_d$ and r_d are the distance vector constituting $G_{j,i}$ and its module, respectively, and V is the volume of the structure.

In this work, we utilized the recently-proposed double-net architecture[10] for constructing the Cu-C-H-O G-NN potential. This approach can inherite the previous trained C-H-O G-NN potential and reuse the related dataset. In the training of Cu-C-H-O G-NN potential, the G-NN potential for C-H-O is incorporated as the auxiliary NN to the four element Cu-C-H-O NN, where the total energy is now expanded as the sum of the atomic energy from both NN. The main advantage of the double-net approach is the reusage of the previously obtained G-NN potentials and thus the reduction in the training time of the new G-NN with more elements. By this approach, the double-network potential is composed of two set of NNs for C, H and O elements and only one NN for Cu element.

2.2 Dataset generation and self-learning of G-NN potential

Undoubtedly, the training dataset largely determines the quality of the potential energy surface (PES) of G-NN. Our previous works have shown that, the stochastic surface walking (SSW) global optimzation[11,12] can fast

generate a global PES dataset, which incorporates different structural patterns on the global PES. Because the vast reaction space of heterogenous reactions, we design an automated procedure for self-learning G-NN potential, not only for relatively simple small molecules (with C, H, O elements) but also for glycerol hydrogenolysis reactions. The procedure is elaborated in the following.

Step 1: Building the initial NN potential The first dataset for Cu-C-H-O is inherited from our previous SW global sampling data for simple reactions on Cu surfaces (~10000 structures) and organic reactions (~12000 structures, on which the C-H-O NN was previously trained)[10]. By carrying out DFT calculations on these data, we trained the first double-net NN potential for Cu-C-H-O system.

Step 2: Self-learning of G-NN potential Using the NN potential generated from the first step, we can accelerate greatly the PES sampling of organic molecules, which provide the chance to sample the reaction space of small molecules and glycerol. In the iterative self-learning procedure, the SSW global optimization using G-NN potential is utilized to expand the PES dataset and the G-NN potential is then gradually improved by learning the new dataset from SSW global optimization. For each cycle, a small additional dataset, less than 500 structures, is obtained from the SSW sampling trajectories, containing the structures on PES either randomly selected or exhibiting unknown atomic environment that are labelled as exotic structures. More than 100 cycles are performed to finally generate the Cu-C-H-O quaternary G-NN potential.

The exotic structures produced from SSW trajectory are often poorly predicted ones from the NN potentials. They need to be added to the training dataset to improve NN predictability. These structures are found to have either one or all of the following features: (i) the value of the structural descriptor for these structures is out of the boundary defined by existing structures in the train dataset, (ii) the second derivative (frequency) of the structure is either too high or too low, and (iii) the energy (per atom) of the structure is far higher or lower than the structures in the dataset. For example, these structures often have highly unreasonable geometry, such as very high coordination, very short bond distance or a ring structure with too high tension. While these structures are generally not low energy minima, they define the high energy boundary of the PES and thus are essential in practical simulations to be avoided.

3. Methodology for reaction rule extraction

The reaction rules are automatically extracted from the reaction data collected from SSW-NN simulation. We followed the shell-based scheme to extract reaction rules as described by Christ al.[13], and Saller and coworkers[14], and an example is illustrated in Figure S2. A reaction rule contains (i) the reaction centers (O1 and H2), which include the atoms and bonds changed in the reaction; (ii) the atoms that are the first neighbors to atoms in the reaction center (the C atom attached to O1). It should be noted that, for the same reasons in generating s-ECFP, all H atoms are explicitly represented and treated as the same as heavy atoms. In this work, the same reaction rule that occurs at least two times in rule dataset will be finally recorded in the reaction database, leading to 5575 rules in total for molecular reaction on Cu surfaces.

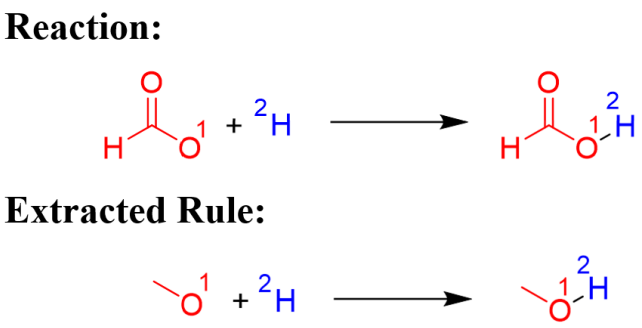


Figure S2. The algorithmic rule extraction scheme as illustrated by the example of the HCOO+H→ HCOOH

4. Training of NN models for R-Pat and K-info units

In training NN models for R-Pat and K-info units, we utilize the keras API as implemented in Tensorflow[15]. The hyperparameters in training R-Pat NN are as follows, number of hidden layers: 2 (512 nodes per layer); Activation function: elu; Dropout fraction 0.2; Loss function type: categorical cross entropy; Optimizer: Adam optimizer (step size 0.001); Batch size: 128. The hyperparameters in training K-Info NN are as follows, Number of hidden layers: 2 (256 nodes per layer); Activation function: sigmoid; Dropout fraction 0.2; Loss function type: Mean squared error; Optimizer: Adam optimizer (step size 0.001); Batch size: 128. The learning curve is plotted in **Figure S3.**

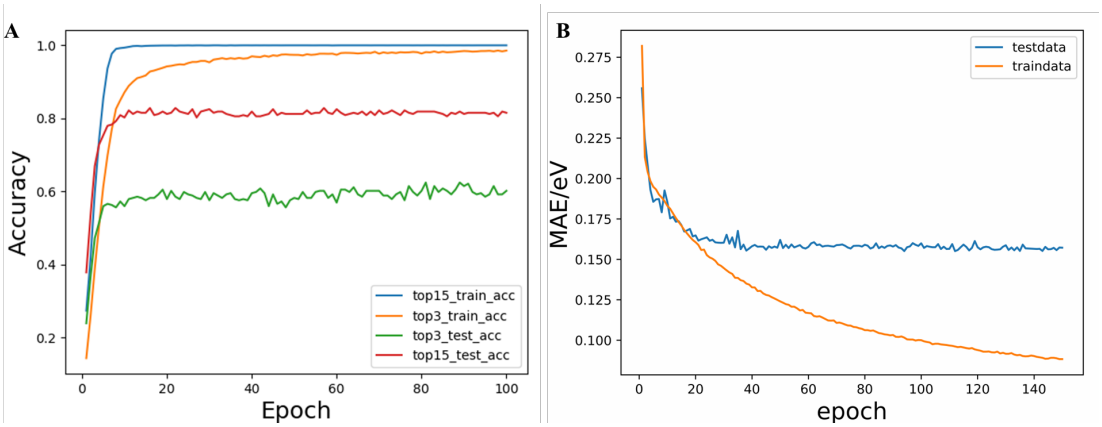


Figure S3. The learning curve of two NN models. A) R-Pat NN B) K-Info NN.

5. Reaction database analysis for catalytic reactions on Cu surfaces

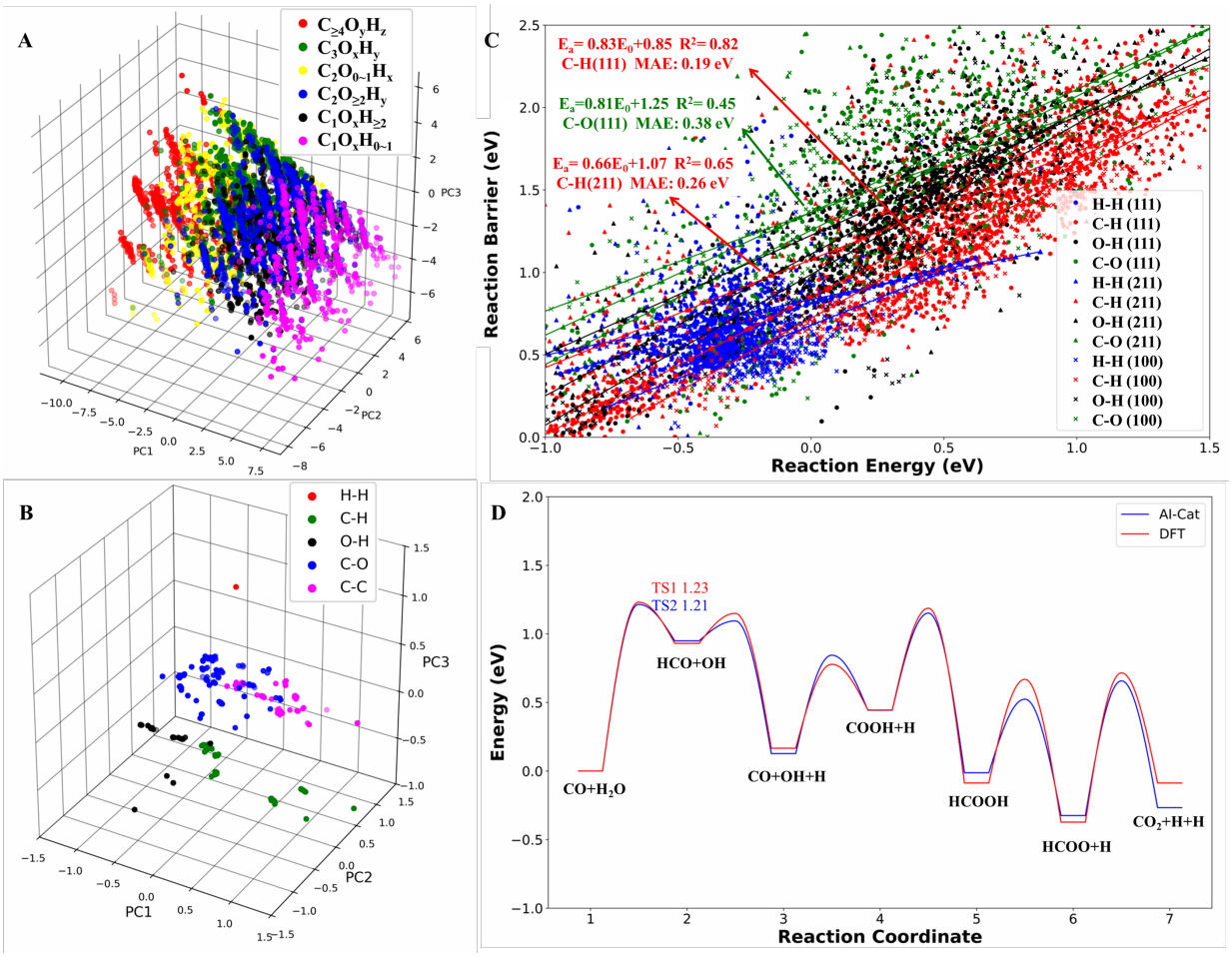


Figure S4 Reaction database for catalytic reactions on Cu surfaces **(A)** Principal components analysis (PCA) for reactants of reaction database; **(B)** PCA for reaction patterns of reaction database; **(C)** the BEP relations between the reaction barrier and the reaction energy for twelve different reactions on Cu surfaces using 6225 simple dissociation reactions in the reaction database; **(D)** the AI-Cat predicted profile for the lowest energy pathway of $CO+H_2O$ reaction on Cu(111) as compared with that from DFT.

To give an overview of the reaction data, we applied the principal components analysis (PCA) on the dataset, which examines the distinction between data. The PCA results projected onto the leading principal components (PC1 to PC3) are shown in Figure S4a and b. Figure S4a shows the results for 7776 reactants, where the input vectors of R-Pat unit are utilized as the basis in PCA. It is found that similar data points tend to cluster with each other, demonstrating our descriptors can distinguish different reactants. PC1 reflects majorly the difference in carbon atom numbers of reactants according to the chemical formula of the largest fragment in reactants, as labelled by different color points. The points at the right-bottom corner correspond solely to the reactants with CO molecule (C_1). In addition, PC2 distinguishes the local chemical environment of these molecules, where the type of coadsorbate and the surface coverage make the difference. Similarly, Figure S4b illustrates the PCA results for 6225 simple dissociation reactions on Cu surfaces by using the s-ECFP of reaction center atoms as the basis in PCA. These dissociation reactions involve only one chemical bond breaking at the TS and thus contain only two reaction center atoms. The data points in different colors in Figure S4b correspond to different chemical bond in dissociation. Not surprisingly, similar reactions tend to cluster, suggesting s-ECFP4 can distinguish different reaction center. For the H-H bond breaking, all reaction patterns are the same and thus only yields one point (red), while the C-O and C-C

bond breaking reactions are more diverse due to more complex chemical environments.

We would also like to emphasize that the BEP relation commonly used for activity estimation can indeed be observed from our dataset, as shown in Figure S4c for the reaction barrier vs. reaction energy for 6225 simple dissociation reactions. However, the linearity in correlation is rather poor, even when the correlation is applied to the same type of bond dissociation reaction, reflecting the reaction complexity in large reaction database (e.g. multiple pathways and the variation of surface sites). This is evidenced by low correlation coefficient R^2 of the fitted lines in the Figure (all fitted functions are listed in Table S2). The mean absolute error (MAE) in the predicted barrier is 0.15, 0.20, 0.19 and 0.38 eV for H-H, C-H, O-H and C-O dissociation reaction on Cu(111) surface, respectively. The C-H dissociation reactions on (111) surface have the best linearity in all fitted functions with R^2 being 0.82. The largest error occurs at the C-O bond dissociation reaction, which is reasonable considering that the chemical environment can be vastly different for different C-O bonds, e.g C-O bond breaking in carboxyl group being very different from that in CO molecule. As a result, the MAE and R^2 of C-O dissociation reaction on (111) surface are 0.38 eV and 0.45 respectively, which are the worst of all fitted functions. For the stepped (211) surface, the BEP linearity is similarly poor, in which the MAE is 0.19, 0.26, 0.26 and 0.37 eV for H-H, C-H, O-H and C-O dissociation reactions, respectively, with the R2 typically below 0.7. In short, although the general presence of BEP relation, the accuracy of MAE up to 0.4 eV and the large R2 deviation suggest a poor reliability in predicting the catalytic activity with BEP correlation even if the reaction energy is available. In contrast, for these simple 6225 dissociation reactions in Figure S4c, the MAE of K-Info model for reaction barrier and reaction energy is 0.055 eV and 0.049 eV respectively, which is improved markedly over the BEP linear relationship.

Table S2 Linear fitted BEP relation for simple dissociation reactions on Cu surfaces in the reaction database. $\mathbf{N_{react}}$ is the number of the same type of reactions that can be utilized for linear fitting.

Reaction Type	Surface	Fitted Linear Func.	MAE/ eV	R^2	N_{react}
H-H dissociation	（111）	Y= 0.449X+0.808	0.151	0.180	573
C-H dissociation	（111）	Y= 0.835X+0.851	0.197	0.820	1110
O-H dissociation	（111）	Y= 0.859X+1.111	0.194	0.697	814
C-O dissociation	（111）	Y= 0.813X+1.254	0.378	0.454	342
H-H dissociation	（100）	Y= 0.646X+0.678	0.154	0.179	298
C-H dissociation	（100）	Y= 0.893X+0.721	0.178	0.804	391
O-H dissociation	（100）	Y= 0.913X+0.987	0.209	0.749	359
C-O dissociation	（100）	Y=0.598X+1.023	0.342	0.428	244
H-H dissociation	（211）	Y= 0.337X+0.835	0.188	0.089	583
C-H dissociation	（211）	Y=0.655X+1.077	0.261	0.647	613
O-H dissociation	（211）	Y=0.722X+1.224	0.264	0.674	473
C-O dissociation	（211）	Y=0.745X+1.364	0.369	0.499	224
others	-	-	-	-	201
Total	-	Y=0.773X+1.023	0.28	0.64	6225

6. Validation of AI-Cat predicted reactions on Cu surfaces

In our AI-Cat model, the R-Pat unit learns 7776 intermediates and 5136 reaction patterns associated with them, 90% of them being set as the training set and the left as the testing set. The probability to occur a reaction pattern is determined by the weighted rate constant in Eq. 1, which is used to supervise the training. After training the model, the top 3 accuracy for predicting most probable reaction pattern is perfect (99%) for the training set and 57% for the testing set; and the top 15 accuracy for the testing set reaches to 79%.

The K-info unit learns the reaction dataset of 30508 reactions on three different Cu surfaces. By splitting the dataset into 90% training set and 10% testing set, the NN finally achieves the accuracy of prediction with MAE of 0.061 and 0.054 eV for the reaction barrier and reaction energy on the training dataset, and MAE of 0.157 and 0.149 eV for the barrier and reaction energy on the testing dataset, respectively. For the simple 6225 dissociation reactions as mentioned above in Figure S4c, the MAE for reaction barrier and reaction energy is 0.055 eV and 0.049 eV respectively, which is improved markedly over the BEP linear relationship.

To better understand the accuracy of AI-Cat model for catalytic reactions with consecutive elementary steps, we have tested 21 different compositions by mixing 7 common small molecules, including H_2, CO, CO_2, H_2O, CH_3OH, CH_4 and CH_2=CH_2 and AI-Cat have yielded correct prediction for the known reactions (see Table S3 for all the data in details). To give an example, we present the AI-Cat predicted results for water gas shift reaction on Cu(111) and benchmarked them against DFT results, as shown in Figure S4d. This reaction has been studied by DFT-based SSW reaction sampling previously[16] and thus is a good benchmarking example. The lowest energy pathway is known to follow the mechanism of $CO+H_2O \rightarrow HCO+OH \rightarrow CO+OH+H \rightarrow COOH+H \rightarrow HCOOH \rightarrow HCOO+H \rightarrow CO_2+H+H$. The major competing pathway is the formation of formate (HCOO) via the hydrogenation of CO_2 product that can alternatively be produced by direct COOH dehydrogenation[17]. Therefore, the key challenge in reaction prediction is thus to predict correctly the small barrier difference between COOH dehydrogenation and COOH hydrogenation, which differs by only 0.15 eV in DFT. Obviously, our AI-Cat model performs well in distinguish the two pathways with the predicted barrier difference being 0.08 eV and thus yields the same mechanism as DFT results. By comparing the energetics in the whole pathway, we found that the MAE of predicted barrier and reaction energy is 0.087 eV and 0.074 eV with respect to the DFT counterparts, whilst the barrier of the rate-determining step, the water dissociation step, from the AI-Cat prediction (1.21 eV) is almost identical to the DFT value (1.23 eV). The good accuracy of AI-Cat prediction also suggests the high accuracy of G-NN PES in reproducing the DFT PES (also see Figure S5 for comparison between G-NN and DFT).

It might be mentioned that the explicit kinetics data from K-Info prediction greatly improves the convergence of MC tree search in finding the correct pathway. We have compared two search modes with or without the kinetics data from K-info unit. With K-info prediction, only 97 nodes are generated in tree search until no more low energy pathway (overall barrier within 0.3 eV above the lowest energy pathway) is identified. In contrast, in the absence of the K-info unit AI-Cat has to enumerate all the likely pathway that gives a few tens thousands of nodes until no more pathways can be found. The searching speed thus differs by more than two orders of magnitude.

Table S4 lists the feasible catalytic reactions predicted by our AI-Cat model with the predicted overall barrier and the reaction site for the rate determining step. The feasible reaction is considered as an exothermal reaction (without zero-point-energy correction) at 0 K with the overall barrier lower than 1.2 eV. Interestingly, these results do reflect the characteristics of Cu catalysts commonly known in heterogeneous catalysis, for example, the low/no activity towards methane conversion; the low catalytic ability towards growing higher hydrocarbons; and the high tendency to retain C-O bond of molecule in catalysis. It is worth mentioning that because our AI-Cat model learns all reaction data on three different Cu surfaces, the rate-determining step predicted generally prefers the open (100) or the stepped (211) surfaces, suggesting the higher activity of these surfaces compared to the close-packed Cu (111) surface. Below we briefly discuss these reactions in the context of the related experimental catalytic conditions in

literatures.

$CO+2H_2 \rightarrow CH_3OH$: The syngas conversion on Cu catalyst is widely utilized in industry, where $Cu/ZnO/Al_2O_3$ catalysts operated under 200-300 ℃ can efficiently covert CO and H_2 to methanol[18]. From our AI-Cat, the rate-determining step occurs at the $CHO+H \rightarrow CH_2O$ with a 0.80 eV overall barrier on Cu (100) surface.

$CH_2=CH_2+H_2 \rightarrow CH_3CH_3$: Ethene hydrogenolysis is known to occur on reduced Cu surface under 150-250 ℃ and 1 atm[19]. From our AI-Cat, the rate-determining step occurs at the first H addition reaction, $CH_2=CH_2+H \rightarrow CH_2CH_3$, with a 0.92 eV overall barrier on Cu(211) surface.

$CO+H_2O \rightarrow CO_2+H_2$: $Cu/ZnO/Al_2O_3$ is widely used as low-temperature water-gas-shift reaction catalyst in industry for producing clean H_2 under 190-230 ℃[20]. From our AI-Cat, the rate-determining step is $COOH+H \rightarrow CO_2+H_2$, with a 0.97 eV barrier on (100) surface. Recent research also shows that Cu(100) surface is the most active facet in low-temperature WGSR[21].

$CO+CH_3OH \rightarrow HC(=O)OCH_3$: Carbonylation of methanol with CO to generate methyl formate (MF), is reported to occur on Cu nanocluster catalyst with 100% MF selectivity under 100-160 ℃, 0.3–3.0 MPa CO[22]. From our AI-Cat, the rate-determining step is synergistic methanol dissociation reaction, $CO+CH_3OH \rightarrow CH_3O+CHO$, with a 1.03 eV barrier on Cu(100) surface.

$CH_2=CH_2+H_2O \rightarrow CH_2CH_3OH$: The hydration of ethylene on metal surface was not investigated for economic reasons. Nevertheless, ethylene and ethanol are common competitive products in the electroreduction of CO_2 on Cu, which indicates the possibility of mutual conversion on Cu surfaces. From our AI-Cat, the rate-determining step is $CH_2CH_3+OH \rightarrow CH_3CH_2OH$, with a 1.06 eV barrier on Cu(211) surface.

Table S3 All the reactions (21 in total) tested on Cu surfaces predicted by AI-Cat, starting from the combination of reactants CO, H_2, CH_3OH, CH_4, CO_2, C_2H_4, and $CH_2=CH_2$. The likely reactions are highlighted, where the overall reaction energy ΔE is exothermic or less than 0.1 eV endothermic. The overall barrier E_a predicted by AI-Cat, the rate-determining step and its reaction site (surface) of the likely reactions are also listed.

No.	Input Name	Predicted Product	E_a/eV	ΔE/eV	Surface	Rate-deter. step
1	$CO+H_2$	CH_3OH	0.80	-0.98	(100)	$CHO+H \rightarrow CH_2O$
2	$CH_2=CH_2+H_2$	CH_3CH_3	0.92	-1.41	(211)	$CH_2=CH_2+H \rightarrow CH_2CH_3$
3	$CO+H_2O$	CO_2+H_2	0.97	-0.43	(100)	$COOH+H \rightarrow CO_2+H_2$
4	$CO+CH_3OH$	$HC(=O)OCH_3$	1.03	-0.35	(100)	$CO+CH_3OH \rightarrow CH_3O+CHO$
5	$CH_2=CH_2+H_2O$	CH_2CH_3OH	1.06	-0.46	(211)	$CH_2CH_3+OH \rightarrow CH_3CH_2OH$
6	CO_2+CH_3OH	$H_3COC(=O)OH$	1.02	-0.09	(100)	$CH_3O+CO_2 \rightarrow H_3COC(=O)O$
7	CO_2+H_2	HCOOH	0.82	0.10	(211)	$CO_2+H_2 \rightarrow COOH+H$
8	CH_3OH+H_2	-	-	-	-	-
9	$CO+CO_2$	-	-	-	-	-
10	$CH_2=CH_2+CO_2$	-	-	-	-	-
11	$CH_2=CH_2+CO$	-	-	-	-	-

12	CH_2=CH_2+CH_3OH	CO+CH_3CH_3	1.76	-1.85	(111)	H_2C=O→HC=O+H
13	H_2O+H_2	-	-	-	-	-
14	H_2O+CO_2	H_2CO_3	0.98	-0.17	(211)	C(=O)(OH)O+H →H_2CO_3
15	CH_3OH+H2O	-	-	-	-	-
16	CH_4+ H_2O	-	-	-	-	-
17	CO_2+CH_4	-	-	-	-	-
18	CH_4+ H_2	-	-	-	-	-
19	CH_4+ CO	-	-	-	-	-
20	CH_4+ CH_3OH	-	-	-	-	-
21	CH_4+ CH_2=CH_2	-	-	-	-	-

Table S4 Feasible reactions on Cu surfaces as predicted by AI-Cat, starting from the combination of reactants CO, H_2, CH_3OH, CH_4, CO_2, C_2H_4 (CH_2=CH_2). The feasible reaction is considered to be those being exothermal (without ZPE) at 0 K with the overall barrier lower than 1.2 eV. The overall reaction barrier E_a together with the rate-determining step and its reaction site (surface) are listed. The references of the related experiments that support AI-Cat prediction are also provided.

Reaction	E_a/eV	Rate-deter. step	Surface	Expt.*
CO+2H_2→CH_3OH	0.80	CHO+H→CH_2O	(100)	1
CH_2=CH_2+H_2→CH_3CH_3	0.92	CH_2=CH_2+H→CH_2CH_3	(211)	2
CO+H_2O→CO_2+H_2	0.97	COOH+H→CO_2+H_2	(100)	3
CO+CH_3OH→HC(=O)OCH_3	1.03	CO+CH_3OH →CH_3O+CHO	(100)	4
CH_2=CH_2+H_2O→CH_2CH_3OH	1.06	CH_2CH_3+OH→CH_3CH_2OH	(211)	-

*Reaction conditions known in experiment:

1. Ref. 18: Cu/ZnO/Al_2O_3 200-300 ℃ and 50-100 atm;
2. Ref. 19: Cu 150-250 ℃ 1 atm;
3. Ref. 20: Cu/ZnO/Al_2O_3 150-300 ℃;
4. Ref. 22: Cu nanocluster 100-160 ℃, 0.3–3.0 MPa CO

7. Benchmark of G-NN PES Accuracy

Our AI-Cat model is trained based on the dataset collected from G-NN potential energy surface (PES). This introduces the first source of error, i.e. the error between G-NN PES and DFT PES. To illustrate this in catalytic reaction prediction, we plot the energy profiles of lowest energy pathway for $CO+H_2O$ reaction on Cu(111) between G-NN and DFT results in Figure **S5**. By comparing the energetics in the whole pathway, we found that the MAE of NN PES barrier and reaction energy is 0.07 eV and 0.09 eV with respect to the DFT counterparts, which is quite typical for G-NN accuracy. The accuracy of G-NN PES is generally good enough to ensure the AI-Cat prediction on energetics.

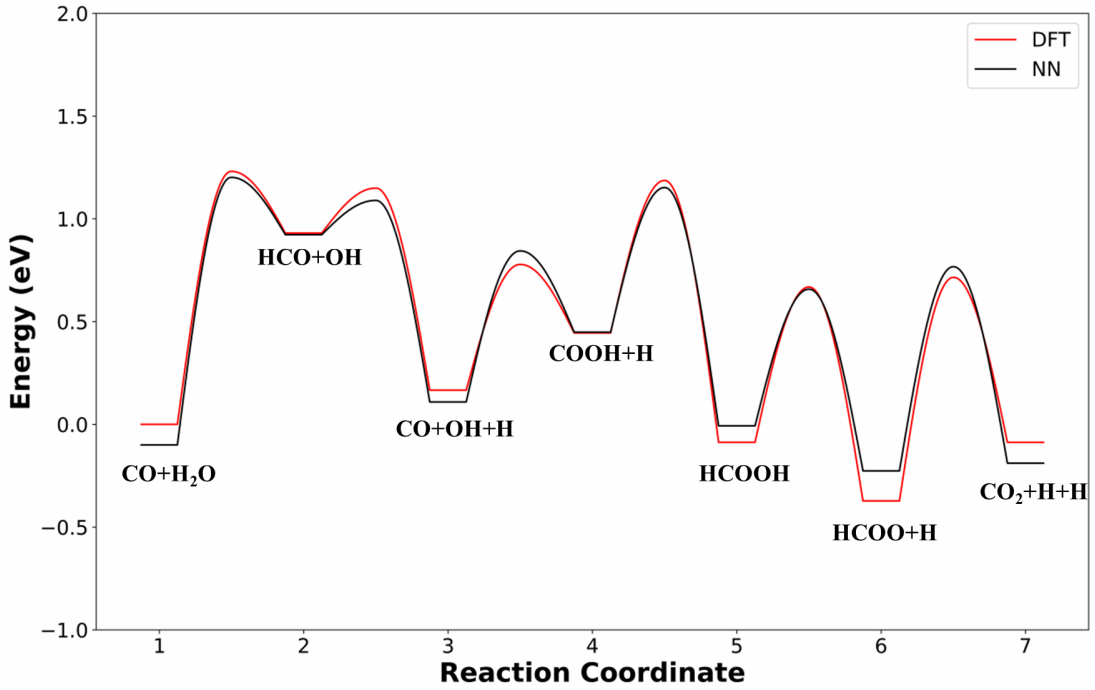


Figure S5 Energy profile of lowest energy pathway for $CO+H_2O$ reaction on Cu(111) from G-NN and DFT calculations.

8. All pathway data on glycerol hydrogenolysis from AI-Cat and from DFT

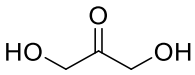
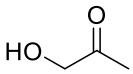
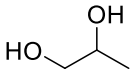
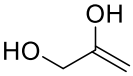
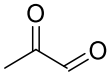
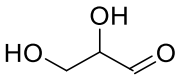
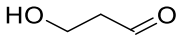
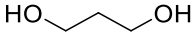
Table S5 Low energy pathway products starting from glycerol and H_2 on Cu(111) predicted from AI-Cat with energetics from AI-Cat and from DFT for comparison. These products belong to exothermal products (after ZPE correction) in energy (0 K) with the overall barrier lower than 1.6 eV. The listed data include the ZPE-corrected overall barrier E_a (AI) (E_a from AI-Cat plus approximated ZPE, i.e. 0.15 eV for X-H bond, X=C, O), the ZPE-corrected overall barrier (E_a) computed from DFT (E_a^1), that with further van der Waals correction (E_a^2) and also with the free energy correction (G_a) that include the thermal correction ΔU(T), the pressure term ΔPV and the entropy term TΔS (see the equations below). The same overall E_a (G_a) of products indicates they share the same rate-determining step.

$E_a = E_{TS} - E_{IS}$

$E_a^1 = E_a + \Delta ZPE$

$E_a^2 = E_{a1} + \Delta E_{vdw}$

$G_a = E_{a2} + \Delta U(T) + \Delta PV - T\Delta S$

No	name	Molecule structure	E_a(AI) /eV	E_{a1} / eV	E_{a2} /eV	G_a /eV
1	dihydroxyacetone		1.10	1.03	0.98	1.06
2	acetol		1.36	1.42	1.32	1.25
3	1,2-PDO		1.36	1.42	1.32	1.25
4	prop-2-ene-1,2-diol		1.36	1.42	1.32	1.25
5	2-oxopropanal		1.36	1.42	1.32	1.25
6	glyceraldehyde		1.45	1.13	1.09	1.28
7	3-hydroxypropanal		1.53	1.53	1.28	1.44
8	1,3-PDO		1.53	1.53	1.28	1.44

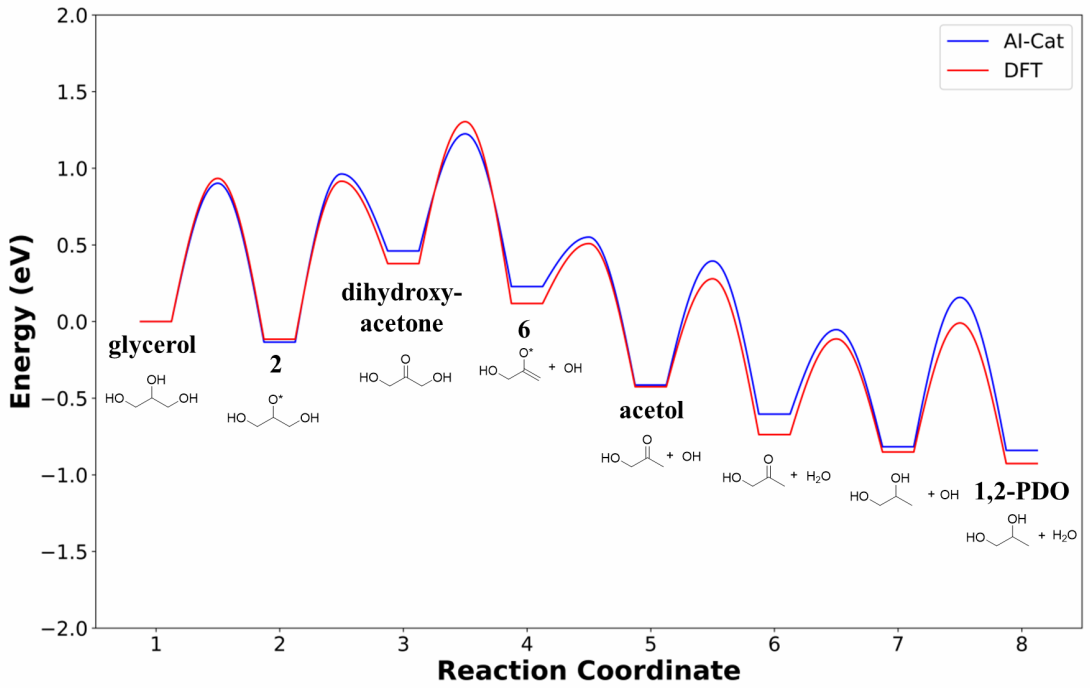


Figure S6 Energy profile of lowest energy pathway from glycerol to 1,2-PDO on Cu(111) from AI-Cat and DFT energetics (with ZPE correction). The MAE of AI-Cat barrier and reaction energy are 0.09 eV and 0.07 eV with respect to the DFT counterparts.

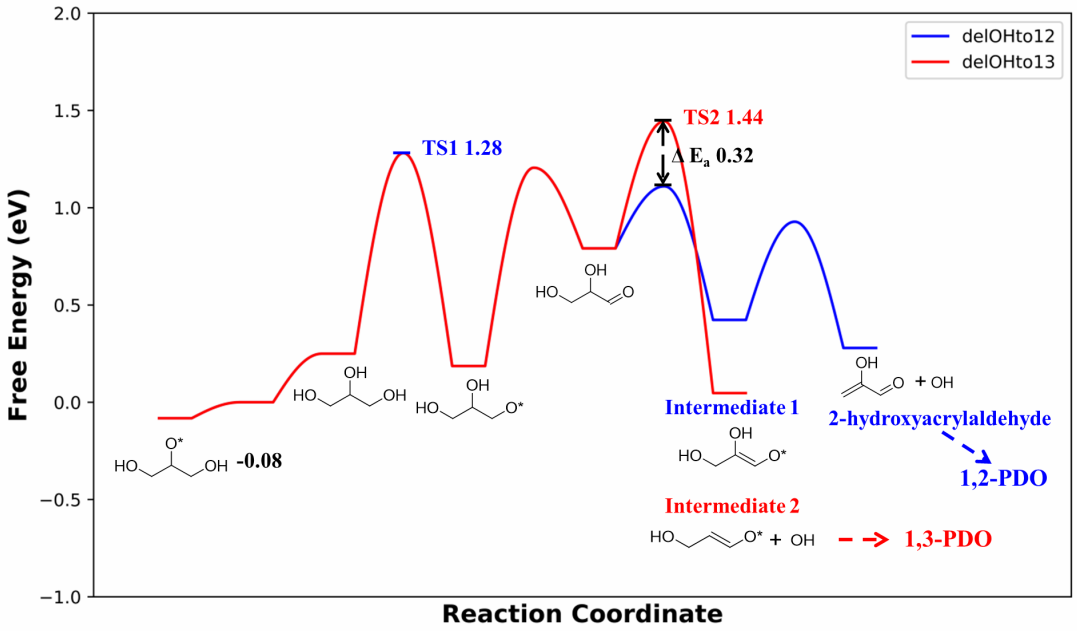


Figure S7 DFT-based Free energy profiles for glyceraldehyde to break its C-O bond at the secondary C atom and the primary C atom that leads to 1,3-PDO and 1,2-PDO, respectively. The energy zero refers to glycerol in the gas phase under 473 K and 0.0065 atm.

The retro-Michael addition pathway bifurcating from glyceraldehyde competes with the dehydrogenation pathway to 1,3-PDO. As shown in Figure **S7**, once the glyceraldehyde is generated, there is a two-step retro-Michael addition dehydration pattern to generate 2-hydroxyacrylaldehyde (blue lines) with a barrier that is 0.32 eV lower than the C-O bond break on secondary C atom (red lines). As we mentioned in section 3.3, for each C-O breaking, there are a series of dehydrogenation/hydrogenation fast equilibrium. Obviously, the 2-hydroxyacrylaldehyde will finally transfer to 1,2-PDO by hydrogenation. As a result, even if glyceraldehyde forms, the presence of a low-barrier dehydration channel via the retro-Michael addition reaction will still favor the selectivity to 1,2-PDO.

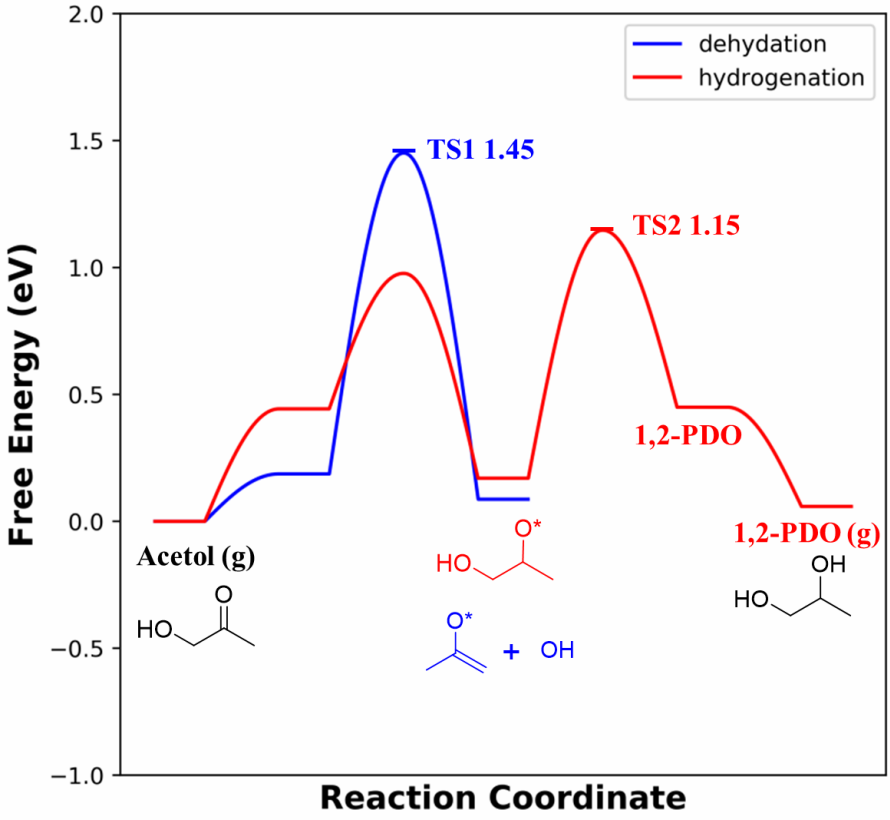


Figure S8 DFT-based Free energy profile for acetol hydrogenation and further dehydration. The energy zero refers to acetol in the gas phase under 473 K and 0.0065 atm.

9. Kinetics data used in microkinetics simulation

Table S6 Free energy barrier and reaction rate constant of elementary reactions at 473 K and a total pressure of 1 atm. (p_{H_2}=1 bar; $P_{glycerol}$= P_{acetol}= $P_{1,2\text{-}PDO}$= $P_{1,3\text{-}PDO}$= $P_{3\text{-}hydroxypropanal}$= P_{water} =0.0065 bar)

No	name	$G_{a,+}$	$G_{a,-}$	k_+	k_-
0	glycerol* ->glycerol+*	0	0.249	9.86E+12	2.19E+10
1	acetol* ->acetol+*	0	0.408	9.86E+12	4.43E+08
2	3-hydroxypropanal* ->3-hydroxypropanal +*	0	0.512	9.86E+12	3.45E+07
3	H_2O* ->H_2O+*	0	0.549	9.86E+12	1.39E+07
4	OH*+H* ->H_2O*+*	0.69	0.871	4.38E+05	5.17E+03
5	H_2+*+* ->H*+H*	0.751	0.831	9.81E+04	1.38E+04
6	glycerol*+* ->2*+H*	0.809	1.14	2.37E+04	7.03E+00
7	2*+* ->8*+H*	0.847	0.414	9.31E+03	3.82E+08
8	8*+* ->6*+OH*	0.813	1.238	2.14E+04	6.35E-01
9	6*+H* ->acetol*+*	0.426	0.864	2.85E+08	6.14E+03
10	acetol*+H* ->11*+*	0.363	0.768	1.34E+09	6.47E+04
11	11*+H* ->1,2-PDO+*+*	0.938	1.088	9.99E+02	2.52E+01
12	glycerol*+* ->3*+H*	1.031	1.095	1.02E+02	2.12E+01
13	3*+* ->9*+H*	1.019	0.415	1.37E+02	3.73E+08
14	9*+* ->10*+OH*	0.653	1.397	1.09E+06	1.28E-02
15	10*+H*->3-hydroxypropanal *+*	0.217	0.69	4.80E+10	4.38E+05
16	3-hydroxypropanal*+H* ->12*+*	0.424	0.826	2.99E+08	1.56E+04
17	12*+H* ->1,3-PDO+*+*	1.082	1.432	2.92E+01	5.44E-03
18	glycerol*+H* ->7*+*	0.837	0.006	1.19E+04	8.51E+12
19	7*+* ->13*+H2O*	0.684	1.581	5.08E+05	1.41E-04
20	13*+H* ->1,2-PDO +*	0.425	1.766	2.92E+08	1.50E-06
21	glycerol*+H* ->1*+*	0.781	0.044	4.70E+04	3.35E+12
22	1*+* ->14*+H2O*	0.719	1.204	2.15E+05	1.46E+00
23	14*+H* ->1,3-PDO +*	0.032	1.559	4.50E+12	2.41E-04

*The structure number corresponds to Figure 5

Reference in SI:

1 S. Sato, M. Akiyama, K. Inui and M. Yokota, *Chem. Lett.*, 2009, **38**, 560–561.

2 S. K. Tanielyan, N. Marin, G. Alvez, R. Bhagat, B. Miryala, R. L. Augustine and S. R. Schmidt, *Org. Process Res. Dev.*, 2014, **18**, 1419–1426.

3 E. S. Vasiliadou, T. M. Eggenhuisen, P. Munnik, P. E. de Jongh, K. P. de Jong and A. A. Lemonidou, *Applied Catalysis B: Environmental*, 2014, **145**, 108–119.

4 X. Zhang, G. Cui, H. Feng, L. Chen, H. Wang, B. Wang, X. Zhang, L. Zheng, S. Hong and M. Wei, *Nat Commun*, 2019, **10**, 5812.

5 L. Huang, Y. Zhu, H. Zheng, G. Ding and Y. Li, *Catal Lett*, 2009, **131**, 312–320.

6 D. Li, Z. Zhou, J. Qin, Y. Li, Z. Liu and W. Wu, *ChemistrySelect*, 2018, **3**, 2479–2486.

7 J. Behler and M. Parrinello, *Phys. Rev. Lett.*, 2007, **98**, 146401.

8 S.-D. Huang, C. Shang, X.-J. Zhang and Z.-P. Liu, *Chem. Sci.*, 2017, **8**, 6327–6337.

9 S.-D. Huang, C. Shang, P.-L. Kang and Z.-P. Liu, *Chem. Sci.*, 2018, **9**, 8644–8655.

10 P. Kang, C. Shang and Z. Liu, *Chinese Journal of Chemical Physics*, 2021, **34**, 583–590.

11 C. Shang and Z.-P. Liu, *J. Chem. Theory Comput.*, 2013, **9**, 1838–1845.

12 C. Shang, X.-J. Zhang and Z.-P. Liu, *Phys. Chem. Chem. Phys.*, 2014, **16**, 17845–17856.

13 C. D. Christ, M. Zentgraf and J. M. Kriegl, *J. Chem. Inf. Model.*, 2012, **52**, 1745–1756.

14 A. Bøgevig, H.-J. Federsel, F. Huerta, M. G. Hutchings, H. Kraut, T. Langer, P. Löw, C. Oppawsky, T. Rein and H. Saller, *Org. Process Res. Dev.*, 2015, **19**, 357–368.

15 Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Y. Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, 2015.

16 X.-J. Zhang, C. Shang and Z.-P. Liu, *The Journal of Chemical Physics*, 2017, **147**, 152706.

17 A. A. Gokhale, J. A. Dumesic and M. Mavrikakis, *J. Am. Chem. Soc.*, 2008, **130**, 1402–1414.

18 Z.-J. Zuo, L. Wang, P.-D. Han and W. Huang, *International Journal of Hydrogen Energy*, 2014, **39**, 1664–1679.

19 D. T. Messervy and K. E. Hayes, *Can. J. Chem.*, 1967, **45**, 629–633.

20 R. J. Madon, D. Braden, S. Kandoi, P. Nagel, M. Mavrikakis and J. A. Dumesic, *Journal of Catalysis*, 2011, **281**, 1–11.

21 Z. Zhang, S.-S. Wang, R. Song, T. Cao, L. Luo, X. Chen, Y. Gao, J. Lu, W.-X. Li and W. Huang, *Nat Commun*, 2017, **8**, 488.

22 L. He, H. Liu, C. Xiao and Y. Kou, *Green Chem.*, 2008, **10**, 619.