Matter, Volume *4*

Supplemental Information

Artificial Intelligence and QM/MM with a Polarizable Reactive Force Field for Next-Generation Electrocatalysts

Saber Naserifar, Yalu Chen, Soonho Kwon, Hai Xiao, and William A. Goddard III

Supplemental Experimental Procedures

1. Developing the vdW parameters of Cu and Au for RexPoN force field

To obtain the vdW parameters of Cu, we used the QM interaction energy of a water molecule with the Cu(100) surface (see Figure S1B). Here, we first minimized the water molecule on the surface using PBE-D3 and then scanned the water molecule along the *z* axis against the Cu surface. Then, the three atomic parameters of RexPoN universal nonbonded curve[1] were obtained by optimizing against the QM potential energy curve. We find the atomic parameters of Cu to be D_e=0.8089 kcal/mol, R_e=2.8985 Å, and L=0.4200. The comparison between RexPoN and QM potential energy curves are shown in Figure S1A.

Similarly, for the vdW parameters of Au, we used the QM interaction energy of a water molecule with the Au(100) surface (see Figure S2B). Using the same procedure as for Cu, we find the atomic parameters of Au to be D_e=2.9387 kcal/mol, R_e=2.5557 Å, and L=0.2547. The comparison between RexPoN and QM potential energy curves are shown in Figure S2A.

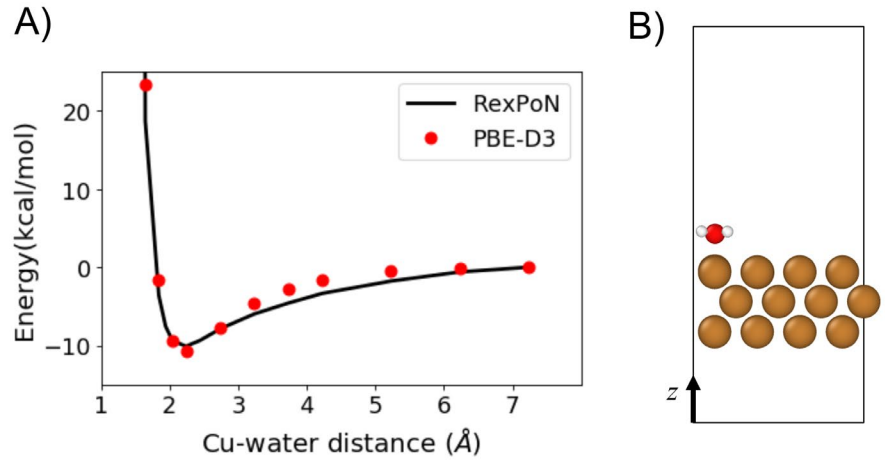

Figure S1. **Development of vdW parameters for Cu atom**. A) Comparison of RexPoN FF and PBE-D3 potential energy curves for the scan of water molecule against Cu(100) surface. B) the snapshot of the system for one of the scan points. The water molecule was scanned along the *z* axis.

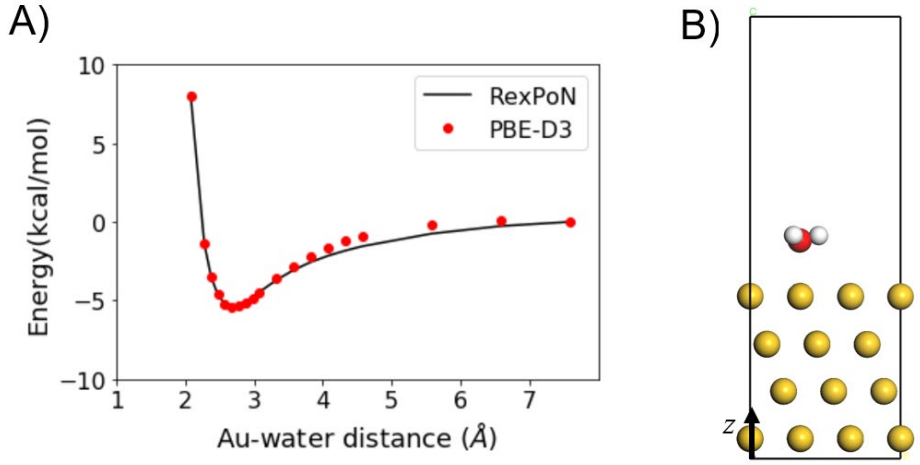

Figure S2. **Development of vdW parameters for Au atom**. A) Comparison of RexPoN FF and PBE-D3 potential energy curves for the scan of water molecule against Au(100) surface. B) the snapshot of the system for one of the scan points. The water molecule was scanned along the *z* axis.

2. ReQM potential energy plots

The changes of the potential energies as a function of time for each of the reaction intermediates during CO2RR (Figure 2 of the manuscript) are shown in Figure S3. The potential energy changes reasonably during the dynamics for all cases with no sudden change in the energies.

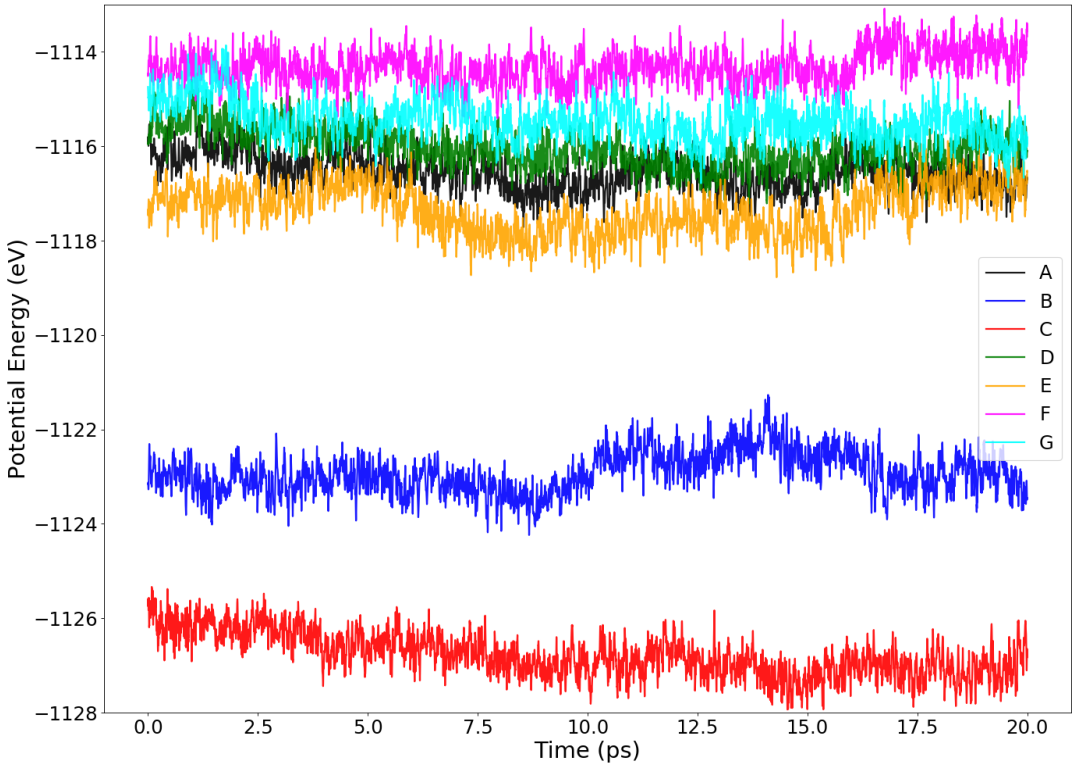

Figure S3. **Potential energies during NVT dynamics**. The change of the potential energy as a function of time for each of the reaction intermediates (Figure 2 of the manuscript) during CO_2RR.

3. Schematic of CO dimerization in explicit solvent

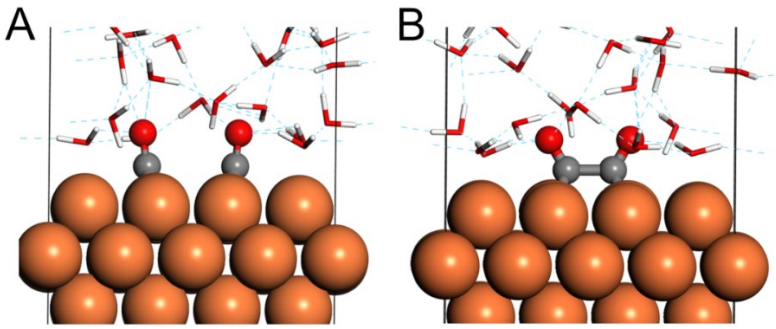

Figure S4. **CO dimerization in explicit solvent.** The snapshots of A) free CO* reactants and B) coupled *OC-CO products during CO dimerization. The CO dimerization is considered as the key step for the high C2 selectivity during CO2RR. Here, the Cu atoms are shown in orange, C in grey, H in white, and O in red. The hydrogen bonds are shown in as dashed red lines.

4. Solvation and minimization of AuNP surface sites in ReQM

4.1. Water box preparation

For the solvation of the AuNPs surface sites we first prepared a water box with 306 molecules at room temperature and 1 atm with density of 0.9965 gr/cm^3. To control the pressure we used a barostat with a relaxation time of 1 ps and to control temperature we used the Nosé-Hoover thermostat with damping time of 100 femtosecond (fs)[2]. We used a flexible model for the water molecules as our ReQM (force field and QM) simulations. We used a time step of 1.0 fs. We performed MD simulations in the NVT ensemble for 1 nanosecond (ns) to fully relax the water molecules at 300 K. The snapshot of the last frame of the equilibration is shown in Figure S5.

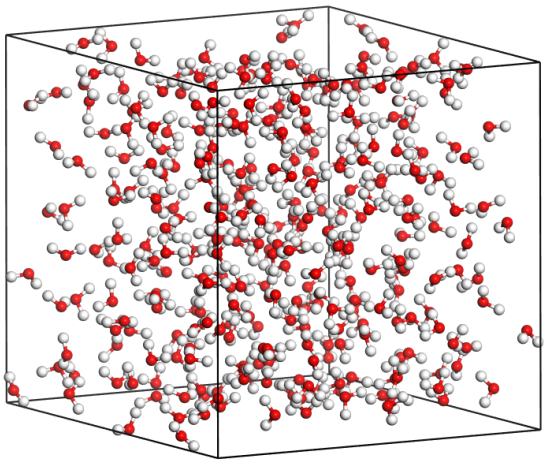

Figure S5. **Water box equilibration with RexPoN force field**. The snapshot of the water box with 306 molecules (20x20x22.96 Å^3) equilibrated at 300 K for 1 ns. This water box is used for the solvation of the AuNP surface sites.

4.2. Surface site insertion in water box

The equilibrated water box described above was used to solvate 2443 AuNP surface sites to train the neural network model. These surface sites were obtained from the work of Chen *et al*[3] , who optimized the position of CO and HOCO using vacuum QM calculations. Each surface complex was inserted in the middle of the water box. Next, any water molecule within the 3.0 Å cut off distance was removed from the system to avoid close contacts. Then, the water molecules were relaxed with RexPoN FF for 50 ps in NVT ensemble at 300 K while the surface atoms (Au, CO, and HOCO) were fixed. We used the same time step, thermostat, and barostat as in the previous section. The vdW, electrostatics, and hydrogen bonding between the water and atoms of the surface are described by RexPoN FF. We used RexPoN FF parameters as described in section 4.1 of the manuscript. The vdW parameters of Au (similar to Cu) were obtained by optimizing vdW parameters of RexPoN against the PBE-D3 potential energy curve of a water molecule scanned against the Au(100) surface (see section S1). After MD, we minimized the last frame to prepare the input structure for ReQM minimization. The structure of solvated Au-CO and Au-HOCO sites for two cases are shown in Figure S6.

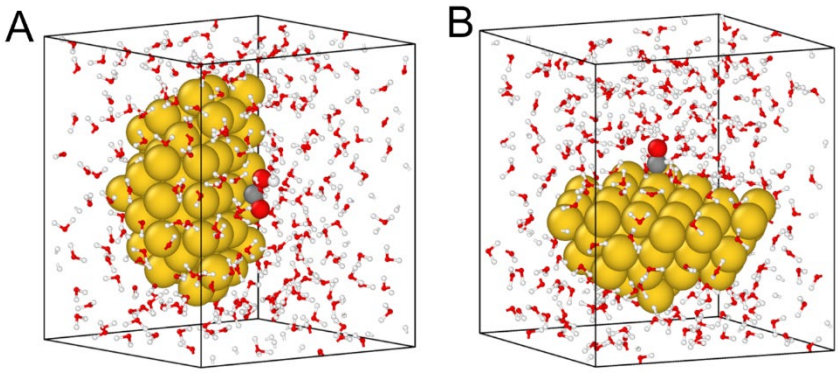

Figure S6. **Solvation of AuNPs surface sites**. The training set structures (1384 Au-CO sites and 1059 Au-HOCO sites) were solvated with water using ReQM. Two examples are shown here **A**) HOCO intermediate and **B**) CO adsorbate. Each cluster was inserted at the center of the equilibrated (at 300K) water box and water molecules within 3.0 Å of the cluster were removed to avoid close contacts. Then, the water molecules were relaxed at 300K while the cluster atoms were fixed. Finally, the coordinates of the adsorbents were minimized in ReQM. Here, the Au atoms are shown in yellow, carbon in grey, oxygen in red, and hydrogen in white. The atoms of the cluster (Au and adsorbent) are shown with larger spheres for better visualization.

4.3. Minimization of AuNP surface sites in ReQM

The final structure from the previous section was used in ReQM to minimize the coordinates of the CO and HOCO molecule on each surface site. Here, the QM region of the ReQM contains Au, adsorbent (CO or HOCO), and the water molecules that within 6 Å of the adsorbent. Although, the RexPoN FF describes hydrogen bonding between water and adsorbents, we include some water molecule in the QM region to allow accurate polarization and charge transfer within the QM region. During the ReQM minimization we relaxed only the position of CO or HOCO while other atoms were fixed. We use 0.05 eV/ Å as the force convergence criterion. We used Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm as implemented in the ASE package and the force convergence criterion of 0.05 eV/Å for the minimization.

5. Neural Network Based Machine Learning Model

We used a neural network based machine learning algorithm similar to the work of Chen *et al*[3]. We provide a brief summary of the neural network model here since more details are provided elsewhere[3]. For each given cluster with N atoms, we compute the interatomic distances (R_{ij}) for N-1 pairs between the Au atoms and the central Au atom (cAu). We transfer the computed R_{ij} values to input features using the following two-body (C2) and three-body (C3) symmetry functions,

$$C2_{m,i} = \sum_j f_m(R_{ij}), \tag{1}$$

$$C3_{mnl,i} = \sum_{jk} f_m(R_{ij}) f_n(R_{ik}) f_l(R_{jk}) \tag{2}$$

where *m, n*, and *l* are the indices of the symmetry functions, *i* is the index of the surface atom, and *j* and *k* are the indices of other Au atoms (see Figure S7A). Here, the C2 terms are constructed by summing over all Au and cAu pairs while the C3 terms are summed over three body terms that have cAu in one corner (see Figure S7C). f is the symmetry function given by a Localized Cosine Piecewise function,

$$f_m(R_{ij}) = \begin{cases} \frac{1}{2}\cos\left(\frac{R_{ij} - d_m}{r}\pi\right) + \frac{1}{2} & |R_{ij} - d_m| < r \\ 0 & Otherwise \end{cases} \tag{3}$$

where d_m and *r* are the center and width of symmetry function (see Figure S7B).

In the above model, we implement 12 C2 and 3 C3 symmetry functions, leading to a total number of 39 input features. We consider that this gives the best balance of dataset size and model complexity. We use these features with a fully connected two-layer neural networks having 30 nodes in the first layer and 50 nodes in the second layer (total of 2801 parameters) to fit two selected physical descriptors: ΔE_{CO}^{sol} and ΔE_{HOCO}^{sol}.

In summary, our model expresses the physical descriptors as a function of two-body and three-body terms with the weight and bias parameters *w* and *b*. Therefore, we can write the neural network function (F_{NN}) as

$$E_i = F_{NN}(C2_{m,i},\ C3_{mnl,i}; w, b) \tag{4}$$

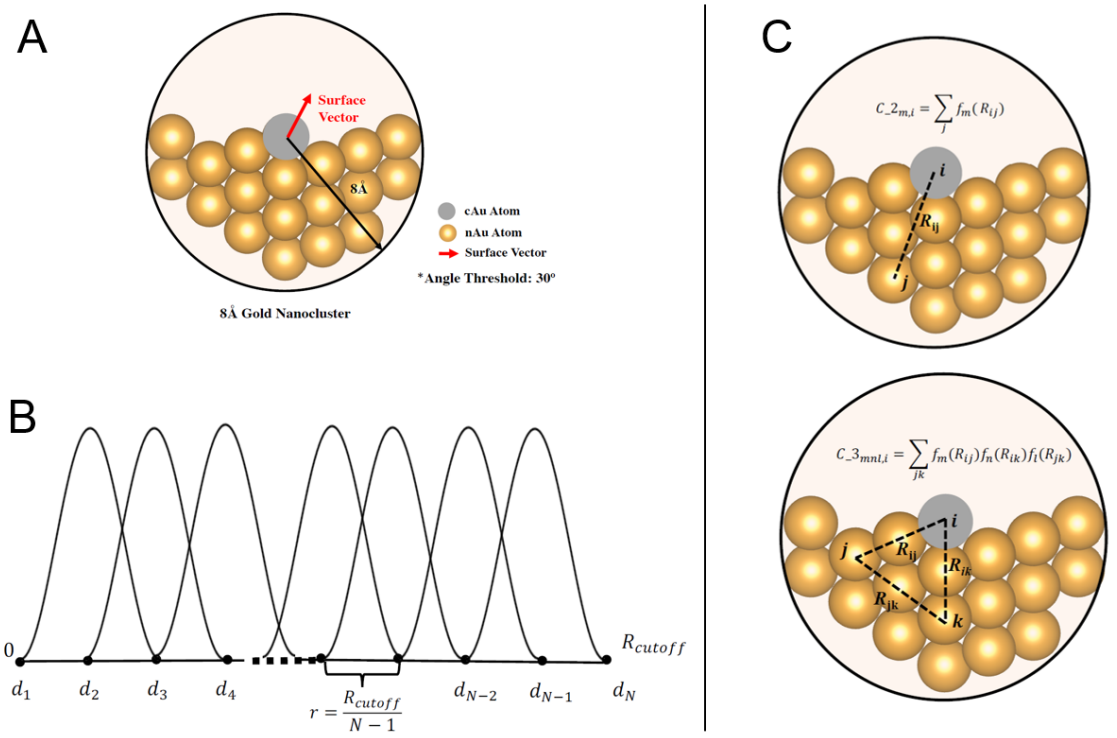

Figure S7. **Transferring geometric features to symmetry functions.** A) The surface vector method was used to identify surface and bulk atoms for an 8 Å nanocluster model. Here, cAu denotes center atom and nAu denotes other atoms in the nanocluster. B) The shape of a localized piecewise cosine symmetry function with width of *r*, center d_m, and R_{cutoff} = 8 Å. C) Representation of C2 and C3 symmetry functions given in Equations 1 and 2. The components of this figure are adopted from Chen *et al*[3].

6. Partition of the data to training, validation, testing sets

Table S1. **Machine learning data set for solvated CO adsorption energy**. Partition of data set to training, validation, and testing sets for solvated CO adsorption energy, ΔE_{CO}^{sol}. The size of the sets for different coordination numbers are set equal to each other to avoid making the model bias toward a specific set. The sites within each group have been selected randomly and the sets are independent from each other.

Data Set	Coordination				Size	Final RMSE (eV)
	6	7	8	9		
Training	296	296	296	296	1184	0.0748
Validation	25	25	25	25	100	0.0758
Testing	25	25	25	25	100	0.0691

Table S2. **Machine learning data set for solvated HOCO formation energy**. Partition of data set to training, validation, and testing sets for solvated HOCO formation energy, ΔE_{HOCO}^{sol}. The size of the sets for different coordination numbers are set close to each other to avoid making the model bias toward a specific set. The sites within each group have been selected randomly and the sets are independent from each other.

Data Set	Coordination				Size	Final RMSE (eV)
	6	7	8	9		
Training	224	214	209	212	859	0.0928
Validation	25	25	25	25	100	0.0969
Testing	25	25	25	25	100	0.1017

7. The change of RMSE as a function of the training epoch

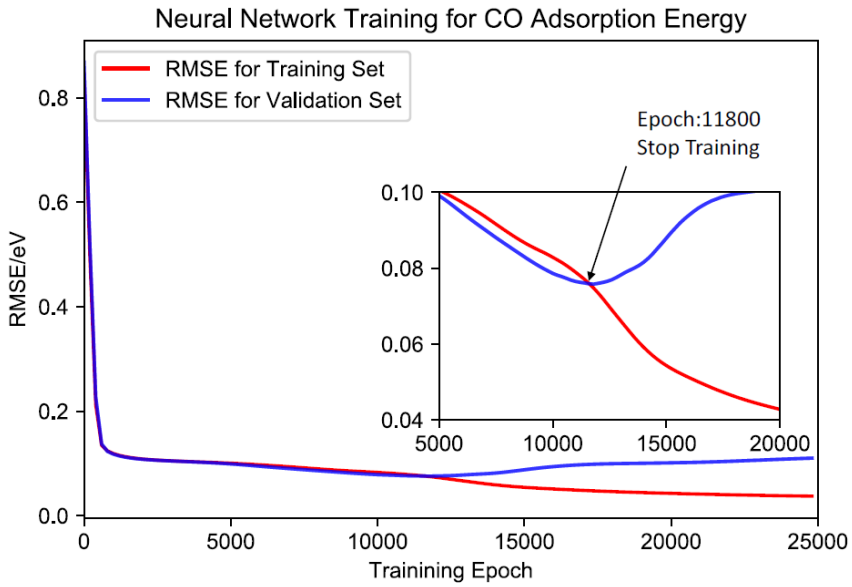

Figure S8. **Convergence of neural network training for ΔE_{CO}^{sol}.** The change of the RMSE as a function of training epoch for ΔE_{CO}^{sol}. The RMSE of validation sets help to avoid overfitting (early-stop). At epoch 11800, the training is stopped as RMSE of validation reaches to the minimum value of 0.0758 eV with the RMSE of training set at 0.0748 eV.

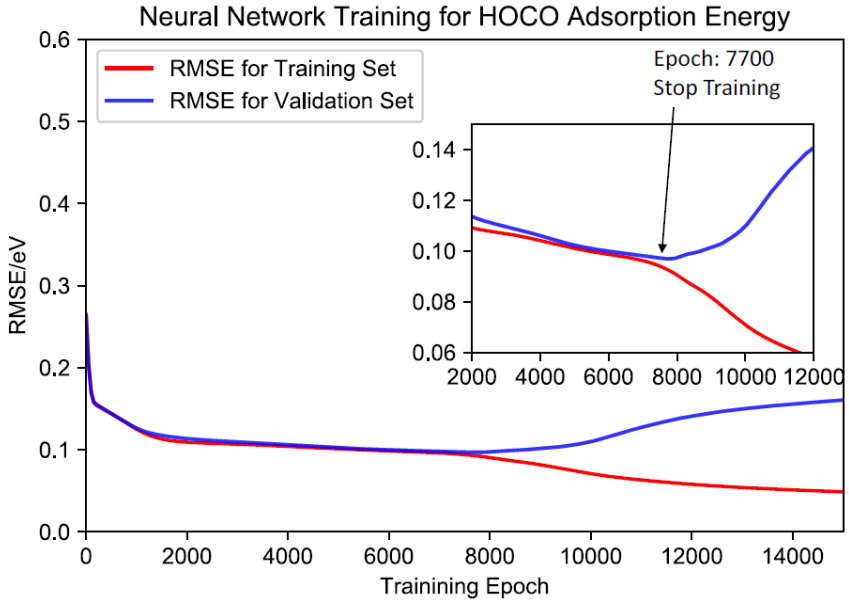

Figure S9. **Convergence of neural network training for ΔE_{CO}^{sol}.** The change of the RMSE as a function of training epoch for ΔE_{CO}^{sol}. The RMSE of validation sets help to avoid overfitting (early-stop). At epoch 7700, the training is stopped as RMSE of validation reaches to the minimum value of 0.0969 eV with the RMSE of training set at 0.0928 eV.

8. Description and structure of seven active groups

As explained in the manuscript, the top 300 sites (out of 11537) can be classified into seven groups. We provide in Figure S10 one representative structure for each group. Chen et al[3] explained the structure of each group as follows. The center atom in Step111 has Au(111) character, the StepUnder111 has under-coordinated Au(111) character, Step110 has 110 character, Step311 has 311 character, StepTB has twin boundary character, StepUnderTB has under-coordinated twin boundary features but with steps around the center atom and SurfaceDefect is the Au(111) surface but with one or two missing atoms around the center atom.

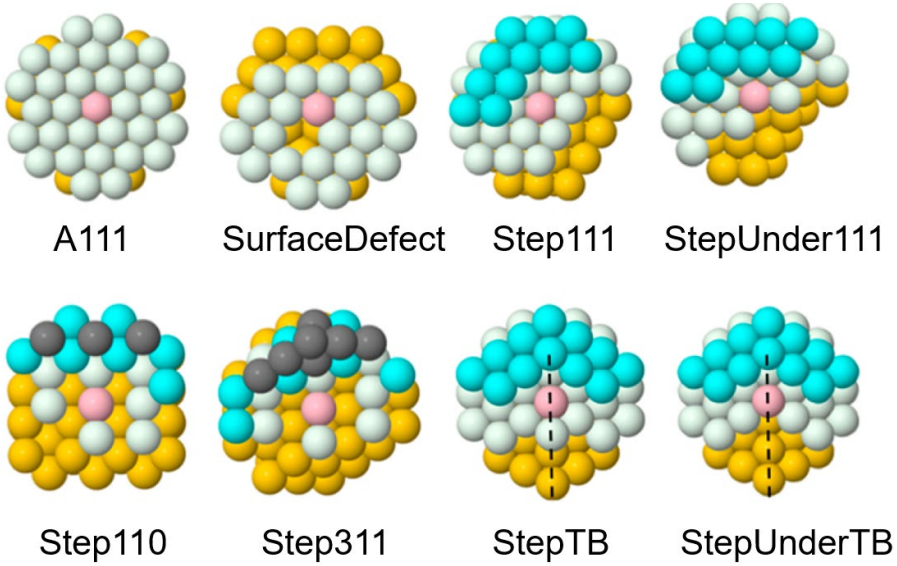

Figure S10. **Representation of AuNP active groups**. A representative structure for each of the seven active groups identified in the top 300 sites. The center atom is shown in pink while other atoms in same layer are shown in white. The color of atoms for the layer below the center atom is gold, for one layer above the center atom is cyan, and for two layers above the center atom is gray. The dashed lines show twin boundaries. This figure was adopted from Figure 4 of Chen *et al*[3].

9. Supplemental References

1. Naserifar, S., Oppenheim, J.J., Yang, H., Zhou, T., Zybin, S., Rizk, M., and Goddard, W.A. (2019). Accurate non-bonded potentials based on periodic quantum mechanics calculations for use in molecular simulations of materials and systems. J. Chem. Phys. *151*, 154111.
2. Hoover, W.G. (1985). Canonical dynamics: Equilibrium phase-space distributions. Phys. Rev. A *31*, 1695–1697.
3. Chen, Y., Huang, Y., Cheng, T., and Goddard, W.A. (2019). Identifying Active Sites for CO_2 Reduction on Dealloyed Gold Surfaces by Combining Machine Learning with Multiscale Simulations. J. Am. Chem. Soc. *141*, 11651–11657.