Supporting Information
The Stability and Dynamics of Zeolite-Confined Gold Nanoclusters

Siddharth Sonti,[†,‖] Chenghan Sun,[†,‖] Zekun Chen,[‡] Robert M. Kowalski,[¶] Joseph S. Kowalski,[§] Davide Donadio,[‡] Surl-Hee Ahn,[*,†] and Ambarish R Kulkarni[*,†]

†*Department of Chemical Engineering, University of California Davis*

‡*Department of Chemistry, University of California Davis*

¶*Department of Chemical Engineering, University of California Los Angeles*

§*Department of Biomedical Engineering, University of California Davis*

‖*Contributed equally*

E-mail: sahn@ucdavis.edu; arkulkarni@ucdavis.edu

Methods

Dataset generation for MLP training

The Vienna *ab initio* simulation package (VASP)[1] was used for DFT calculations of both initial structural relaxations and AIMD simulations. The Revised PBE from Hammer et al.[2] (RPBE) exchange-correlation functional in generalized gradient approximation (GGA) was employed. Dispersion corrections were considered with the DFT-D3 method with Becke-Johnson damping (D3BJ). Only the single (Γ)-point was used for K-space sampling considering the overall large sizes of the unit cell of Au@zeolite configurations (max size

of 303 atoms for 15Au@MFI). A plane-wave cutoff of 400 eV was used in all calculations, and the Kohn-Sham equations were solved self-consistently until electronic energies were converged to 10^{-6} eV. These values are similar to the parameters we adopted in previously work.[3]

To generate the initial candidates of Au@zeolite pore-confined configurations and ascertain the appropriate cluster sizes of the Au NCs for each zeolite, we devised an Au@zeolite configuration random generation algorithm anchored in Monte Carlo simulation principles.[4] As a prerequisite step, the zeolite frameworks were first optimized to the force constraints of 50 meV/Å using the same pseudo potentials, cutoff energy, and a single k-point, as previously outlined. Within the algorithm, the action space encompasses three distinct moves: the addition of a single Au atom into a unit cell of a given zeolite, its reinsertion, and its deletion. The associated probabilities for these actions are respectively set at 0.5, 0.3, and 0.2, which were picked to achieve optimal balance between configuration diversity and convergence within short time periods. For the addition, the newly introduced Au atom is ensured to be positioned within the same zeolite pore at a distance ranging between 2-4 Å from the nearest extant Au, Si, or O atom. The reinsertion comprises by two steps: random deletion of an existing Au atom followed by the addition of a new Au atom into the pore. If the algorithm gravitates toward the deletion probability, an existing Au atom is randomly removed from the configuration. The simulation procedure was iterated for a maximum of 1000 steps until it converged to the designated max size of Au NCs confined in each zeolite.

As a result of the Au@zeolite configuration random generation process, we found the Au NCs ranges from 1 to 15 atoms were successfully confined into AEI, BEA, CHA, LTA, MAZ, MOR, and RHO zeolite frameworks. In the context of the MFI zeolite framework, Au NCs encompassing a range of 1 to 12 atoms exhibited successful confinement. Similarly, within the MWW and SOD zeolite frameworks, Au NCs with size of 1 to 11 atoms were confined. The deliberate inclusion of a diverse initial set of reference data serves the purpose of enhancing the robustness and stability of the initial trained MLP. It is pertinent to note that while

our primary emphasis is directed towards the investigation of the diffusion characteristics of the 1-10Au@LTA systems, the inclusion of other zeolites in the initial datasets enabled us to develop an MLP which is transferable with respect to different topologies.

All of our initially simulated pore-confined multi-Au@zeolite configurations were fully relaxed and geometrically optimized with a force convergence criteria of 50 meV/Å before performing the AIMD simulations. As we primarily focused on developing MLP for studying pore-to-pore diffusion of multi-Au@LTA system, structural optimizations were performed for both the pore-confined and channel-confined (Au NCs were situated traversing the channels of LTA, same at each axis) LTA configurations, where the channel-confined structures provide AIMD initial trajectories originating from transition state conformations. Instead of imposing constraints on individual Au atoms, our approach entailed on utilizing channel-confined structures that have undergone a near-complete optimization. This approach is favored over employing fully optimized structures, given that the latter will eventually undergo confinement within the nearest pore. In the end, the gas phase Au NCs optimizations with sizes ranging from 1-15 atoms were performed using the same VASP parameters mentioned earlier. To mitigate the possibility of erroneous image-image interactions, we ensured that the dimensions of the simulation cell boxes are sufficiently spacious.

The NVT AIMD simulations at temperatures of 300, 800, and 1300 K were used to generate the initial reference data for the model training with the VASP parameters as previously stipulated. A broad temperature range was chosen to enhance the transferability of the model across varied physical conditions. Configurations were extracted from AIMD trajectories, with a minimum simulation time of 2 picoseconds (ps) and a time step of 0.5 femtoseconds (fs). Extended simulation times up to 5 ps were adopted for Au@LTA systems. Snapshots were extracted from these trajectories at intervals of 10 time steps. These snapshots were subsequently utilized to train an initial MLP model, which in turn facilitated prolonged classical MD simulations. The initial training dataset encompassed an ensemble of structures, including the following four systems: 1. Pore-confined 1-15 Au@Zeolites systems;

2. Channel-confined 1-10Au@LTA systems; 3. Frameworks of ten types of pure siliceous zeolites, and 4. Au NCs ranging in size from 1 to 15 atoms. As a result, a total of over 17,000 configurations were sampled as the reference training set for the SchNet model.

Training of polarizable atom interaction neural network (PaiNN) model

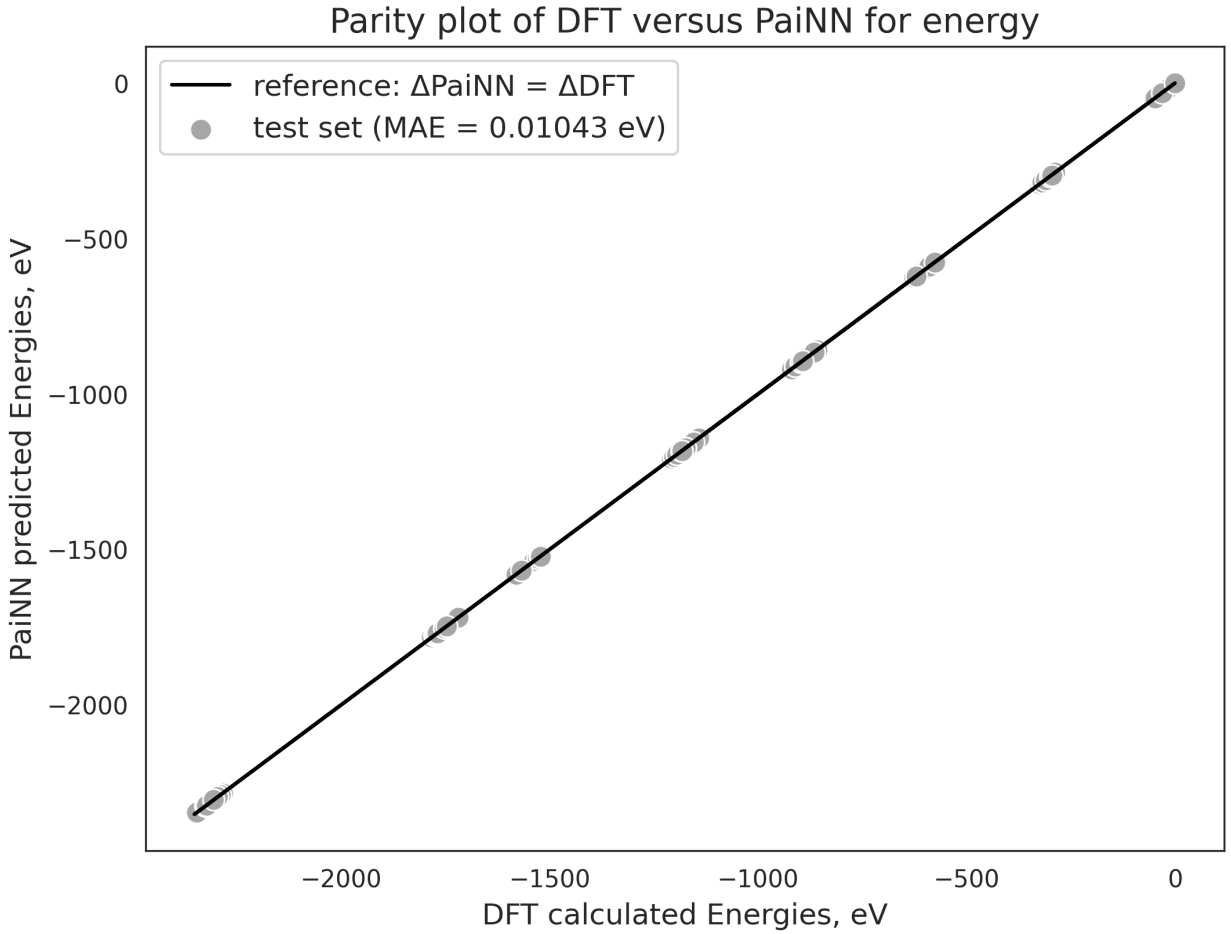

Figure (S1) Parity plots comparing energies predicted using polarizable atom interaction neural network (PaiNN) model with corresponding DFT-calculated references.

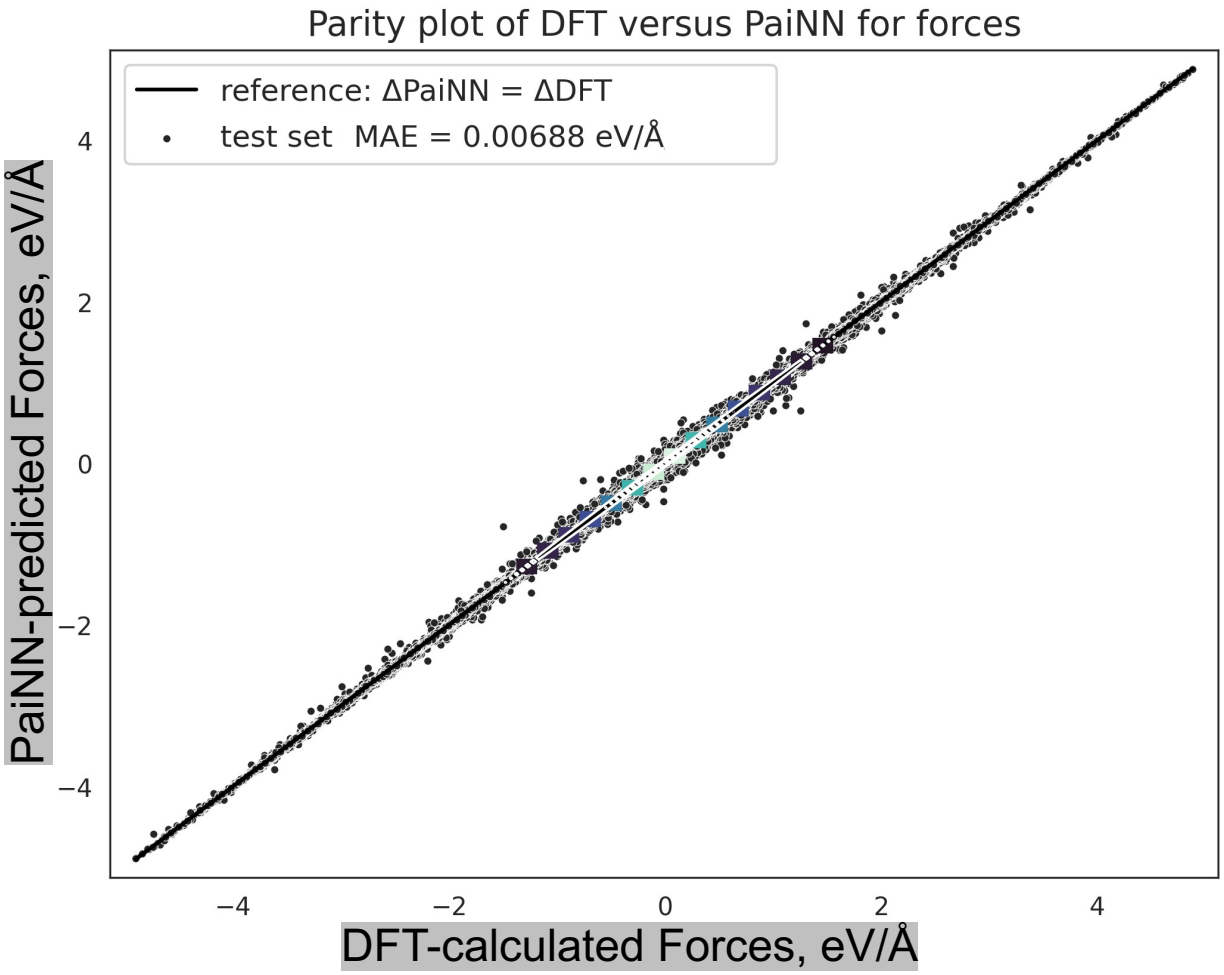

Figure (S2) Parity plots comparing forces predicted using polarizable atom interaction neural network (PaiNN) model with corresponding DFT-calculated references.

An auxiliary MLP model that used to extend the classical MD simulation length for 1-10Au@LTA conformations at the training data augmentation stage was trained via the polarizable interaction neural network (PaiNN) representation implemented in the SchNetPack (version 1.0) Package[5] with Pytorch extensions.[6] PaiNN is an equivariant message-passing neural network, which handles quantum interactions for molecule structures through a graph setup. Within the PaiNN representation, atom features are iteratively updated by the interaction between itself and neighboring atoms through continuous-filter convolutions scheme,[7,8] and its equivariant representations over angular features facilitate the propagation of scalar and tensorial properties of molecular structures.

The PaiNN model training was done over a random split part of the DFT reference data with 80% data points as training set and 10% data points as validation and test set, respectively. We adopted similar hyper-parameters (6 interaction layers, 128 atom-wise feature vectors and convolution filters, and 60 Gaussians for pairwise distance expansion) used by Erlebach et al.,[9] who developed neural network potentials for siliceous zeolites using the PaiNN architecture with the loss function trade-off of energy and forces to be 0.01 and 7.5Å cutoff. Minimization of the loss function used mini-batch gradient descent along with the AdamW optimizer[10] and batch size of 4 structures. The initial learning rate was 10^{-4} with a decay rate of 0.5 until a minimum value of 10^{-9}, if the loss function for the validation set did not decrease in 15 subsequent epochs. The PaiNN model for each iteration was trained for 300 epochs, unless matching with the criteria of early-stopping.

Training of neuroevolution potential (NEP) model with active learning

Table (S1) NEP Hyper-parameters Optimization Study Part I (approx. 22 hours per 20000 steps using one NVIDIA RTX A4000 GPU) for our NEP model development. Otherwise stated in the following table, details of the training processes remain the same as main manuscripts. All train-test performances are adopted from the end of 20000-step runs. Bold font indicates the protocol used in training the NEP5 model.

Protocols	E_{train} $(eV/atom)$	F_{train} $(eV/Å)$	E_{test} $(eV/atom)$	F_{test} $(eV/Å)$
Loss weight: 0.5 / 0.5	0.01332	0.13584	0.01641	0.13740
Loss weight: 0.4 / 0.6	0.02041	0.13407	0.02201	0.13367
Loss weight: 0.3 / 0.7	0.01646	0.11089	0.02023	0.11268
Loss weight: 0.2 / 0.8	0.01843	0.11140	0.01741	0.11336
Loss weight: 0.1 / 0.9	**0.01623**	**0.10374**	**0.01624**	**0.10526**
Loss weight: 0.05 / 0.95	0.03403	0.10012	0.03726	0.09976
Loss weight: 0.01 / 0.99	0.05929	0.09723	0.06343	0.09830

Table (S2) NEP Hyper-parameters Optimization Study Part II (approx. 22 hours per 20000 steps using one NVIDIA RTX A4000 GPU) for our NEP model development. Otherwise stated in the following table, details of the training processes remain the same as main manuscripts. All train-test performances are adopted from the end of 20000-step runs.

Protocols	E_{train} $(eV/atom)$	F_{train} $(eV/\text{Å})$	E_{test} $(eV/atom)$	F_{test} $(eV/\text{Å})$
Loss weight: 0.1 / 0.9, $R_{cut}^{radial} = R_{cut}^{angulr} = 8\text{Å}$	0.01972	0.10041	0.02030	0.10255
Loss weight: 0.1 / 0.9, $R_{cut}^{radial} = R_{cut}^{angulr} = 6.5\text{Å}$	0.01988	0.09739	0.02025	0.09746
Loss weight: 0.1 / 0.9, neuron=30	0.02310	0.08948	0.02506	0.09107
Loss weight: 0.1 / 0.9, neuron=70	0.01970	0.11076	0.02259	0.11122

The recently developed neuroevolution potential (NEP) approach has been widely applied for various MD studies, including investing the phonon mean free path for metal-organic frameworks, probing Tungsten under the radiation damage, and modeling the thermodynamics abnormalities of water and ice.[11–13] The NEP can be used as a classical force field in Graphics Processing Units Molecular Dynamics (GPUMD)[14,15] (version 3.8). An NEP potential constitutes a fully connected feed-forward neural network (FFNN), which has information in the model flowing in only one direction. The input of such an FFNN consists of the radial and angular descriptors. These descriptors are rotation, translation and permutation invariant functions of atomic distances and serve the purpose of describing two-body and three-body interactions for configurations sampled in the training set.[15] Representing the radial descriptors gives:

$$r_n^i = \sum_{j \neq i} \sum_{k=0}^{N_{\text{bas}}^{radial}} c_{nk}^{ij} F_k\left(R_{ij}\right) \ \ with\ 0 \leq n \leq n_{max}^{radial}, \tag{1}$$

where the term r_n^i is defined as the n^{th} radial component for atom i, R_{ij} renders the atomic distances between two types of atoms i and j, n_{max}^{radial} symbolizes the upper limit of radial basis set components indexed by n, and N_{bas}^{radial} represents the number of radial basis functions. The tunable parameters set c_{nk}^{ij} depend on prior defined indices i, j, n, as well as

index k, which indicates the k^{th} order polynomial embedded in term $F_k(R_{ij})$ and updated during the training of our NEP potential. Inside the above radial descriptor expression, the term $F_k\left(R_{ij}\right)$ has the following form:

$$F_k\left(R_{ij}\right) = \frac{1}{2}\left[T_k\left(2\left(\frac{R_{ij}}{R_{cut}^{radial}} - 1\right)^2 - 1\right) + 1\right] f_{\text{cut}}\left(R_{ij}\right). \tag{2}$$

In Equation 2, T_k stands for the Chebyshev polynomial. $R_{\text{cut}}^{\text{radial}}$ represents the radial cutoff. $f_{\text{cut}}\left(R_{ij}\right)$ is a cutoff function to smoothly decay the r_n^i to 0 as the atom pairs are gradually separated beyond $R_{\text{cut}}^{\text{radial}}$. In principle, any cutoff function can be applied, while the traditional cutoff function for the Tersoff potential[16] is used to formulate r_n^i. After defining the radial term, one can express the angular components as follow:

$$A_{nl}^i = \sum_{m=-l}^{l} (-1)^m \Lambda_{nlm}^i \Lambda_{nl(-m)}^i. \tag{3}$$

In equation 3, n and l are integers and upper bounded by two user-defined parameters n_{max}^A and l_{max}^{3b}, which are set to control number of maximum angular components, and the degree l of polynomials to describe three-body interactions, respectively. The total angular component is a summation over $-l$ to l as indexed by m. Expanding Λ on polar and spherical coordinates yields:

$$\Lambda_{nlm}^i = \sum_{j \neq i} c_{nk}^{\prime\, ij} F_k'\left(R_{ij}\right)\left(R_{ij}\right) Y_{lm}\left(\theta_{ij}, \phi_{ij}\right), \tag{4}$$

where $Y_{lm}\left(\theta_{ij}, \phi_{ij}\right)$ represents the spherical harmonics function. It is worth to note that the $c_{nk}^{\prime\, ij}$ and $F_k'\left(R_{ij}\right)$in equation 4 share the same form as their counterparts in equation 1, but $R_{\text{cut}}^{\text{radial}}$ is replaced by an angular cutoff $R_{\text{cut}}^{\text{angular}}$. In analogous to A_{nl}^i, angular descriptors depicting the four-body interactions can be shown as the product sums of A_{nl}^i when permuting all possible n, l, m indices:

$$A^i_{nl_1l_2l_3} = \sum_{m_1=-l_1}^{l_1} \sum_{m_2=-l_2}^{l_2} \sum_{m_3=-l_3}^{l_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \times \Lambda^i_{nl_1m_1} \Lambda^i_{nl_2m_2} \Lambda^i_{nl_3m_3}. \quad (5)$$

The angular components can be extended to describe up to five-body interactions. However, as shown in multiple studies deploying NEP potentials, incorporating five-body terms in the angular components has little improvement on the model accuracy.[17–19] Thus, to balance accuracy and efficiency, we chose to leave out five-body descriptions when constructing our NEP potential. With radial and angular components being decided, the FFNN, in which hyperbolic tangent acts as the activation function, is optimized via separable natural evolution strategy (SENS).[20] Since virial data is not accessible in each of the training configuration, the loss function only consists of the sum of the root mean square error (RMSE) of energy and force, adjusted by their weighting factors λ_e and λ_f. To prevent overfitting, both norm-1 (λ_1) and norm-2 (λ_2) regularizations are pre-estimated via the weighted sum of the loss function upon convergence and imposed during training.[21,22] Among different parameters used to construct the NEP potential, $R^{\text{radial}}_{\text{cut}}$ and $R^{\text{angular}}_{\text{cut}}$, as well as λ_e and λ_f remain the most impactful ones: $R^{\text{radial}}_{\text{cut}}$ and $R^{\text{angular}}_{\text{cut}}$ dictate the range of which the potential senses different Au NCs, while λ_e and λ_f control the static weighting of each energy and force data point during the optimization. Thus, we executed a series of model hyper-parameters sensitivity studies (more details can be found in Table S1) to determine the optimal parameters and summarized in table S3.

Table (S3) Optimal hyper-parameters for the NEP potential used for Au@zeolite systems.

R^{radial}_{cut}(Å)	$R^{angular}_{cut}$(Å)	N^{radial}_{bas}	$N^{angular}_{bas}$	n^{radial}_{max}
7.5	7.5	12	12	8
$n^{angular}_{max}$	$l^{max}_{3b}/l^{max}_{4b}/l^{max}_{5b}$	N_{neu}	λ_e	λ_f
8	4/2/0	50	0.1	0.9
λ_1	λ_2	N_{bat}	N_{pop}	N_{gen}
0.05	0.05	1000	50	100000

The adaptive sampling framework was used to screen out duplicate configurations in the data set so that they are not repeated in the MLP training data set. We systematically

introduced training configurations from the adaptive metadynamics simulations[23] by employing an "on-the-fly" active learning strategy, which stems from the query-by-committee (QBC).[24–26] In recent years, many different active learning strategies[27–29] have been developed to address the issues related with sampling efficiency and uncertainty quantification in MLP developments. However, QBC was employed in our adaptive sampling framework as it is designed for models where an explicit uncertainty estimation is not available, thus made direct fit for neural networks. This strategy is achieved by training a couple of committee models, where each model is developed with the same dataset but different train/validation splits, as well as different initialization of hyper-parameters. In NEP, the maximum standard deviations of forces predicted from each committee was used as the uncertainty estimate metric to decide whether a candidate configuration should be included into the next iteration of training set, and can be quantified by the following equation:

$$\sigma_f = \max_i \sqrt{\sigma_{i,x}^2 + \sigma_{i,y}^2 + \sigma_{i,z}^2}, \tag{6}$$

where σ_i for each atom indexed by i along each $x - y - z$ Cartesian direction symbolizes the model variance computed among the committee models. To develop our NEP potential, we carried out multiple rounds of QBC along with the iterative model refinements (first iteration as NEP1, second as NEP2 and so on), during each iteration we constructed a committee consisting of five potentials. The candidate configurations were generated using the first potential, followed by evaluations from the other four potentials. This assessment aimed to rigorously ascertain the extent of uncertainty associated with each candidate, thereby facilitating an informed determination regarding its inclusion within the cohort of candidates under consideration. To optimize the performance of the first NEP potential, we employed the optimized NEP model hyper-parameters detailed in Table S3. We used the entire previous reference database (100%) as the training set (with validation handled internally via NEP model training) to train this NEP potential from scratch, aiming to maximize its accu-

racy for candidate generation. For each committee model, training was conducted for 5000 generations as used by the SNES algorithm,[20] employing the optimized parameters from Table S3, except for setting the norm-1 (λ_1) and norm-2 (λ_2) regularization values to -1, indicating automatic determination based on actual loss terms for physical quantities (energy and forces). The auto-adjusted penalty terms also provide different sets of parameter initialization for each of the committee potential. To maximize diversity among the committee ensemble, iterations NEP2-NEP4 utilized only a random subset of 75% of the reference data from the previous iteration for constructing each committee model. The remaining 25% served as a test set to monitor committee performance. In the final iteration (5), 95% of the reference data from the previous iteration (4) was used, as the number of sampled configurations had converged.

Table (S4) Details of the data involved in training different NEP iterations (AIMD/metadynamics), and the type of simulations performed with each NEP iteration.

NEP iteration	AIMD	Metadynamics	Active learning	Temperature	Transferability
NEP1	Yes (2 ps)	No	No	300 K	No
NEP2	No	Yes (10 ps)	Yes	300 K	No
NEP3	No	Yes (100 ps)	Yes	300 K, 500 K	Yes
NEP4	No	Yes (200 ps)	Yes	300 K, 500 K	Yes
NEP5	No	Yes (500 ps)	No	300 K, 500 K, 700 K	Yes

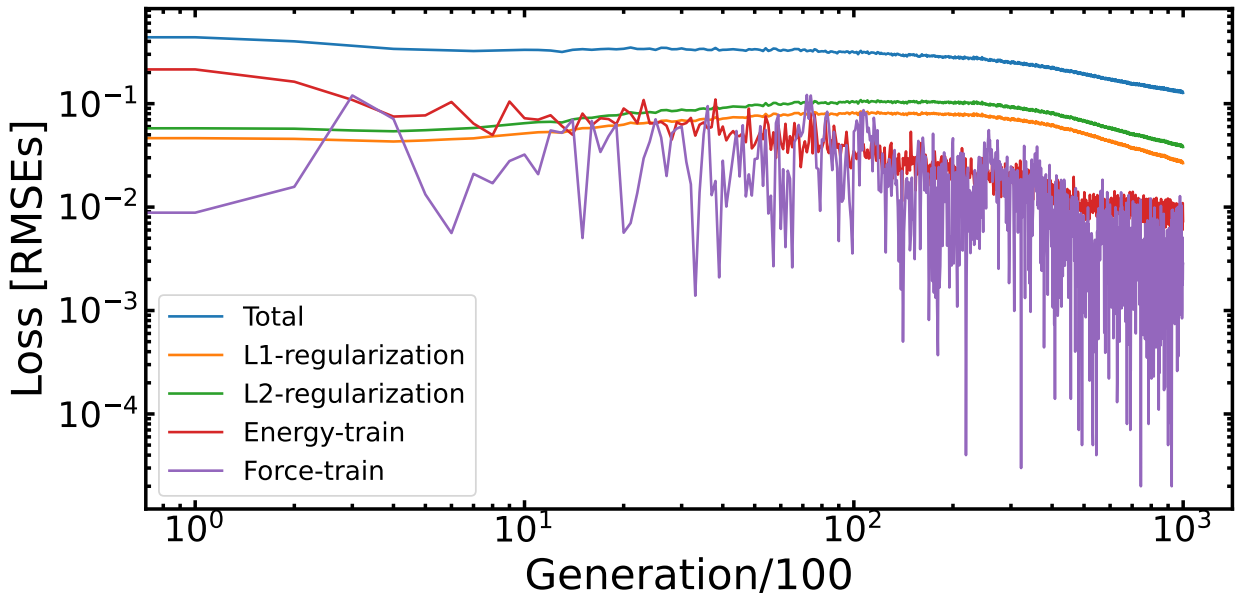

Figure (S3) Progression of loss function for per hundred generations.

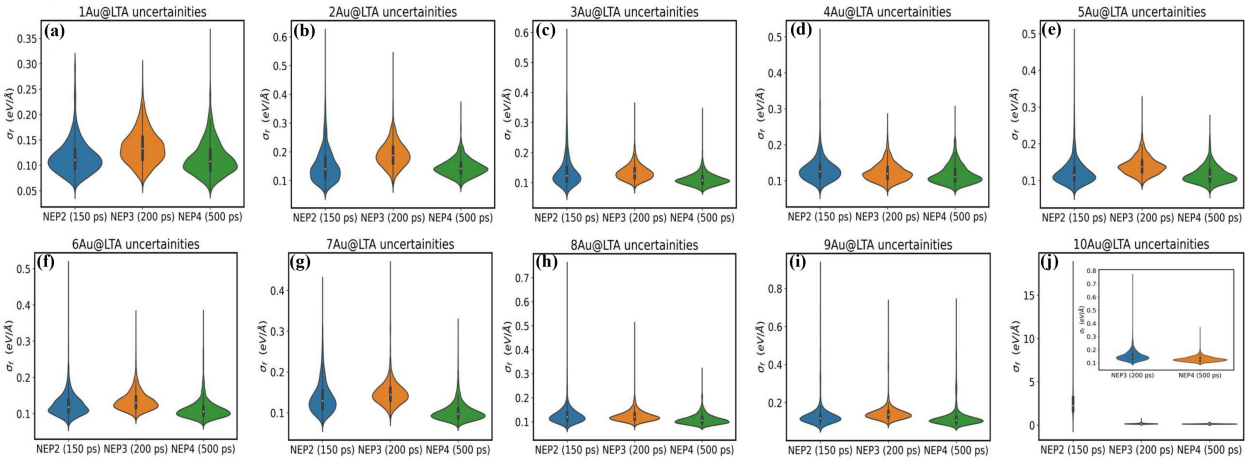

Figure (S4) The σ_f uncertainity estimation for the 2^{nd} to 4^{th} iterations of NEP in Au_{1-10}@LTA.

Results and Discussion

Transferability of the NEP model across different temperatures

To test the transferability of the final NEP MLP with respect to temperature and to investigate any potential sampling biases stemming from its training before applying to longer

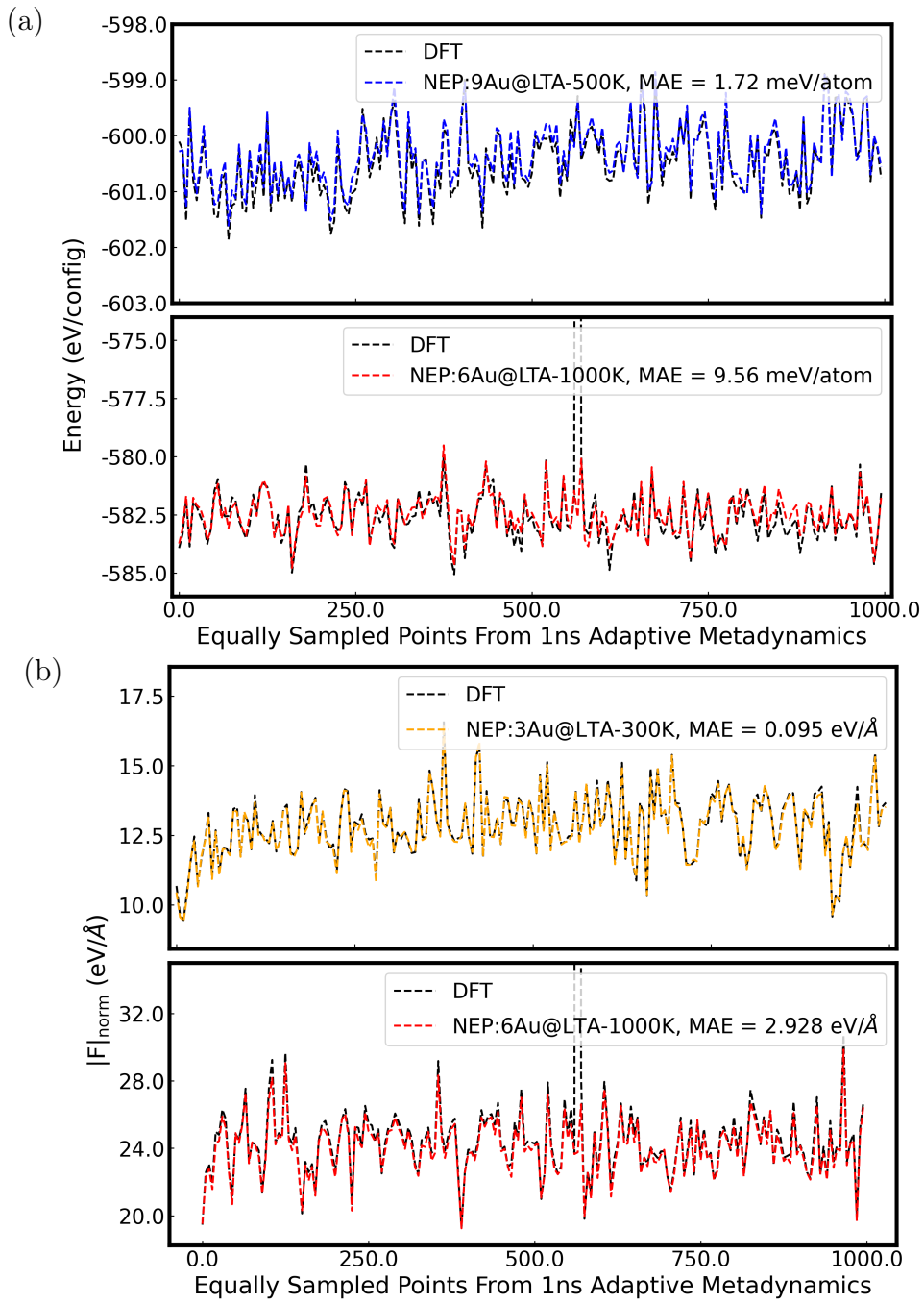

Figure (S5) Comparison of the parity between NEP and DFT predictions as a function of time (or a line plot) for 3/6/9 Au@LTA systems at 300 K, 500 K, 700 K, and 1000 K. (a) The potential energy line plot for the best (blue curve, 9Au@LTA at 500 K) and worst (red curve, 6Au@LTA at 1000 K) MAEs. (b) The force norm line plot for the best (yellow curve, 3Au@LTA at 300 K) and worst (red curve, 6Au@LTA at 1000 K) MAEs. The black line in all plots represent the DFT values.

Table (S5) Summary of MAEs of energy and force norm for transferability of Au@LTA systems across different temperatures.

Au Cluster	Temperature (K)	MAE of Energy (meV/atom)	MAE of Force Norm (eV/Å)
3-Au	300 K	4.52	0.095
6-Au	300 K	4.92	0.122
9-Au	300 K	2.15	0.124
3-Au	500 K	4.53	0.099
6-Au	500 K	3.74	0.132
9-Au	500 K	1.72	0.133
3-Au	700 K	5.23	0.133
6-Au	700 K	3.48	0.180
9-Au	700 K	1.78	0.185
3-Au	1000 K	N/A	N/A
6-Au	1000 K	9.56	2.928
9-Au	1000 K	4.18	0.280

timescales, we performed simulations for 3/6/9Au@LTA systems for 1 ns at four different temperatures (300 K, 500 K, 700 K, and 1000 K) to test its performance across a broad temperature range. 200 configurations were sampled at equal time intervals (equivalent to 1 configuration per 5 ps) from each simulation to run DFT single point calculations. Figures S5(a) and S5(b) show the parity between NEP and DFT predicted potential energies and forces respectively as a function of simulation time (also called a line plot), for systems with the highest and lowest MAE. Table S5 shows the MAEs of all the systems that were simulated. Considering that forces are three-dimensional in Cartesian coordinates, we evaluated the force norms and presented them in line plots. We observed an excellent agreement between the potential energies and force norms predicted by the NEP and DFT, with simulations at 300 K, 500 K and 700 K having lower MAEs compared to 1000 K due to the MLP having limited exposure ($\sim$9% of the overall dataset) to data above 500K.

Scalability of the NEP model across different zeolite topologies

The scalability of the NEP MLP across different purely silicious zeolite topologies was tested by performing adaptive metadynamics simulations for 200 ps using the NEP4-MLP model. The simulations were done for 3/6/9 Au NCs confined within BEA, CHA, MFI

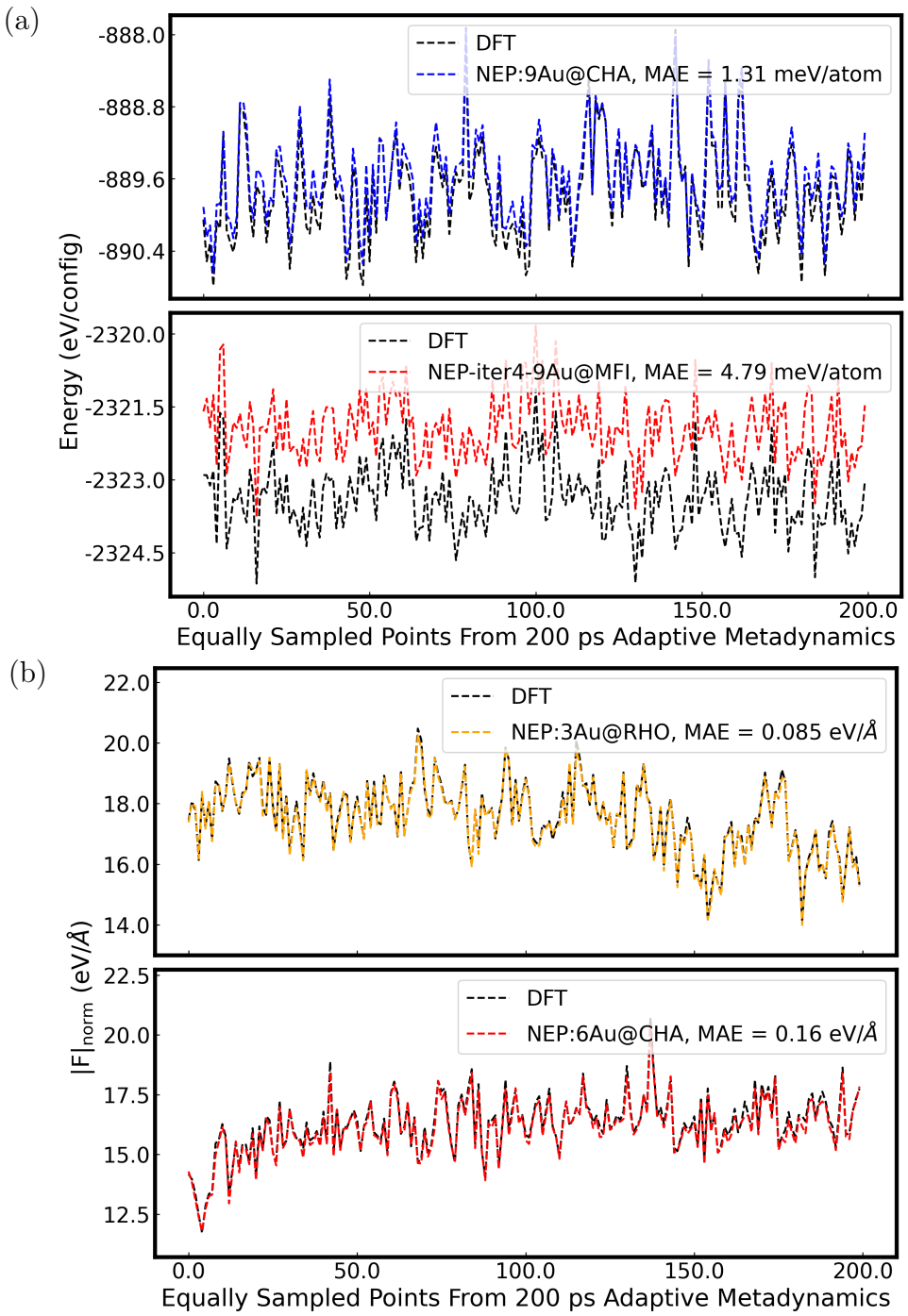

Figure (S6) Comparison of the parity between NEP and DFT predictions as a function of time (or a line plot) for 3/6/9 Au@zeolite systems (BEA,MFI,CHA) and 3/4/5 Au@zeolite systems (RHO) at 300 K. (a) The potential energy line plot for the best (blue curve, 9Au@CHA) and worst (red curve, 9Au@MFI) MAEs.(b) The force norm line plot for the best (yellow curve, 3Au@RHO) and worst (red curve, 6Au@CHA) MAEs. The black line in all plots represent the DFT values.

Table (S6) Summary of MAEs of energy and force norm for scalability of the NEP MLP across different purely silicious zeolite topologies. All the simulations were performed under 300K.

Au Cluster	Zeolite	MAE of Energy (meV/atom)	MAE of Force Norm (eV/Å)
3-Au	BEA	2.72	0.103
6-Au	BEA	2.08	0.109
9-Au	BEA	1.40	0.113
3-Au	CHA	2.42	0.097
6-Au	CHA	1.72	0.160
9-Au	CHA	1.31	0.131
3-Au	MFI	2.59	0.089
6-Au	MFI	4.54	0.107
9-Au	MFI	4.79	0.103
3-Au	RHO	1.53	0.085
4-Au	RHO	1.57	0.106
5-Au	RHO	2.39	0.091

zeolites, and 3/4/5 Au NCs in the RHO zeolite. These zeolites are characterized by larger pore sizes and consist of 192 atoms in BEA, 108 in CHA, 288 in MFI, and 144 in RHO as compared to 72 atoms in LTA. We uniformly sampled 200 configurations from the 200 ps simulation (equivalent to 1 configuration per 1 ps). This test presented a significant challenge, as these meso to macro porous zeolites were introduced to MLP training in only one iteration and constituted approximately only 4% of the final dataset. Nonetheless, our results showed an excellent agreement between the NEP predictions and DFT for both the potential energy and force norm line plots (see Figure S6 and Table S6). We observed a consistent, yet small bias in the error for the potential energy of 3/6/9Au@MFI systems. This due to the substantial number of atoms within the MFI framework, limiting the chemical locality information that can be leveraged for similar features across configurations. However, the accuracy in the forces norm plot remains accurate and consistent across all four zeolite types, as our MLP development prioritized force loss terms over energy terms, as forces are critical in an MD simulation.

Radial distribution function (RDF) of Au interactions

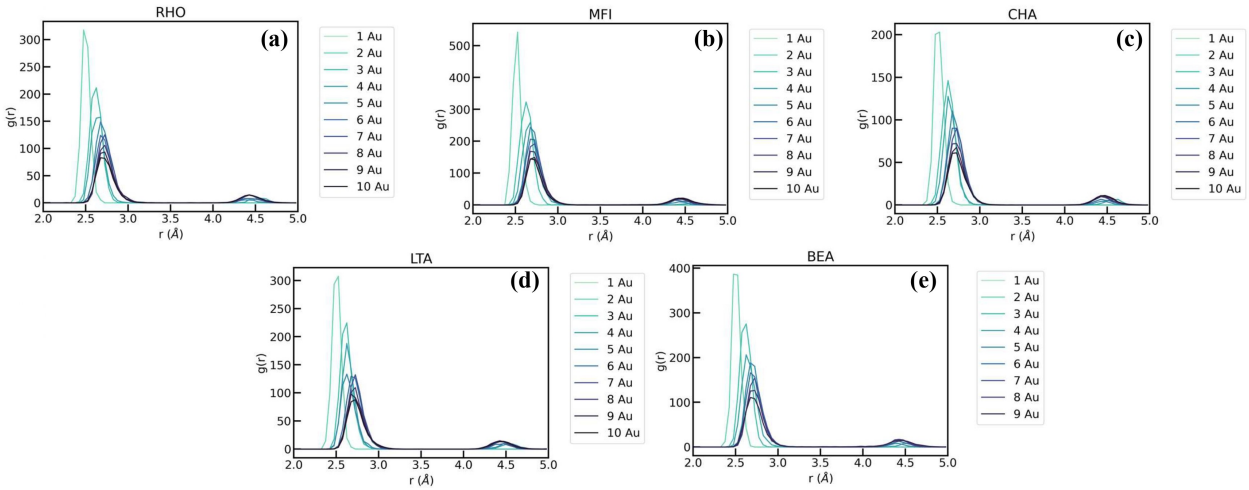

Figure (S7) The Au-Au RDF of Au_{1-10} in a) RHO, b) MFI, c) CHA, d) LTA, e) BEA.

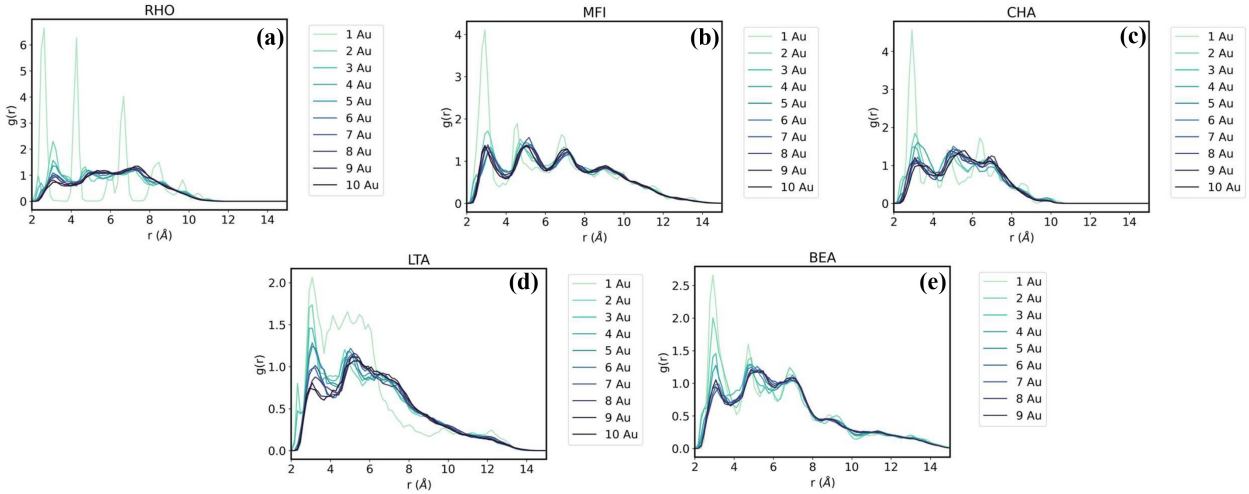

Figure (S8) The Au-O RDF of Au_{1-10} in a) RHO, b) MFI, c) CHA, d) LTA, e) BEA.

Averaged Free energy surfaces of Au_3@LTA at different temperautures

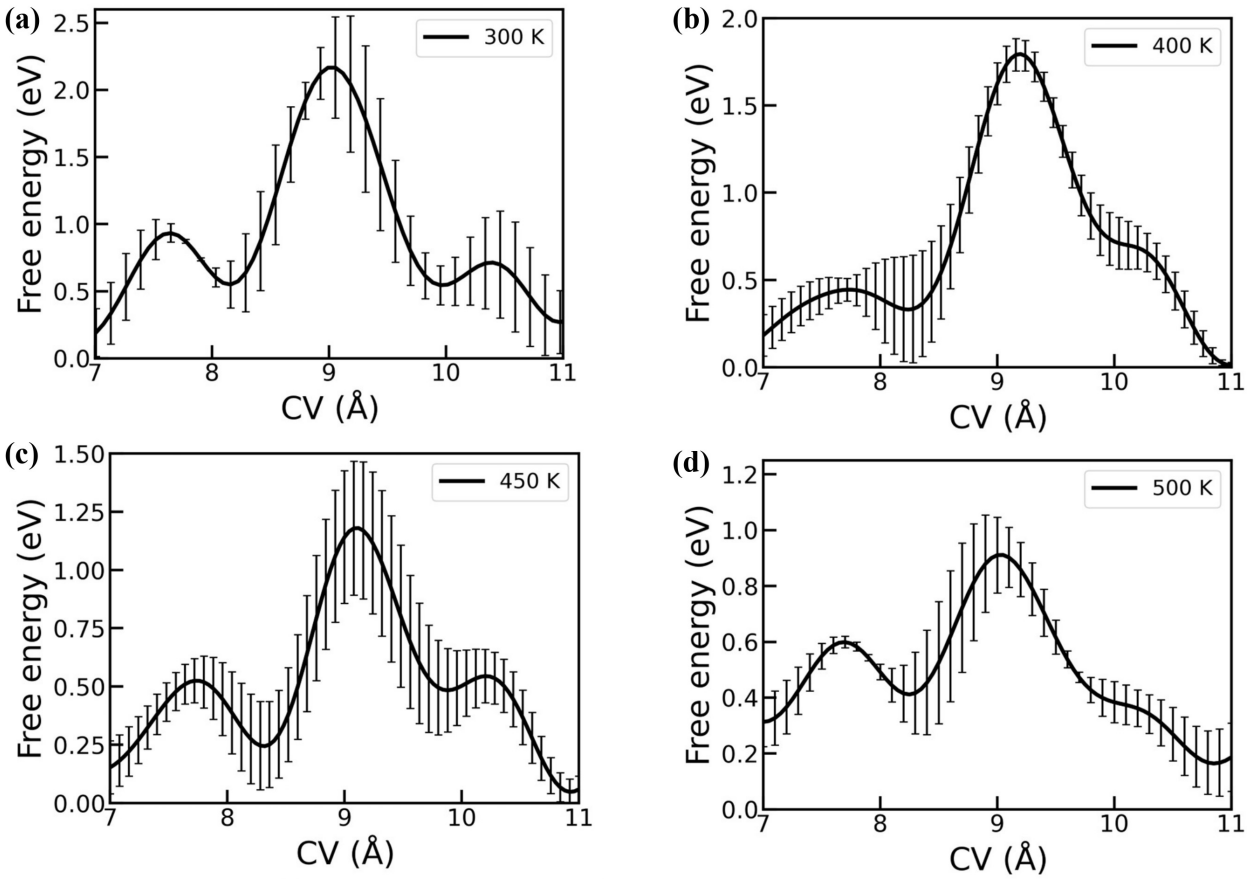

Figure (S9) The averaged free energy surfaces over 3 independent runs of Au_3@LTA at a) 300 K, b) 400 K, c) 450 K, d) 500 K

Dynamics of Au_5 in LTA and CHA

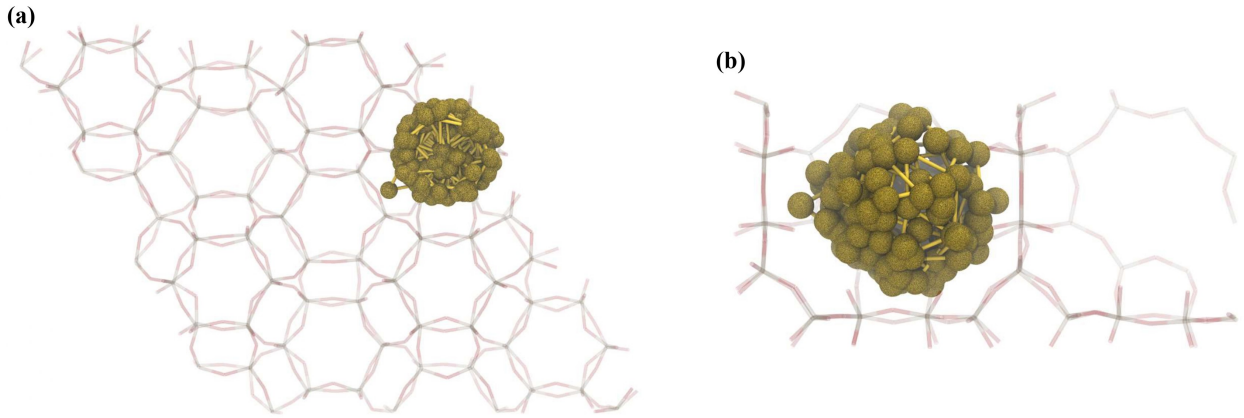

Figure (S10) (a) The positions occupied by an Au_5 nanocluster in a CHA zeolite from a 10 ns metadynamics simulation conducted at 500 K. (b) The positions occupied by an Au_5 nanocluster in an LTA zeolite from a 10 ns metadynamics simulation conducted at 500 K.

The 10 ns simulations for Au_5@CHA and Au_5@LTA showed that Au_5 is more restricted in CHA compared to LTA, as CHA has a smaller pore size (Figure S11). The implication of this results is the possibility of Au_5@CHA being a more accessible catalytic site for reactants, subject to the space available in the pore for reactants to diffuse.

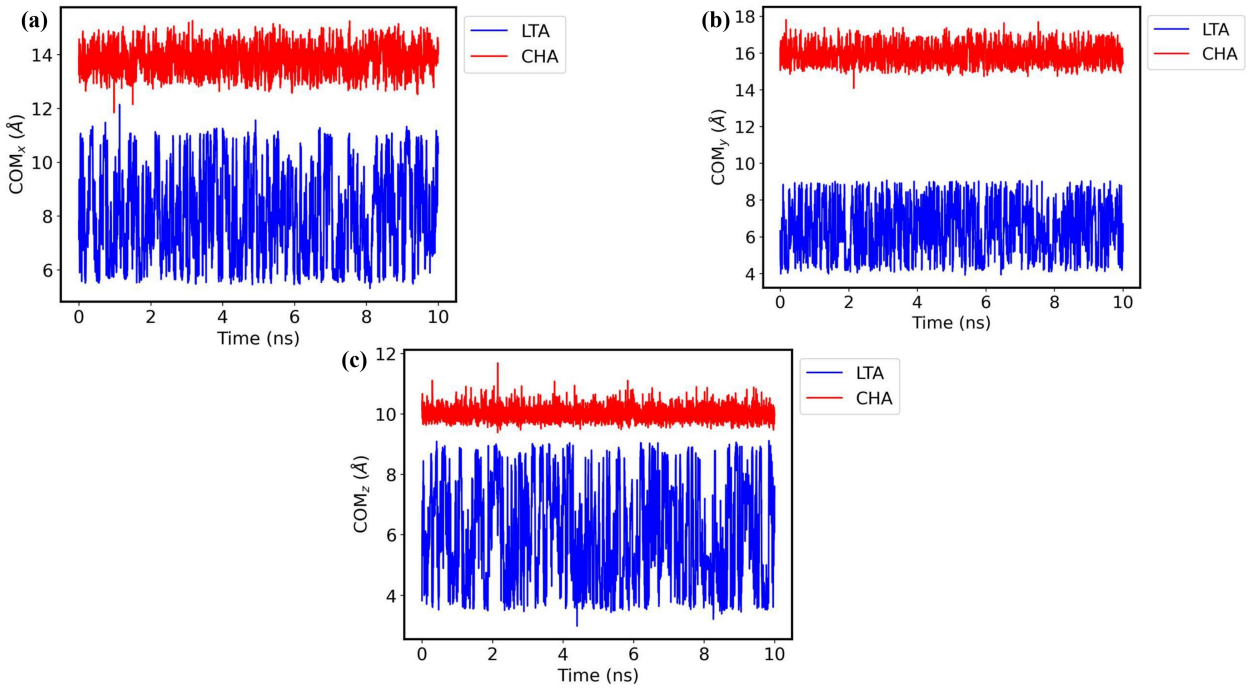

Figure (S11) The variations of centre of mass (COM) of Au_5 in CHA (red) and LTA (blue) as a function of simulation time in (a) x-direction (b) y-direction (c) z-direction.

References

(1) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical review B* **1996**, *54*, 11169.

(2) Hammer, B.; Hansen, L. B.; Nørskov, J. K. Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Physical review B* **1999**, *59*, 7413.

(3) Sours, T. G.; Kulkarni, A. R. Predicting structural properties of pure silica zeolites using deep neural network potentials. *The Journal of Physical Chemistry C* **2023**, *127*, 1455–1463.

(4) James, F. Monte Carlo theory and practice. *Reports on progress in Physics* **1980**, *43*, 1145.

(5) Schütt, K.; Kessel, P.; Gastegger, M.; Nicoli, K.; Tkatchenko, A.; Muller, K.-R. SchNetPack: A deep learning toolbox for atomistic systems. *Journal of chemical theory and computation* **2018**, *15*, 448–455.

(6) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; others Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* **2019**, *32*.

(7) Schütt, K.; Kindermans, P.-J.; Sauceda Felix, H. E.; Chmiela, S.; Tkatchenko, A.; Müller, K.-R. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in neural information processing systems* **2017**, *30*.

(8) Schütt, K.; Unke, O.; Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. International Conference on Machine Learning. 2021; pp 9377–9388.

(9) Erlebach, A.; Nachtigall, P.; Grajciar, L. Accurate large-scale simulations of siliceous zeolites by neural network potentials. *npj Computational Materials* **2022**, *8*, 174.

(10) Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* **2017**,

(11) Ying, P.; Liang, T.; Xu, K.; Zhang, J.; Xu, J.; Zhong, Z.; Fan, Z. Sub-micrometer phonon mean free paths in metal-organic frameworks revealed by machine-learning molecular dynamics simulations. *arXiv preprint arXiv:2306.02091* **2023**,

(12) Liu, J.; Byggmastar, J.; Fan, Z.; Qian, P.; Su, Y. Large-scale machine-learning molecular dynamics simulation of primary radiation damage in tungsten. *arXiv preprint arXiv:2305.08140* **2023**,

(13) Chen, Z.; Berrens, M. L.; Chan, K.-T.; Fan, Z.; Donadio, D. Thermodynamics of Water and Ice from a Fast and Scalable First-Principles Neuroevolution Potential. **2023**,

(14) Fan, Z.; Chen, W.; Vierimaa, V.; Harju, A. Efficient molecular dynamics simulations with many-body potentials on graphics processing units. *Computer Physics Communications* **2017**, *218*, 10–16.

(15) Fan, Z.; Wang, Y.; Ying, P.; Song, K.; Wang, J.; Wang, Y.; Zeng, Z.; Xu, K.; Lindgren, E.; Rahm, J. M.; others GPUMD: A package for constructing accurate machine-learned potentials and performing highly efficient atomistic simulations. *The Journal of Chemical Physics* **2022**, *157*, 114801.

(16) Dodson, B. W. Development of a many-body Tersoff-type potential for silicon. *Physical Review B* **1987**, *35*, 2795.

(17) Xu, K.; Hao, Y.; Liang, T.; Ying, P.; Xu, J.; Wu, J.; Fan, Z. Accurate prediction of heat conductivity of water by a neuroevolution potential. *The Journal of Chemical Physics* **2023**, *158*.

(18) Dong, H.; Cao, C.; Ying, P.; Fan, Z.; Qian, P.; Su, Y. Anisotropic and high thermal conductivity in monolayer quasi-hexagonal fullerene: A comparative study against bulk phase fullerene. *International Journal of Heat and Mass Transfer* **2023**, *206*, 123943.

(19) Ying, P.; Dong, H.; Liang, T.; Fan, Z.; Zhong, Z.; Zhang, J. Atomistic insights into the mechanical anisotropy and fragility of monolayer fullerene networks using quantum mechanical calculations and machine-learning molecular dynamics simulations. *Extreme Mechanics Letters* **2023**, *58*, 101929.

(20) Schaul, T.; Glasmachers, T.; Schmidhuber, J. High dimensions and heavy tails for natural evolution strategies. Proceedings of the 13th annual conference on Genetic and evolutionary computation. 2011; pp 845–852.

(21) Hoerl, A. E.; Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67.

(22) Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software* **2010**, *33*, 1.

(23) Branduardi, D.; Bussi, G.; Parrinello, M. Metadynamics with Adaptive Gaussians. *Journal of Chemical Theory and Computation* **2012**, *8*, 2247–2254.

(24) Krogh, A.; Vedelsby, J. Neural network ensembles, cross validation, and active learning. *Advances in neural information processing systems* **1994**, *7*.

(25) Bloodgood, M. Support vector machine active learning algorithms with query-by-committee versus closest-to-hyperplane selection. 2018 IEEE 12th International Conference on Semantic Computing (ICSC). 2018; pp 148–155.

(26) Schran, C.; Brezina, K.; Marsalek, O. Committee neural network potentials control generalization errors and enable active learning. *The Journal of Chemical Physics* **2020**, *153*.

(27) Gubaev, K.; Podryabinkin, E. V.; Shapeev, A. V. Machine learning of molecular properties: Locality and active learning. *The Journal of chemical physics* **2018**, *148*.

(28) Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. Less is more: Sampling chemical space with active learning. *The Journal of chemical physics* **2018**, *148*.

(29) Chen, S.; Cao, H.; Ouyang, Q.; Wu, X.; Qian, Q. ALDS: An active learning method for multi-source materials data screening and materials design. *Materials & Design* **2022**, *223*, 111092.