

# **Harnessing Physics-inspired Machine Learning to Design Nanocluster Catalysts for Dehydrogenating Liquid Organic Hydrogen Carriers**

Chuhong Lin<sup>1</sup>, Bryan C. S. Lee<sup>1</sup>, Uzma Anjum<sup>1</sup>, Asmee M. Prabhu<sup>1</sup>, Rong Xu<sup>1</sup>, and Tej S. Choksi<sup>1\*</sup>

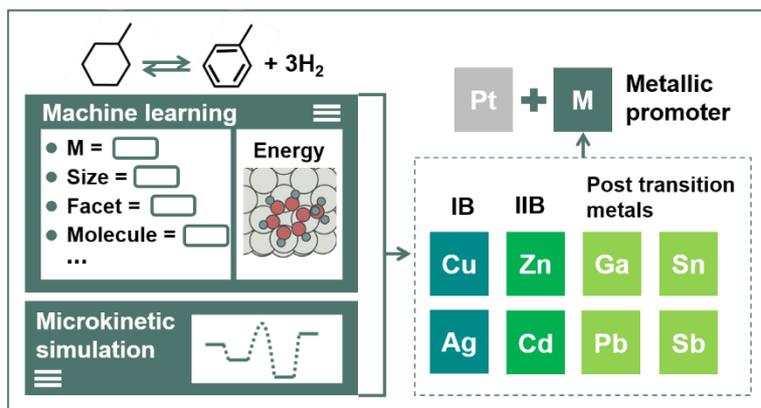
<sup>1</sup>School of Chemistry, Chemical Engineering and Biotechnology, 62 Nanyang Drive, Nanyang Technological University, Singapore, 637665, Singapore.

\*Corresponding author, email: [tej.choksi@ntu.edu.sg](mailto:tej.choksi@ntu.edu.sg)

## Abstract

Using liquid organic hydrogen carriers for the trans-oceanic shipment of hydrogen requires selective and low-cost dehydrogenation catalysts. Machine learning methods can accelerate the discovery of these catalysts. The state-of-the-art machine learning methods are however limited by challenges associated with building predictive models for large cyclic intermediates that adsorb and react on low-symmetry active sites. Focusing on methyl cyclohexane dehydrogenation to toluene, an industrially relevant hydrogen carrier, we introduce a machine learning approach to accelerate the design of selective and cost-effective catalysts. Using inputs to a gaussian process regression model that are inspired by physical theories of chemisorption, we predict the adsorption energies of large hydrocarbon intermediates that are encountered during methyl cyclohexane dehydrogenation. Across bimetallic active sites of nanoclusters having varied shapes and compositions, our model yields mean absolute errors of 0.11 – 0.25 eV on test sets and utilizes under 100 datapoints per reaction intermediate. This model is integrated with a microkinetic model to identify promising catalysts. Modifying Pt nanoclusters with IB, IIB, and post-transition elements like Cu and Sn increases dehydrogenation rates, reduces unselective reactions, and lowers Pt utilization, consistent with prior experiments. This work presents a scalable, and efficient framework for designing bimetallic catalysts for dehydrogenating hydrogen carriers.

## Table of Contents Graphic



## 1 Introduction

Liquid organic hydrogen carrier technology is central to the trans-oceanic shipment of hydrogen<sup>1,2</sup>. Discovering cost-effective catalysts that selectively dehydrogenate the hydrogen carriers while minimizing side-reactions is essential for the widespread deployment of this technology. Currently, these catalysts are discovered using trial-and-error methods that modify existing catalysts that were developed for other dehydrogenation reactions like naphtha reforming and light alkane dehydrogenation<sup>3</sup>. First principles calculations and more recently, machine learning methods have introduced accelerated paradigms for catalyst discovery<sup>4,5,6,7</sup>. It is inevitable that such computational methods will be essential in discovering catalysts for dehydrogenating hydrogen carriers that need to meet stringent stability and selectivity requirements. Yet, the computational guided discovery of catalysts has thus far been limited to reactions involving small molecules like oxygen electrolysis<sup>8</sup>, ammonia synthesis<sup>9</sup>, ammonia oxidation<sup>10</sup>, methane oxidation<sup>11</sup>, CO and CO<sub>2</sub> hydrogenation<sup>12,13</sup>, C<sub>2</sub>-C<sub>3</sub> alkane dehydrogenation<sup>14,15,16,17</sup>, and acetylene semi-hydrogenation<sup>18</sup>. Establishing similar predictive models for catalytic reactions involving larger molecules as seen in dehydrogenating hydrogen carriers is not feasible using the current state-of-the-art because of four limitations. First, larger reaction intermediates like unsaturated cyclic hydrocarbons adsorb across multiple atoms of an active site, with the active sites in turn, possessing low symmetry. Second, the diverse structural (e.g., coordination numbers) and chemical environments (e.g., composition of first nearest neighbors) further complicate accurate predictions of adsorption energies using machine learning models that have been otherwise effective for small molecules<sup>19</sup>. Third, the high computational cost of building datasets involving such larger reaction intermediates like cyclic hydrocarbons precludes the use of data-heavy approaches like deep learning<sup>20</sup>. Fourth, the reaction kinetics cannot always be represented by one or two rate determining steps as is done for reactions involving small molecules like N<sub>2</sub> and O<sub>2</sub>. Instead, reaction kinetics can only be calculated after knowing the energetics of all reaction intermediates. To overcome these four challenges, a new machine learning paradigm is required that employs minimalistic training datasets, is transferable across different active site motifs in structurally and chemically complex systems and is efficient to deploy across large cyclic hydrocarbons.

In response to these challenges, we introduce a machine learning approach that guides design of catalysts used for the dehydrogenation of methyl cyclohexane to toluene. This particular liquid organic hydrogen carrier is selected because it has been demonstrated in pilot plant and more or less ready to be scaled up by industry<sup>21,22</sup>. The endothermic dehydrogenation reaction necessitates high temperatures, typically above 500 K, and requires a catalyst<sup>23</sup>. While Pt is a promising catalyst for methyl cyclohexane dehydrogenation, its high cost and low selectivity to toluene hinder the wide-spread applications<sup>24</sup>. Too-strong adsorption of toluene leads to undesirable side products like benzene, bibenzyl-toluene, and coke, diminishing the efficacy of this technology<sup>25</sup>. The quest of discovering tailored catalysts that maintain or enhance reactivity and selectivity while minimizing the Pt loading remains a pressing concern.

One promising strategy to address this challenge is to use bimetallic nanoclusters (Pt-M) where Pt is combined with earth-abundant elements denoted as M. The introduction of element M modulates the electronic environment of the active sites and can improve catalytic performance while lowering the Pt content. Prior experimental studies have reported that using Pt-Sn, Pt-Pb, and Pt-Zn bimetallic catalysts improves the rate and selectivity of methyl cyclohexane dehydrogenation reaction<sup>26,27</sup>, suggesting that alloying with late transition metals and metalloids is a viable pathway for discovering new catalysts.

The Pt-M alloys span a wide-range of morphologies and compositions that complicates the selection of metal M. Trial-and-error based experiments are both time- and resource-intensive. To streamline the catalyst design efforts for methyl cyclohexane dehydrogenation, we introduce an innovative machine learning framework. This framework capably predicts the adsorption energies of unsaturated diverse C<sub>7</sub> hydrocarbons that are encountered as reaction intermediates. These adsorption energies are inserted into a microkinetic model. The machine learning model assesses the reaction energetics and rates for methyl cyclohexane dehydrogenation over bimetallic active sites exhibiting diverse coordination environments and compositions. Using physics-inspired model inputs both reduce the computational cost of training the model while yielding seamless transferability across different nanoparticle shapes and compositions beyond the training dataset. Our machine learning framework identifies that elements with a full d-band can increase reactivity and prevent overly strong adsorption of toluene thus reducing unwanted byproducts such as benzene. This paradigm can not only be used for designing tailored catalysts for dehydrogenating hydrogen carriers but has the potential to streamline similar catalyst design efforts for reactions involving large organic molecules as encountered during biofuel production and plastic depolymerization.

## 2 Methods

### 2.1 Density functional theory (DFT) calculations

Density Functional Theory calculations were conducted using the Quantum ESPRESSO package<sup>28</sup> within the Atomic Simulation Environment (ASE) environment<sup>29</sup>. Total energies were self-consistently calculated using the Perdew–Burke–Ernzerhof exchange-correlation functional<sup>30</sup>. Long-range dispersion corrections were accounted for using Grimme’s D3 approach<sup>31</sup>. Core states were simulated with Vanderbilt ultra-soft pseudopotentials<sup>32</sup>. Kinetic energy cutoffs of 500 eV for plane waves and 5000 eV for electron densities were applied. All calculations were spin-paired. Electronic states were smeared by 0.1 eV using the Fermi-Dirac distribution, to enable convergence of the total energies. Nanoparticles (cubo-octahedral, decahedral, and octahedral) and gas-phase molecules were simulated in supercells having sizes such that periodic images were separated by at least 9 Å of vacuum along the x, y, and z directions. Total energies were converged to a numerical tolerance of 10<sup>-5</sup> eV. Forces were minimized

during geometry optimization until a threshold of 0.05 eV Å<sup>-1</sup>. All atoms of the adsorbates and the nanoparticles were relaxed during this geometry optimization. Transition state energies were calculated using the machine-learning enhanced nudged elastic band (ML-NEB) method introduced by Garrido Torres et al.<sup>33</sup> These transition states were confirmed using a vibrational frequency analysis by verifying the presence of an imaginary mode along the reaction coordinate. The vibrational frequencies were calculated through numerical differentiation of the forces using a second-order finite difference approach with a step size of 0.015 Å. For these vibrational frequency calculations, the metal atoms of the nanoparticle were fixed in their optimized geometries, with only the adsorbate atoms being relaxed. Zero-point energies (ZPE) were determined using these vibrational frequencies. Entropies of gas molecules were calculated using the ideal gas assumption. The vibrational entropies of adsorbates were computed under the harmonic approximation. Vibrational frequencies less than 12 cm<sup>-1</sup> and imaginary modes that were not along the reaction coordinate, were replaced by 12 cm<sup>-1</sup> based on prior literature study<sup>34</sup>. Input and output files the DFT calculations are available on GitHub via [https://github.com/CLinSim/ML\\_LOHC](https://github.com/CLinSim/ML_LOHC). Images of atomic structures were generated with scripts that are also available on the same GitHub website. Additional details are provided in the Supporting Information.

## **2.2 Gaussian process regression (GPR) model to estimate adsorption energies of reaction intermediates of methyl cyclohexane dehydrogenation**

We first constructed a dataset comprised of adsorption energies of methyl cyclohexane dehydrogenation intermediates that were adsorbed on different Pt-M nanoclusters. Individual GPR models were trained for each C<sub>7</sub> reaction intermediate. The features employed reflect properties inferred from physics-based models and are discussed in section 3.1. The kernel function and hyperparameters of the GPR model were trained using the fitrgp function in MATLAB. Feature inputs for the GPR model are computed using the composition and the coordination environment of an active site and its first nearest-neighbors. MATLAB scripts employed for calculating the features, performing training, validating, and testing of the GPR model, establishing the active learning workflow, and for visualizing the results are available on Github via [https://github.com/CLinSim/ML\\_LOHC](https://github.com/CLinSim/ML_LOHC).

## **2.3 Microkinetic modelling**

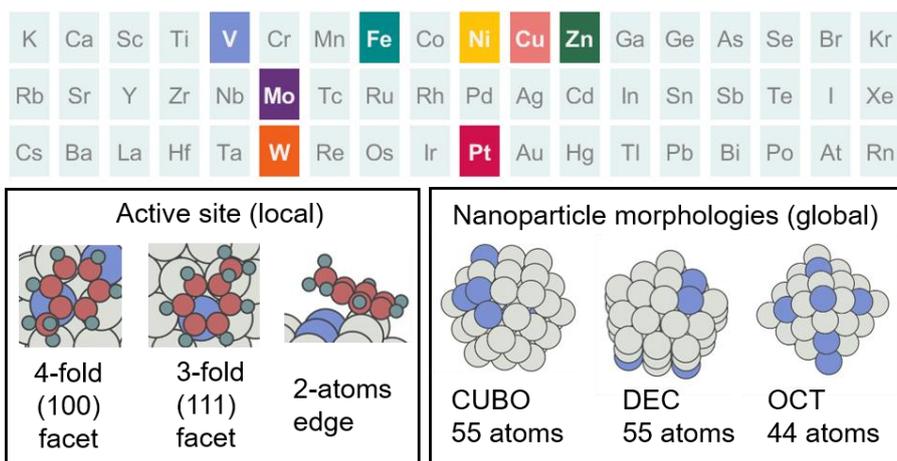
We constructed a dual site microkinetic model to assess the impact of varying the composition of bimetallic nanoparticles on the rates of methyl cyclohexane dehydrogenation. C<sub>7</sub> hydrocarbons and atomic hydrogen was assumed to adsorb on different active sites. The microkinetic model includes elementary steps for methyl cyclohexane adsorption, six successive C-H dissociations, toluene desorption, and H-H association. Reaction energies of elementary steps on a given active site of a bimetallic nanoparticle were determined using the GPR model that is discussed in section 2.2. Intrinsic activation barriers were computed using a Brønsted–Evans–Polanyi (BEP) relationship; such linear relationships are broadly

established for C-H activation<sup>35</sup>. The BEP relationship was determined using intrinsic activation barriers that were computed on Pt<sub>55</sub> nanoparticles using DFT. The reaction network was simulated in MATLAB under a fixed conversation of 5% with 95 kPa methyl cyclohexane, 5 kPa toluene and 15 kPa hydrogen at a temperature of 600 K. The pseudo steady state hypothesis was employed for adsorbed species. Where possible, numerical simulations were confirmed using an analytical expression for the reaction rates. Further details of the microkinetic model are found in the Supporting Information.

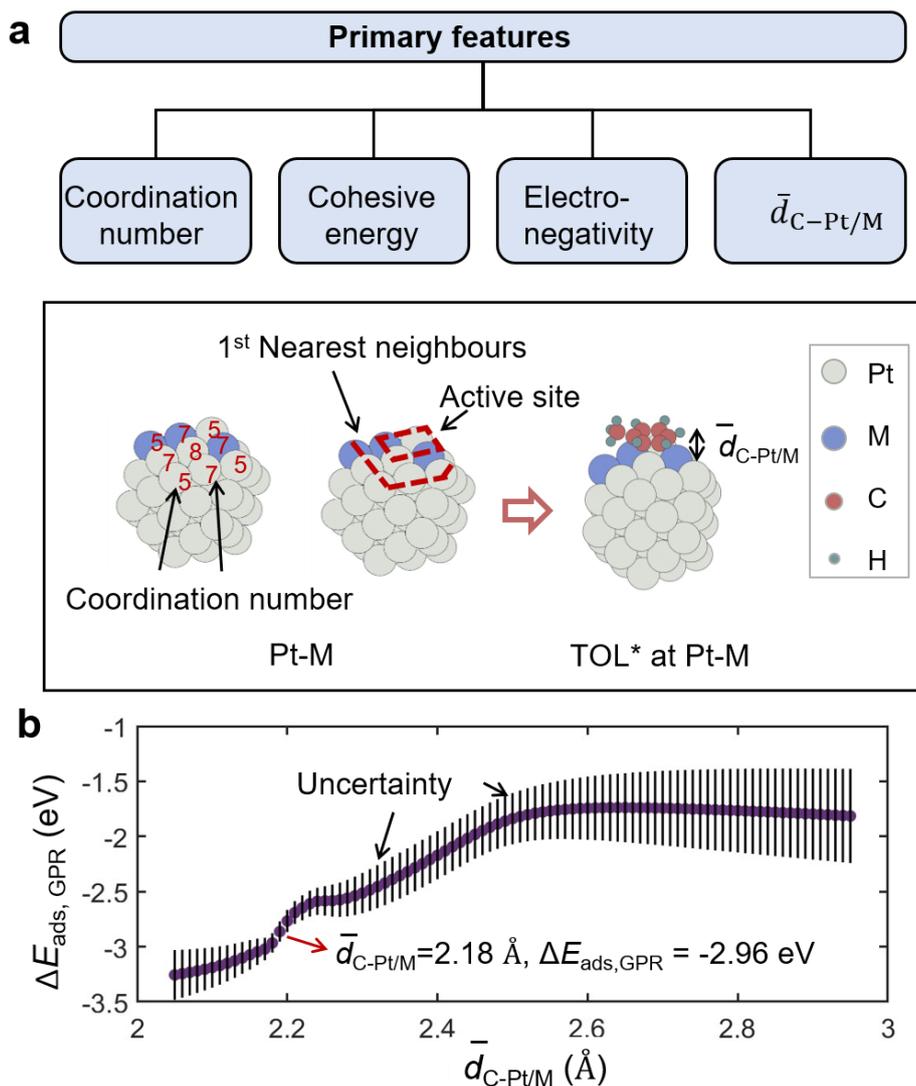
### 3 Results

#### 3.1 Gaussian process regression (GPR) model for predicting the adsorption energy of toluene

The GPR model was first built for predicting the adsorption energy of toluene ( $\Delta E_{\text{ads,TOL}}$ ) across a diverse range of Pt-M bimetallic nanoclusters. This choice is guided by the importance of adsorbed toluene in dictating the selectivity of methyl cyclohexane dehydrogenation. The strong adsorption of toluene increases the propensity of deleterious C-C scission that forms benzene<sup>25</sup>. Figure 1 shows the compositional and configurational space from which bimetallic clusters are selected to train the GPR model for predicting  $\Delta E_{\text{ads,TOL}}$ . This training set encompasses variations in the doping element M, the local coordination- and compositional-environment of the active site and its first nearest neighbors, and the morphologies of nanoclusters. The active site is characterized by the metal atoms that are proximal to the unsaturated carbon atoms of the adsorbate. Tuning the composition of the active site and its first nearest neighbors has the highest impact on the chemisorption energies of bimetallic catalysts<sup>36</sup>. Changes in the composition of the second nearest neighbors and beyond have a progressively diminished impact on chemisorption energies. This is because of the well-established screening effects shown by d-electrons<sup>37,38</sup>. Hence, we restrict our focus to tuning the composition of elements that constitute the active site and its first nearest neighbors.



**Figure 1.** The range of Pt-M bimetallic nanoparticles (where M is the doping element) that form the dataset used to train the GPR model for predicting the adsorption energy of toluene. The training dataset spans different doping metals (colored elements are the doping elements used as M), different local active site configurations on (100) and (111) facets of the nanoparticle, and nanoparticle morphologies exhibiting varied shapes and sizes.



**Figure 2** (a) The primary features used in the GPR model are the coordination number, cohesive energy, electronegativity, and  $\bar{d}_{C-Pt/M}$ , the average bond length between all the unsaturated carbons with their nearest metal atoms. (b) One example of calculating the adsorption energy of toluene using the trained GPR model ( $\Delta E_{ads, GPR}$ ) under a series of  $\bar{d}_{C-Pt/M}$  from 2.0 to 3.0 Å. The Pt-M nanocluster, where M is V that is used in this example is illustrated in (a). The value of  $\Delta E_{ads, GPR}$  is shown by the purple line and the uncertainty at each  $\bar{d}_{C-Pt/M}$  is plotted using the vertical black lines. The corresponding  $\Delta E_{ads, GPR}$  with the least uncertainty, indicated by the red arrow, is considered as the predicted adsorption energy.

Intermediates like toluene adsorb across two to four metals atoms of the active sites in multi-dentate adsorption modes. These metal atoms span diverse coordination numbers and compositions, resulting in heteroatomic active sites. A single descriptor for chemisorption will not work for such structurally and compositionally complex heteroatomic active sites. Moreover, linear models like scaling relationships lose their

accuracy on heteroatomic active sites<sup>39</sup>. To overcome the challenges associated with multi-dentate adsorption at such complex active sites, we employ Gaussian Process Regression (GPR). The features of the GPR model are selected based on insights from physics-based models of chemisorption. These features include the bulk cohesive energy of metal atoms ( $E_{\text{coh}}$ ) and the electronegativity according to the Pauling scale ( $\chi$ ), as presented in Figure 2a.  $E_{\text{coh}}$  represents the stability of elements that constitute the active site. The choice of  $E_{\text{coh}}$  is inspired by the inherent relationship between stability and reactivity as shown in computational models like the “alpha-scheme”<sup>40,41</sup> and as inferred from calorimetry measurements<sup>42</sup>.  $\chi$  is selected as an electronic descriptor for metal atoms<sup>43</sup> and is an important feature in other physics-based and machine learning models of chemisorption of small adsorbates<sup>44,45</sup>. The strength of the chemical bond between the adsorbate and the active site, an essential aspect of the molecule-metal interaction<sup>45</sup>, is characterized by the average carbon-metal bond length, denoted as  $\bar{d}_{\text{C-Pt/M}}$  and included as a primary feature. Recognizing that adsorption energies can be structure-sensitive<sup>46</sup>, we calculated these atomic properties weighted by the coordination environment of the active site. A three-dimensional fingerprint [ $F_1$ ,  $F_2$ ,  $F_3$ ] was established for each Pt-M active site.  $F_1$ ,  $F_2$ , and  $F_3$  are defined by the following equations:

$$F_1 = \frac{1}{n_{\text{as}}E_{\text{coh,bulk}}} \sum_i^{\text{as}} \left( E_{\text{coh},i} \times \sum_j^{\text{1NN}} E_{\text{coh},j} \right) \quad (1)$$

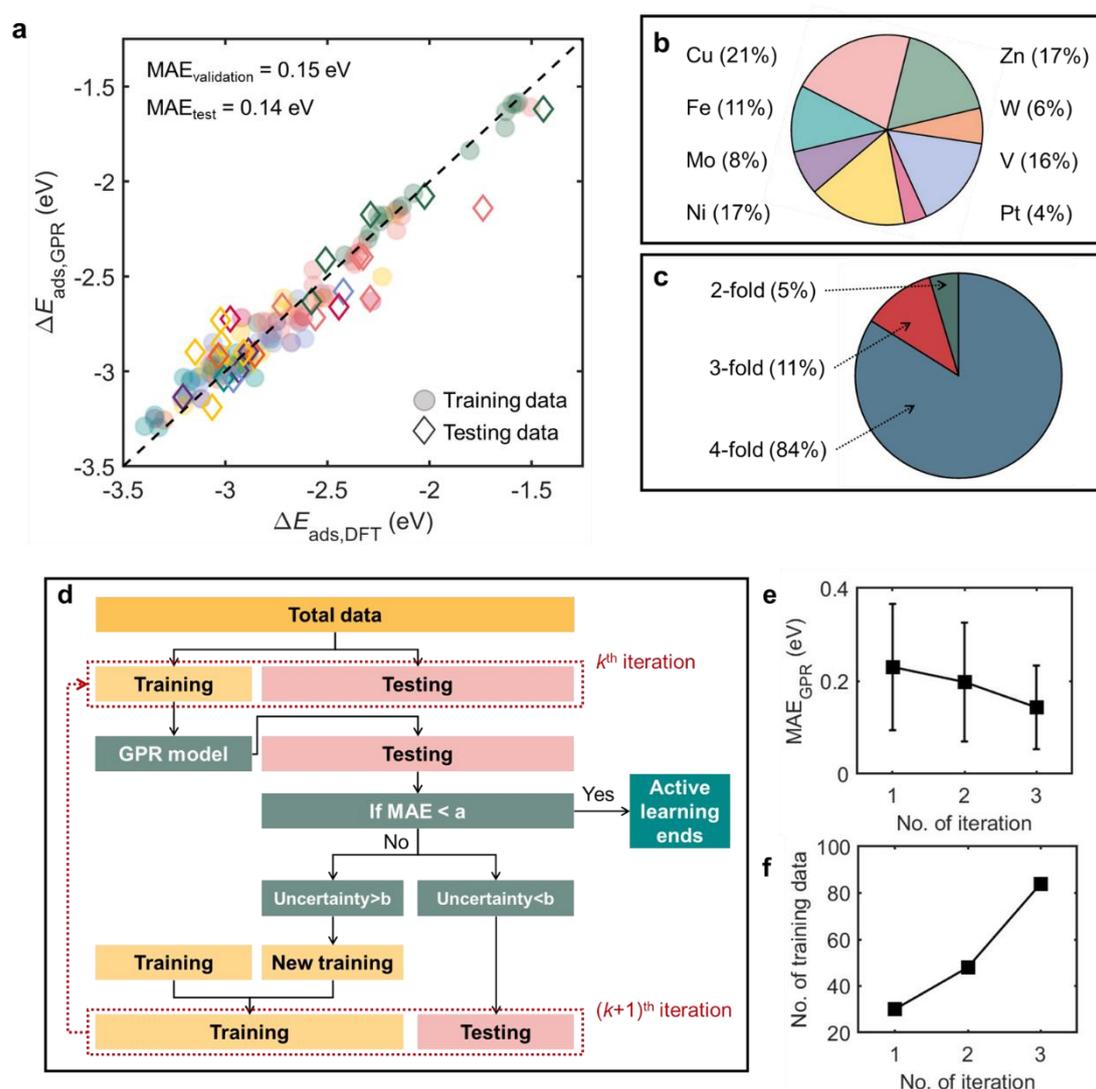
$$F_2 = \frac{1}{n_{\text{as}}\chi_{\text{bulk}}} \sum_i^{\text{as}} \left( \chi_i \times \sum_j^{\text{1NN}} \chi_j \right) \quad (2)$$

$$F_3 = \frac{\sum_i^{\text{as}} d_{\text{C-Pt/M}}}{n_{\text{as}}} \quad (3)$$

In Eqn. 1-3, “as” refers to the atoms belonging to the active site, “1NN” refers to the first nearest neighbor metal atoms,  $n_{\text{as}}$  is the total number of active-site atoms, and  $n_{\text{1NN}}$  is the total number of 1NN atoms. A detailed example of calculating these features is shown in Figure S3 of the supporting information. The values of  $E_{\text{coh}}$  and  $\chi$  are obtained from established databases<sup>47,48</sup>.  $E_{\text{coh,bulk}}$  and  $\chi_{\text{bulk}}$  are the cohesive energy and electronegativity values of the bulk element in the bimetallic nanocluster. The product of the features of atoms in the active site and the first nearest neighbors in Eqn. 1 and 2 represent the pair-wise interactions between atoms in the active site and the first nearest neighbors.

For training the model, we applied GPR with exponential kernel functions, using the fingerprint [ $F_1, F_2, F_3$ ] extracted from DFT-optimized geometries over a dataset consisting of 131  $\Delta E_{\text{ads,TOL}}$ . The only feature that requires DFT-optimized geometries in the fingerprint is,  $\bar{d}_{\text{C-Pt/M}}$ . However, when we apply the GPR model to a new active site outside the training dataset,  $\bar{d}_{\text{C-Pt/M}}$  is not known a priori. Thus,  $\bar{d}_{\text{C-Pt/M}}$  is treated as a variable in such situations. We harness the intrinsic uncertainty estimation capabilities of GPR to find the specific value of  $\bar{d}_{\text{C-Pt/M}}$  that minimizes the uncertainty

in predicting  $\Delta E_{\text{ads,TOL}}$ . Thus, no DFT-derived inputs are needed when predicting adsorption energies using our GPR model. Figure 2b illustrates the application of the GPR model for predicting  $\Delta E_{\text{ads,TOL}}$  on a Pt<sub>52</sub>-Mo<sub>3</sub> nanocluster, wherein two Pt atoms in the active site and one Pt atom in the nearest neighbor of the active site are replaced by Mo atoms. We computed both  $\Delta E_{\text{ads,TOL}}$  and its uncertainty across  $\bar{d}_{\text{C-Pt/M}}$  ranging from 2.1 to 2.9 Å. The  $\Delta E_{\text{ads,TOL}}$  corresponding to the smallest uncertainty was determined as the predicted adsorption energy. For the Pt<sub>52</sub>-Mo<sub>3</sub> nanocluster, the averaged bond length and the adsorption energy were predicted to be 2.18 Å and -2.96 eV, closely matching the DFT-calculated results of 2.16 Å and -2.89 eV respectively.



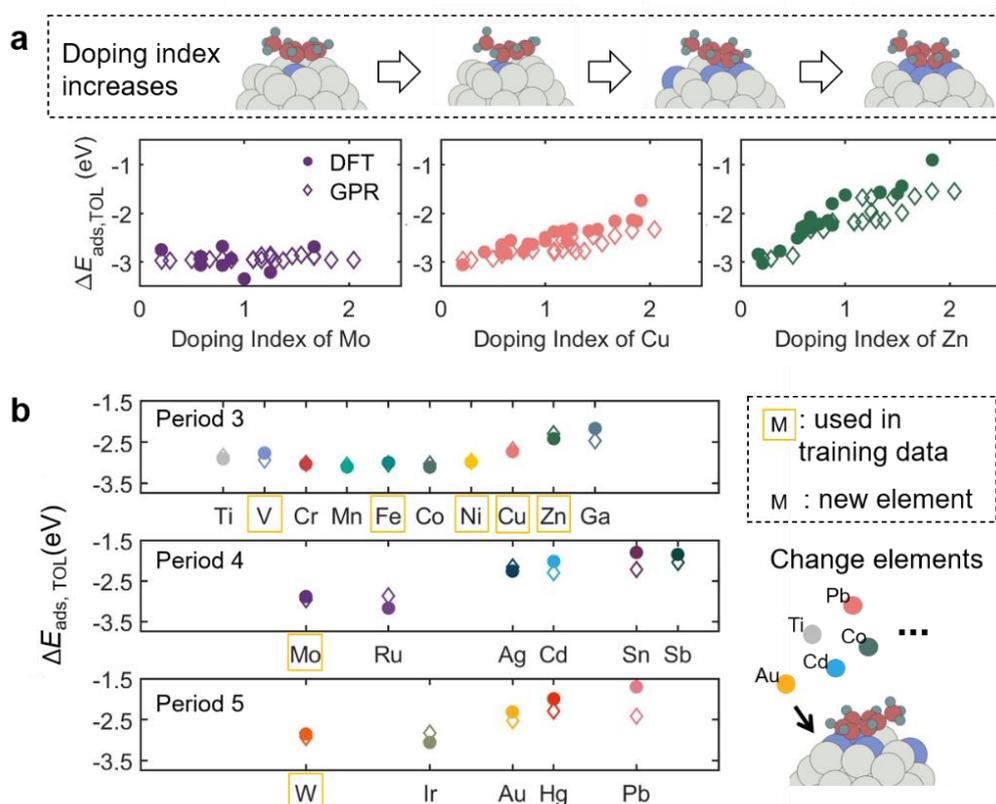
**Figure 3** (a) Parity plot of the GPR model for predicting the adsorption energy of toluene. The 131 data points were randomly divided into training, validation, and testing datasets. The mean absolute error across these datasets is stated. (b) and (c) show the distribution in the elemental composition and the distribution of the active sites for adsorbed toluene. (d) Workflow of the active learning simulator. 30 structures

were selected randomly from the 131 data points to train a GPR model. The GPR model was used to predict the  $\Delta E_{\text{ads}}$  for the remaining data points. The data points exhibiting an uncertainty  $> b$  ( $b$  is 0.15 eV) were calculated by DFT and added to the training set. The training and testing process was iteratively continued until the MAE of the testing data was smaller than  $a$  ( $a$  is 0.15 eV). (e) MAE on the test set as a function of iterations of the active learning simulator. (f) Number of training data points as a function of iterations of the active learning simulator.

Figure 3a presents a parity plot of the GPR model that is constructed using a dataset of 131  $\Delta E_{\text{ads,TOL}}$  that correspond to various Pt-M configurations shown in Figure 1. We divided the total dataset into training and testing subsets using an 80-20 split. Model training ensued using a five-fold cross-validation approach on the training set. Further details about the training, validation, and testing process are stated in the Supporting Information. To rigorously assess the model's accuracy, we computed the mean absolute error (MAE) not from a single division of the 131  $\Delta E_{\text{ads,TOL}}$  but over 100 random splits, as shown in the trends in Figure S5. Averaged over 100 iterations, the MAE on the training dataset, validation dataset, and testing dataset are 0.06 eV, 0.15 eV, and 0.14 eV, respectively. The resultant validation and testing MAE values are consistently 0.15 eV or lower, which is comparable to, or surpasses, the precision of existing models for smaller adsorbates on nanoclusters<sup>49,50,51</sup>. The similarity of MAEs between the validation and testing data confirms the absence of overfitting. The distributions of the doping element M and the active site geometries within the dataset of 131 points are depicted in Figures 3b and 3c. Due to the selection of physics-inspired features, the GPR model makes accurate prediction across diverse doping elements and different active site geometries.

While the dataset of 131 points is already modest when compared to other neural network based machine learning models<sup>52</sup>, we demonstrated the potential to further condense the training set size via active learning. We clarify that this technique was not utilized while assembling the dataset for  $\Delta E_{\text{ads,TOL}}$  but it is anticipated to be beneficial for retraining the model for other cyclic hydrocarbons in future studies. Furthermore, this active learning simulation determines the minimal dataset size required for building a GPR model that predicts  $\Delta E_{\text{ads,TOL}}$  within a given precision. We randomly selected 30 structures from the 131 points as an initial training dataset, with the remainder serving as testing dataset. Testing data with uncertainties exceeding 0.15 eV were computed via DFT and incorporated into the training pool. This iterative process continued until all testing data achieved an MAE below 0.15 eV. The active learning simulator's workflow is outlined in Figure 3d, and the iterative reduction of MAE against the training dataset size is plotted in Figure 3e. Remarkably, after three iterations, we attained a testing MAE under 0.15 eV using a training set of only 98 data points for  $\Delta E_{\text{ads,TOL}}$ . The result proves that accurate predictions for intricate multidentate adsorptions on heteroatomic active sites of bimetallic nanoclusters are feasible with fewer than 200 DFT calculations (98 calculations on the clean nanocluster and an additional 98 calculations for toluene adsorbed on the nanocluster).

### 3.2 Transferability of the GPR model to new active sites, elements, and adsorbates



**Figure 4** (a) Dependence of the adsorption energy of toluene on the doping index for 55-atom Pt-Mo, Pt-Cu and Pt-Zn nanoparticles having CUBO morphologies. This doping index represents the concentration of M in the active site and its 1<sup>st</sup> nearest neighbors. (b) The adsorption energy of toluene is predicted on active sites containing doping elements outside the training set. The structure of the nanocluster is shown on the right. In both (a) and (b), solid circles were adsorption energies calculated by DFT and open diamonds were adsorption energies predicted by GPR. A parity plot comparing the model against DFT calculated adsorption energies is shown in the Supporting Information (Figure S6).

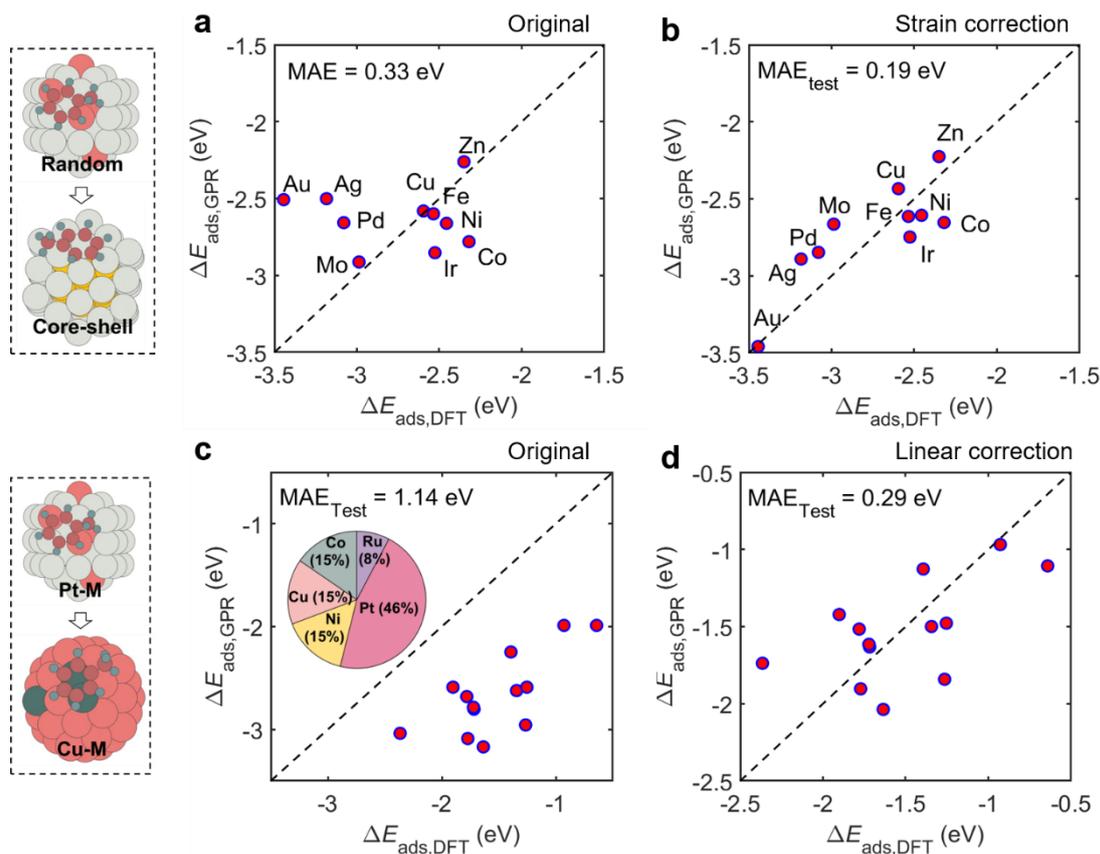
Given the small MAE of the GPR model, we now test the veracity with which this model determines directional trends in  $\Delta E_{\text{ads, TOL}}$  as the identity and concentration of doping elements is varied. To quantitatively encapsulate both the concentration and spatial distribution of M atoms relative to the active site, we introduced a metric termed as the doping index in Eqn. (4).

$$\text{Doping Index} = \sum_i^{M \text{ atoms}} w_i \times \left( \frac{CN_i}{12} \right) \quad (4)$$

In Eqn. (4),  $w_i$  is the coefficient indicating the location of M, defined as 0.5 if M is a first nearest neighbor and 1 if M is part of the active site. An illustration of calculating the doping index for a Pt<sub>52</sub>Cu<sub>3</sub> nanocluster is provided in the Supplementary Information.

Employing a variety of Pt-M structures not previously used in the training dataset, we generated structural models of nanoclusters with differing doping indices and estimated their  $\Delta E_{\text{ads,TOL}}$  using the GPR model. Figure 4a illustrates the sensitivity of  $\Delta E_{\text{ads,TOL}}$  to the modification of the local environment around the active site with three different elements Mo, Cu and Zn. Mixing Pt with Mo exhibited negligible effects on  $\Delta E_{\text{ads,TOL}}$ ; increasing Cu concentrations proportionally weakened  $\Delta E_{\text{ads,TOL}}$ ; while minor amounts of Zn were sufficient to significantly weaken  $\Delta E_{\text{ads,TOL}}$ . With increasing doping index, the trends in the  $\Delta E_{\text{ads,TOL}}$  predicted by the GPR model are consistently aligned with DFT calculated adsorption energies. This directional alignment validates the capability of the GPR model in discerning the impact on  $\Delta E_{\text{ads,TOL}}$  through subtle changes in composition of the active site and its first nearest neighbors. For higher doping index values of Zn, a slight deviation in predicted  $\Delta E_{\text{ads,TOL}}$  relative to DFT results is observed. This deviation is attributed to the restructuring of the Pt-Zn nanocluster at high Zn concentrations. Such restructuring in nanoclusters after adsorption is not accounted in our GPR model, but could potentially be addressed through machine learned force fields<sup>53</sup>.

We conducted further tests of the GPR model on Pt-M alloys where M is taken from the 3<sup>rd</sup> to 5<sup>th</sup> periods of the periodic table, as depicted in Figure 4b. The close match between the GPR-predicted  $\Delta E_{\text{ads,TOL}}$  against DFT calculated values validate our approach. Interestingly, this model is extrapolatable to elements outside the training set. Pt-Pb however, emerges as an exception. Pb provides a sufficiently strong repulsion towards toluene that the interaction between the Pt-Pb site and toluene is almost physisorption. Since the GPR model is trained on structures with chemisorption, the higher errors for physisorption are expected. Screening across various metals, we find that the elements with a full d-band, such as Cu, Zn, Sn and Pb, weaken the adsorption energy of toluene, thus enabling swift desorption of the dehydrogenated product and preventing side-reactions to benzene. Hence such elements can be considered as promoters for methyl cyclohexane dehydrogenation.



**Figure 5** (a, b) Transferability of the GPR model to 55-atom core-shell nanoclusters with Pt in the shell. The core elements are marked at each point. The adsorption energies of toluene in (a) are predicted using the GPR model presented in Figure 3. In (b), the fingerprints are corrected for the effects of strain resulting in an improved agreement between the GPR-predicted and DFT-derived adsorption energies. (c, d) Transferability of the GPR model to Cu-rich bimetallic nanoclusters (Cu-M). The distribution of M is plotted in the pie chart. The toluene adsorption energies in (c) are all predicted using the same GPR model presented in Figure 3. In (d), an additional linear correction is applied to each predicted adsorption energy, where the coefficients ( $p_1$ ,  $p_2$ ) in the linear fitting  $y = p_1 * x + p_2$  are obtained using a leave-one-out method. The MAE in (d) refers to the leave-one-out error.

Figure 4 confirms that the GPR model can competently describe how changes in composition around a single active site influences  $\Delta E_{\text{ads,TOL}}$ . We now test the validity of the model against changes in composition at the level of an entire nanocluster. Figure 5a shows that the GPR model trained on randomly doped Pt-M nanoclusters yields significantly higher error on core-shell Pt-M nanoclusters (Pt as the shell and different M as the core) resulting in a MAE of 0.33 eV. While the GPR model can correctly simulate the ligand effects on adsorption (Figures 3 and 4), this model does not include the effects of strain on adsorption energies. Given the mismatched lattice constants of elements in the core and the shell<sup>54</sup>, such effects induced by strain need explicit consideration. We observed a positive correlation between the prediction error

and the geometric area of the active site, as shown in Figure S7. This correlation indicates that the high error is because the effects of strain are ignored. We therefore adjusted the fingerprints by incorporating the influence of geometric area of the active site as shown in Eqn. (5). The geometric area of the active sites of core-shell nanoclusters and Pt<sub>55</sub> are derived using their lattice constants and are listed in the Supporting Information. Figure 5b shows that after adjusting the fingerprints, MAE reduces to 0.17 eV, confirming the validity of our model to core-shell structures after correcting for strain effects.

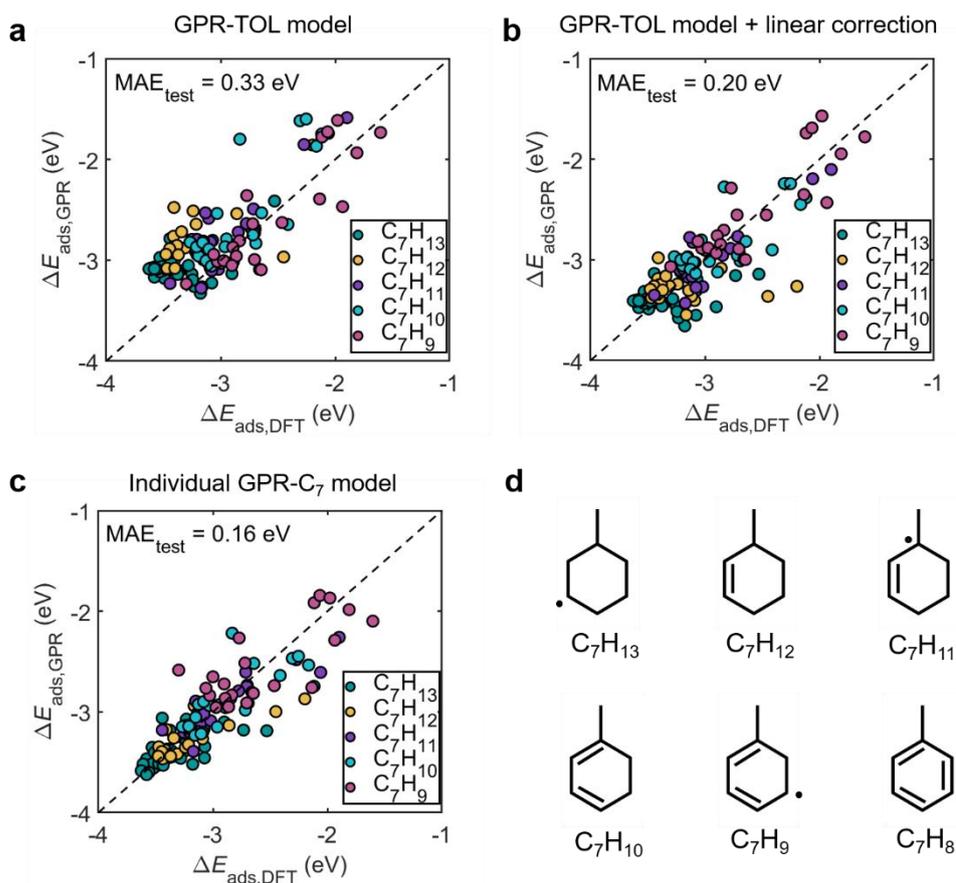
$$\Delta E_{\text{ads, strain corrected}} = \Delta E_{\text{ads, original}} \left( \frac{\text{Area}_{\text{Pt55}}}{\text{Area}_{\text{core-shell}}} \right)^2 \quad (5)$$

The transferability of the GPR model trained on Pt-rich nanoclusters to Cu-rich Cu-M nanoclusters is tested in Figure 5c. We calculated  $\Delta E_{\text{ads, TOL}}$  on 13 Cu-M structures, including different morphologies and compositions (monometallic Cu, Cu-Pt, Cu-Co, Cu-Ru and Cu-Ni). Although the MAE is high in Figure 5c, this error is systemic and the GPR-predicted  $\Delta E_{\text{ads, TOL}}$  is highly correlated with the DFT calculated  $\Delta E_{\text{ads, TOL}}$  for the considered structures. Therefore, a linear function is used to correct the predicted adsorption energy as shown in Eqn. (6).

$$\Delta E'_{\text{ads, GPR}} = p_1 \Delta E_{\text{ads, GPR}} + p_2 \quad (6)$$

In Eqn. (6),  $\Delta E_{\text{ads, GPR}}$  is the adsorption energy predicted by the GPR model and  $\Delta E'_{\text{ads, GPR}}$  is the result after linear fitting. The parameters  $p_1$  and  $p_2$  are fitted to the DFT calculated Cu-M data. In Figure 5d, we use the leave-one-out approach to validate the linear fitting. Each time one of the 13 data points was used as the testing datapoint, and  $p_1$  and  $p_2$  are trained to the remaining 12 training data. The MAE in Figure 5d is averaged over the total 13 tests. After a simple linear correction, the MAE decreases from 1.04 eV to 0.29 eV, affirming the utility of such linear corrections to extend the validity of the GPR model to bimetallic compositions that go beyond the training set.

Figure S8 shows the histograms of  $F_1$  and  $F_2$  in Cu-rich and Pt-rich datasets. The difference in  $F_1$  and  $F_2$  between the two datasets explains the increase of MAE when transferring the GPR model trained on Pt-rich nanoclusters to Cu-rich. Pt and Cu present distinct cohesive energies (Pt: 5.30 eV/atom, Cu: 3.49 eV/atom) and electronegativities (Pt: 2.28, Cu: 1.90). The GPR model is expected to perform even better on alloys with similar cohesive energies and electronegativities as Pt.



**Figure 6** Transferability of the GPR model that is trained to the adsorption energies of toluene, to predict the adsorption energies of other  $C_7$  intermediates. (a) Adsorption energies of  $C_7$  species are predicted using the GPR model trained to the adsorption energies of toluene (GPR-TOL model); (b) Adsorption energies of  $C_7$  species are predicted using a linear correction to outputs from the GPR-TOL model; (c) Adsorption energies of  $C_7$  species are predicted using individual GPR models that are trained separately to each  $C_7$  species. For (b) and (c), the training-testing dataset was split using the leave-one-out method. The testing MAE stated is calculated by averaging the absolute error over all possible leave-one-out combinations in the data set. (d) The structures of different  $C_7$  intermediates considered.

Thus far, a GPR model trained on physics-inspired fingerprints predicts adsorption energies of large adsorbates like toluene with accuracies on-par with other machine learning models built for simpler adsorbates<sup>6, 10, 51</sup>. Extensive testing of this model proved its transferability across different active site structures and diverse nanocluster compositions. We finally extend this GPR model trained on  $\Delta E_{\text{ads,TOL}}$  to predict the adsorption energies of other  $C_7$  reaction intermediates seen during methyl cyclohexane dehydrogenation. These intermediates range from  $C_7H_{13}$  to  $C_7H_9$ . In Figure 6, the adsorption energy of reaction intermediate  $C_7H_x$  was referred to the energy of gas phase toluene and  $(x-8)/2 H_2$  as defined in Eqn.(S1b). The MAE is found to be 0.33 eV averaging over all the  $C_7H_x$  adsorbates as shown in Figure 6a. The

relatively large MAE is caused due to the structural difference between toluene and other  $C_7H_x$  adsorbates. It is reasonable that the transferability of hyperparameters trained for one adsorbate to other adsorbates is determined by the structural similarity between the two adsorbates, i.e. sharing the same functional group or adsorbing in a similar configuration. For example, the GPR model trained on toluene, can however be directly used to predict the adsorption energy of benzene (see Figure S9) with a MAE of 0.18 eV. Thus, the GPR model can be applied to molecules of similar structures without further training of the hyperparameters. However, the model is less accurate if the adsorption configuration of intermediates (e.g.,  $C_7H_{12}$  and  $C_7H_{13}$ ) changes. To reduce the error, we used a linear regression to correlate the model-predicted adsorption energies (GPR-TOL) with DFT-derived values for each reaction intermediate. This approach resembles what was used in Figure 5 for extending a model trained on Pt-rich nanoclusters to Cu-rich nanoclusters. The predicted adsorption energies after this linear regression are plotted in Figure 6b with a reduced MAE of 0.20 eV. The MAE after the linear regression ( $MAE_{\text{linear}}$ ) for each intermediate is listed in Table 1.

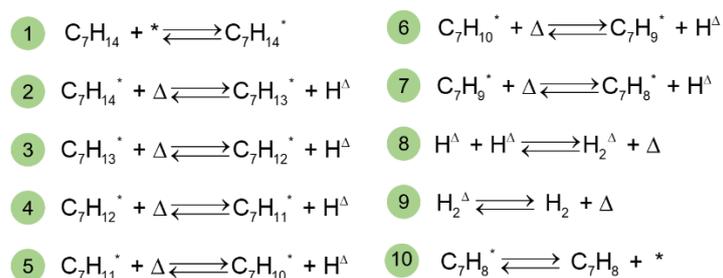
**Table 1** The total number of adsorption energies of each reaction intermediate on bimetallic Pt nanoclusters that are calculated using DFT. The MAEs for different models are also stated.

	$C_7H_{13}$	$C_7H_{12}$	$C_7H_{11}$	$C_7H_{10}$	$C_7H_9$
No. of data	48	27	26	26	24
$MAE_{\text{linear}}$ (eV) (GPR-TOL model + linear correction)	0.18	0.24	0.16	0.19	0.25
$MAE_{\text{GPR, validation}}$ (eV) (Individual GPR- $C_7$ model)	0.18	0.19	0.13	0.26	0.27
$MAE_{\text{GPR, test}}$ (eV) (Individual GPR- $C_7$ model)	0.12	0.13	0.12	0.19	0.29

In addition to such linear regressions, training individual GPR models for each  $C_7$  intermediate can also be built to increase the predictive accuracy. Using individual GPR models, the mean MAE across all the  $C_7$  adsorbates decreases from 0.33 eV to 0.16 eV, as shown in Figure 6c. Further details about the accuracy of the individual GPR models are provided in Table 1. For  $C_7$  adsorbates other than toluene, we only calculated 25-40 adsorption energies. Hence the MAEs between the individually trained GPR models and the linear regression appear to be similar. But the prediction accuracy of the GPR models can be further improved with larger datasets. The active learning procedure developed in Figure 3b can help to accelerate the training of such models. Figure 6c, together with Figure 5b and 5d, confirm that our physics-inspired

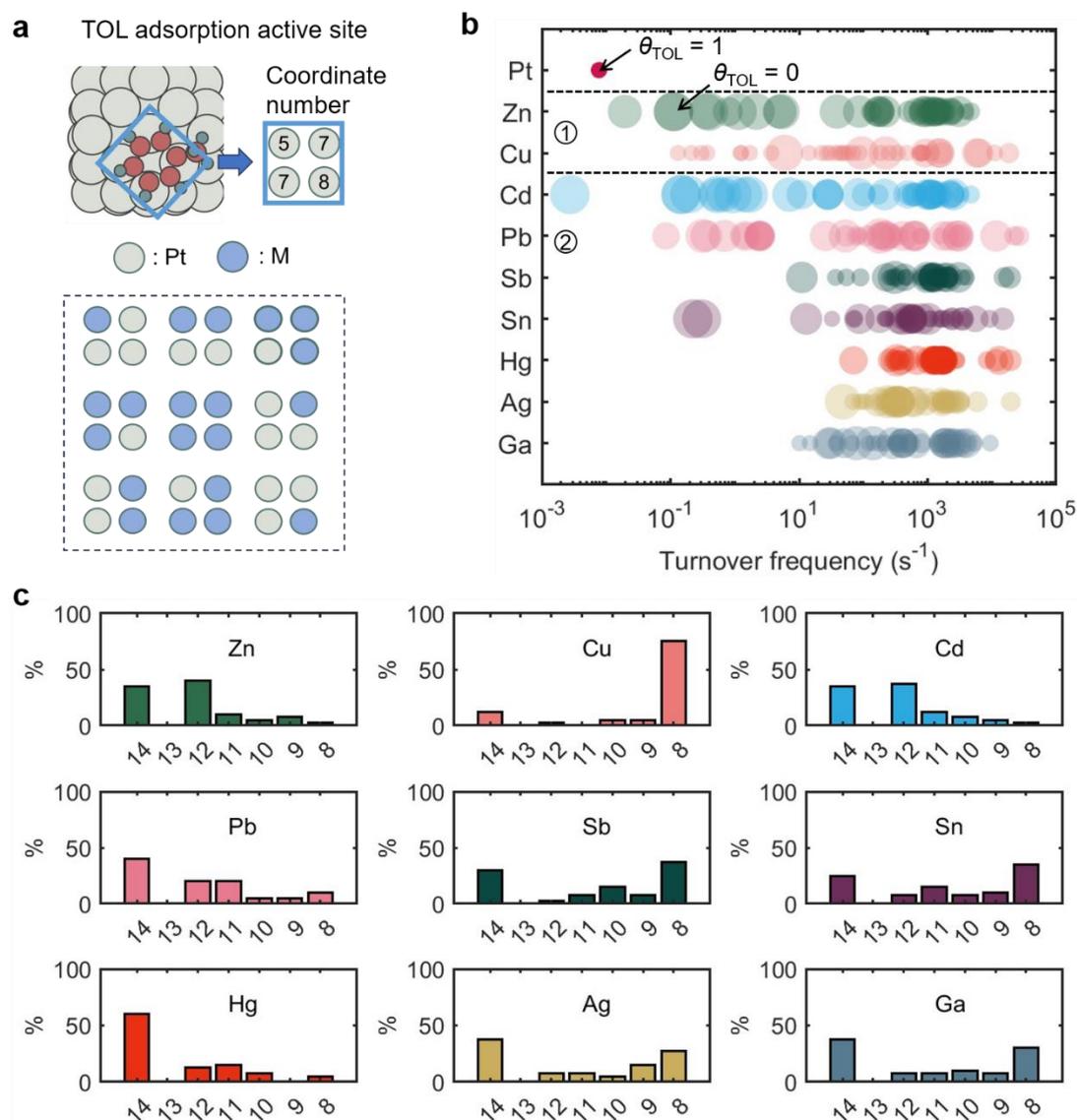
primary features accurately represent the interaction between unsaturated carbon atoms and low-coordinated metal atoms that are part of heteroatomic active sites. Upon constructing and validating the GPR model, we will now harness this model to understand how tuning the composition of the active site and the first nearest neighbors impacts the kinetics of methyl cyclohexane dehydrogenation.

### 3.3 Screening the kinetics of bimetallic active sites for methyl cyclohexane dehydrogenation



**Figure 7** Elementary reaction steps for MCH dehydrogenation. The active site for the adsorption of  $C_7$  hydrocarbons is represented by  $*$  while the active sites for the adsorption of atomic hydrogen ( $H^\Delta$ ) is  $\Delta$ .

A microkinetic model is developed to calculate the reaction kinetics. The energetics inputted into the microkinetic model are obtained from the GPR models that were formulated and validated in sections 3.1 and 3.2. The reaction pathway of methyl cyclohexane dehydrogenation is shown in Figure 7, containing 10 elementary steps and 9 reaction intermediates. The 10 elementary steps include the adsorption and desorption of methyl cyclohexane, toluene, and  $H_2$ , the C-H cleavage steps and the H-H recoupling step. Since the microkinetic model is applied to evaluate the performance of different bimetallic catalysts, we simulated the turnover frequency under a fixed 5% conversion, where the pressures of methyl cyclohexane, toluene and hydrogen are constant as 95 kPa, 5 kPa and 15 kPa, respectively. Dual-site model was applied in the simulation where  $C_7$  and H species adsorb at different active sites, marked by  $*$  and  $\Delta$ , respectively.



**Figure 8** (a) Nine Pt-M active sites of 55-atom nanoclusters having a CUBO morphology. Nine doping elements (M) are considered. For each Pt-M active site structure shown in (a), four compositions of Pt and M in the first nearest neighbors are selected, leading to 36 structures per Pt-M system. (b) Turnover frequencies for methyl cyclohexane dehydrogenation are computed at 5% conversion at 600 K, 95 kPa methyl cyclohexane, 5 kPa toluene, and 15 kPa hydrogen. The size of each data point represents the surface coverage of  $C_7H_8^*$  ( $\theta_{TOL}$ ) at steady state. Monometallic Pt is used as the reference. Zn and Cu in ① are the elements used in the training dataset while the other elements in ② are outside the training dataset. Larger marker sizes indicate a lower  $\theta_{TOL}$ ; (c) Distribution of the most abundant surface species plotted for each doping element M. 14  $\rightarrow$  8 on the x-axes refer to percentage coverage of  $C_7H_{14}^* \rightarrow C_7H_8^*$ .

Utilizing the GPR models shown in Figure 6b, we determined the reaction energy for each elementary step of the methyl cyclohexane dehydrogenation pathway. To simplify

the catalyst screening process, we assumed that zero-point energies and entropies of adsorbed species are insensitive to the chemical composition of the Pt-M active site. Thus, these values are determined on the reference active site on the Pt<sub>55</sub> nanocluster. These assumptions are valid because contributions from zero-point energies are typically below 0.10 eV and changes in vibrational entropy during surface reactions involving chemisorbed species are typically small. These assumptions are supported by Table S5 and by previous studies in the literature<sup>13,40</sup>. Since we are focusing on Pt-rich alloys, and since H<sup>Δ</sup> diffuses rapidly at 600 K<sup>55</sup>, the adsorption energy of H<sup>Δ</sup> was approximated to that observed on Pt<sub>55</sub>. Given that C<sub>7</sub>H<sub>14</sub><sup>\*</sup> is physisorbed, the adsorption energy at a different Pt-M active site was regarded as the same as the adsorption energy of C<sub>7</sub>H<sub>14</sub><sup>\*</sup> on Pt<sub>55</sub>. The activation barriers are inferred using a Brønsted–Evans–Polanyi (BEP) relation, with a BEP coefficient of 0.74 derived from dehydrogenation steps on Pt<sub>55</sub> (Figure S2). This BEP coefficient is indicative of a late transition state for C-H bond breakage—a commonality in light alkane dehydrogenation reactions<sup>56,57</sup>. Such BEP relationships are established approaches for estimating the activation barriers of C-H scission reactions that have been generalized across different types of active sites, including bimetallic alloys<sup>58</sup>. We determined the reaction kinetics at different Pt-M active sites, with the terrace site at Pt<sub>55</sub> chosen as the reference active site. 36 configurations of Pt-M are generated for nine M elements. These elements include Ag, Cu and Zn, which weaken the adsorption energy of toluene (Figure 4b). As shown in Figure 8a, we first select 9 different active site motifs. For each active site we randomly replaced 1, 4, 7 and 10 Pt atoms from the first nearest neighbors by M. Notably, the predictions assume a pristine nanocluster surface; however, microkinetic simulations reveal a full layer of adsorbed toluene covering Pt<sub>55</sub>. Hence, the effects of surface coverage on the desorption energy of toluene are explicitly included. An empirical term representing coverage effects is computed to be 0.53 eV based on analyzing the coverage-dependent adsorption energies of toluene. The positive value of this term implies a weakening in the adsorption energy of toluene with increasing surface coverage, as shown Figure S10. The magnitude of this coverage effect for toluene is similar to what has been observed in the literature for the adsorption energy of phenol under high surface coverages<sup>59,60</sup>. Here we assumed that the coverage effect is always 0.53 eV, irrespective of the identity of the C<sub>7</sub> species that are most abundant under pseudo steady state conditions. Future work will refine this term through experimental validation<sup>61</sup>. For each active site, we calculated the turnover frequency and the surface coverage of adsorbed toluene,  $\theta_{\text{TOL}}$ . The  $\theta_{\text{TOL}}$  is connected to the likelihood of benzene formation, since a high  $\theta_{\text{TOL}}$  increases the rate of deleterious C-C scission forming CH<sub>3</sub><sup>\*</sup> and adsorbed benzene.

In Figure 8a, the magnitude of  $\theta_{\text{TOL}}$  is inversely proportional to the size of the datapoints. For catalysts exhibiting turnover frequencies resembling monometallic Pt active sites, the surfaces are covered by adsorbed toluene and the reaction rate is hindered by the sluggish desorption of toluene. For catalysts exhibiting turnover frequencies that are higher than monometallic Pt active sites, the desorption of toluene is faster and is no longer the only rate-determining step. The rate determining steps are qualitatively inferred by inspecting the forward, reverse, net rates, and the surface coverages.

Figure 8b shows the distribution of the most abundant surface species ( $C_7H_x^*$ , where  $x$  ranges from 8 to 14). The reference Pt site was not plotted as adsorbed toluene was the most abundant species on Pt. Replacing Pt environments with M enables the faster desorption of toluene, but at the expense of increasing the activation barriers for C-H bond scission, as inferred from the BEP relationship. Hence a higher concentration of M leads to sluggish kinetics of C-H bond scission mainly from the first and second dehydrogenation step. As inferred from Figure 8b, the dehydrogenation of  $C_7H_{10}$  can also be rate-limiting for certain elements such as Sb, Sn, Pb, Ag and Ga.

In general, we found that modifying Pt nanoclusters with elements from group IB and IIB, and post-transition metals results in weakened interactions of toluene with the active site and, consequently, higher reaction rates. Such alloys also reduce the cost of the catalyst by reducing the composition of Pt in the catalyst formulation.

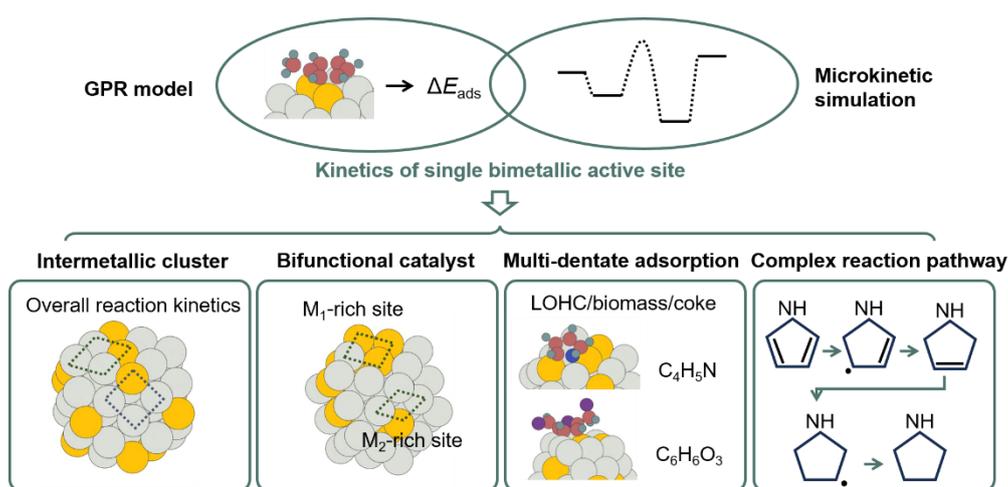
## 4 Discussion

Cyclic reaction intermediates seen in dehydrogenation of hydrogen carriers like methyl cyclohexane adsorb in multidentate configurations on active sites. For bimetallic nanoclusters, such active sites are moreover heteroatomic in composition and span diverse coordination numbers. These complexities have impeded the use of *in silico* methods to design effective catalysts for dehydrogenating liquid organic hydrogen carriers. We overcome these challenges using a novel catalyst design workflow that leverages the uncertainty quantification capabilities of Gaussian Process Regression (GPR). By selecting physics-inspired atomic properties as primary features, our machine learning model predicts adsorption energies of reaction intermediates using markedly smaller datasets as accurately as state-of-the-art approaches that are built for smaller adsorbates on less complex active sites<sup>12,51,62</sup>. Such physics-inspired features also lead to a model that is transferable across variations in the structure and composition of not only the active site but also of the entire nanocluster. Combining these GPR models with a microkinetic model, we developed a machine-learning assisted catalyst screening method to identify Pt-based bimetallic nanoclusters for the dehydrogenation of methyl cyclohexane. A high-throughput screening over different transition metals and active site configurations, we found that elements from groups IB and IIB, and post-transition metals are promising promoters for methyl cyclohexane dehydrogenation.

Our findings align with experimental observations of the dehydrogenation of both light alkanes and methyl cyclohexane. Similar increases in selectivity were obtained by facilitating the desorption of unsaturated hydrocarbon products through alloying with group 1B elements. During propane dehydrogenation<sup>63</sup>, Pt-Cu and Pt-Ag exhibited a 96% selectivity towards propylene that exceeds the 90% selectivity seen on monometallic Pt. Experimental studies also reported Pt/Cu single-atom alloys as coke-resistant C-H activation catalysts<sup>64</sup>. Similarly, post-transition metals (Pb, Sn, Sb) were reported as promoter to Pt nanoparticles for ethane dehydrogenation<sup>65</sup>.

For methyl cyclohexane dehydrogenation, experimental studies reported that the desorption of toluene is the rate-determining step on Pt catalyst<sup>66,67</sup>, an observation that is congruent with our microkinetic analysis on active sites located at the (100) facet of nanoclusters. Furthermore, including other transition metal, like Cu, Zn, Ga and Sn, facilitated the desorption of toluene and limited the formation of coke<sup>68,27,69</sup>. These experimental findings are also consistent with our computational predictions. Beyond the elements with a full d band, enhanced reactivity and selectivity were obtained by alloying Pt with Fe, Co, Mo, Mn, and Ti<sup>27,70</sup>. In contrast, our GPR model predicts that the metallic form of these elements will not weaken the adsorption energy of toluene relative to monometallic Pt (Figure 4), and thus these elements should not promote the reactivity and selectivity of methyl cyclohexane dehydrogenation. The deviation for Fe and Co may be caused by the spin polarization which was included in the DFT calculation and the GPR model. Mo, Mn, and Ti however exist in the form of oxides under the reaction temperature of 500~650 K during methyl cyclohexane dehydrogenation. Unlike their metallic counterparts, such oxide clusters may weaken the adsorption energy of toluene. Furthermore, interactions of these oxides with the catalyst support (usually Al<sub>2</sub>O<sub>3</sub> or SiO<sub>2</sub>), can influence the acidity of the support. Tuning the acidity of supporting material has been reported to influence the reactivity and selectivity of dehydrogenation reactions<sup>71</sup>. Although the effects of the support are not discussed in the present work, they are a natural next extension of our computational framework. Herein, we aim to overcome the first step of building an in-silico catalyst design paradigm that is capable of identifying bimetallic active sites for such complex reaction pathways.

In general, our computational findings align with strategies employed in the dehydrogenation of light alkane and methyl cyclohexane over bimetallic catalysts, underscoring the utility of a secondary metal components in weakening product-catalyst interactions.



**Figure 9** Applications of our machine learning model to catalytic reactions involving large molecules. The capabilities of computing reaction energetics at each active site of the nanocluster can be harnessed to determine the overall reaction kinetics across

*the nanoparticle. Explicit considerations of bifunctional coupling between different active sites of the nanoparticle are possible. This framework is translatable to other liquid organic hydrogen carriers and to cyclic hydrocarbons that are precursor to coke.*

We outline the applications of our method to other catalytic systems in Figure 9. The GPR model predicts active-site-specific adsorption energies of cyclic organic species. This capability can be leveraged to map the reaction energetics, and when paired with a microkinetic model, the reaction kinetics over all the possible active sites of a nanocluster. These active site distributions can either be characterized from experiments or generated using theoretical models of nanoclusters<sup>72</sup>. Moreover, the active-site-specific reaction energetics provided by the GPR model can be used to determine the extent of bifunctional coupling between active sites, as has been postulated for bimetallic alloys<sup>73</sup>. In addition to evaluating the reaction kinetics of methyl cyclohexane dehydrogenation intermediates, the active learning simulator can be used to retrain the model for other cyclic organic species. The reaction energetics of such adsorbates are useful in understanding the deactivation rate of bimetallic sites through coking. The formation of coke is the most frequent cause of catalyst deactivation in industrial processes<sup>74</sup>. As coke molecules are formed from unsaturated species like alkenes, aromatics, cyclic alkenes, and cyclic dienes<sup>74</sup>, the likelihood of poisoning a bimetallic active site can be also predicted by our GPR model. A similar in silico paradigm can be applied directly to identify improved catalysts for the reverse reaction, the hydrogenation of toluene to methyl cyclohexane<sup>75</sup>, and other heterogeneous catalysis of large organic compounds having complex reaction pathways.

## 5 Conclusions

The computational-guided design of tailored catalysts for dehydrogenating liquid organic hydrogen carriers is limited by the complexities of predicting adsorption energies of unsaturated cyclic hydrocarbons on low-symmetry active sites. To circumvent these limitations, we present a physics-inspired machine learning model that predicts the adsorption energies of species encountered during the dehydrogenation of methyl cyclohexane, a hydrogen carrier that is deployed at industrial scale. We formulate a gaussian process regression (GPR) model that leverage its intrinsic uncertainty quantification capabilities and uses inputs that are inspired by physical theories of chemisorption. The model inputs include cohesive energies of bulk metals, electronegativity, and carbon-metal bond lengths. Across structurally and chemically complex active site environments in bimetallic alloys, this GPR model rapidly and accurately predicts the adsorption energies of reaction intermediates observed during methyl cyclohexane dehydrogenation. Notwithstanding the complexities of multi-dentate configurations of  $C_7H_x$  ( $x = 8 - 13$ ) species that adsorb on low-symmetry active sites, the model predicts adsorption energies with mean absolute errors ranging from 0.12 eV to 0.29 eV on blind test sets. The physics-inspired

inputs ensure seamless transferability of this model across active sites located in nanoclusters having varied morphologies and compositions.

The energetics derived from the GPR model are used to compute rate constants of elementary steps of methyl cyclohexane dehydrogenation. Using this microkinetic model, we perform a high-throughput virtual screening of bimetallic nanocluster catalysts for methyl cyclohexane dehydrogenation. Our model identifies Pt-based nanoclusters alloyed with metals featuring a complete d-band (group IB, IIB, and post transition elements) as optimal catalysts, aligning with experimental observations in the literature. Such bimetallic compositions not only accelerate the rate of toluene desorption but concomitantly lower the surface coverage of toluene. The latter prevents deleterious side reactions like the demethylation of toluene to benzene. The scalable and efficient approach introduced in this work marks a significant advancement in the predictive design of catalysts for reactions involving large cyclic hydrocarbons that react on low-symmetry active sites. The framework introduced in this work can be translated to design tailored catalysts for dehydrogenating other types of liquid organic hydrogen carriers, thus accelerating the deployment of technologies that enable the trans-oceanic shipment of low-carbon hydrogen.

## Acknowledgements

This work was funded by the Low Carbon Energy Research Funding Initiative administered by the Agency for Science, Technology and Research (A\*STAR), with the award number being U2102d2013. We would like to acknowledge the High-Performance Computing Centre of Nanyang Technological University Singapore, for providing the computing resources, facilities, and services that have contributed significantly to this work. The computational work for this article was partially performed on resources of the National Supercomputing Centre, Singapore ([www.nsc.sg](http://www.nsc.sg)).

## Author Contributions

Author contributions are defined based on the CRediT (Contributor Roles Taxonomy) and listed alphabetically.

Conceptualization: CL and TC conceptualized the machine learning assisted catalyst screening approach. Data curation: CL and UA constructed the code and data repository. Formal analysis: AP, BL, CL, and UA performed the DFT calculations and CL did the kinetic analyses. Investigation: AP, BL, CL, and UA collected the computational data; CL built the machine learning model and microkinetic model. Methodology: CL and TC designed the methodologies. Project administration: RX and TC managed and coordinated the research activities. Resources: TC provided the computational resources. Software: CL developed the codes and algorithms used in the computational studies, implementing Quantum ESPRESSO and MATLAB for the investigation. Supervision: TC supervised the research. Visualization: CL created all

the figures and illustrations. Writing-original draft: CL and TC drafted the main text of the original manuscript, supporting information, and data repository. Writing-review & editing: All authors reviewed and edited the manuscript.

## References

1. Preuster, P., Papp, C. & Wasserscheid, P. Liquid Organic Hydrogen Carriers (LOHCs): Toward a Hydrogen-free Hydrogen Economy. *Acc. Chem. Res.* **50**, 74–85 (2017).
2. Lee, K., Jing, Y., Wang, Y. & Yan, N. A unified view on catalytic conversion of biomass and waste plastics. *Nat. Rev. Chem.* **6**, 635–652 (2022).
3. Pacheco, M. A. & Petersen, E. E. Reaction kinetics of methylcyclohexane dehydrogenation over a sulfided Pt + ReAl<sub>2</sub>O<sub>3</sub> reforming catalyst. *J. Catal.* **96**, 507–516 (1985).
4. Mou, T. *et al.* Bridging the complexity gap in computational heterogeneous catalysis with machine learning. *Nat. Catal.* **6**, 122–136 (2023).
5. Montemore, M. M., Nwaokorie, C. F. & Kayode, G. O. General screening of surface alloys for catalysis. *Catal. Sci. Technol.* **10**, 4467–4476 (2020).
6. Wexler, R. B., Gautam, G. S., Stechel, E. B. & Carter, E. A. Factors Governing Oxygen Vacancy Formation in Oxide Perovskites. *J. Am. Chem. Soc.* **143**, 13212–13227 (2021).
7. Liu, J.-X., Richards, D., Singh, N. & Goldsmith, B. R. Activity and Selectivity Trends in Electrocatalytic Nitrate Reduction on Transition Metals. *ACS Catal.* **9**, 7052–7064 (2019).
8. Back, S., Na, J. & Ulissi, Z. W. Efficient Discovery of Active, Selective, and Stable Catalysts for Electrochemical H<sub>2</sub>O<sub>2</sub> Synthesis through Active Motif Screening. *ACS Catal.* **11**, 2483–2491 (2021).
9. Fuller, J., An, Q., Fortunelli, A. & Goddard, W. A. I. Reaction Mechanisms, Kinetics, and Improved Catalysts for Ammonia Synthesis from Hierarchical High Throughput Catalyst Design. *Acc. Chem. Res.* **55**, 1124–1134 (2022).

10. Pillai, H. S. *et al.* Interpretable design of Ir-free trimetallic electrocatalysts for ammonia oxidation with graph neural networks. *Nat. Commun.* **14**, 792 (2023).
11. Jia, H. *et al.* Computational Discovery of Codoped Single-Atom Catalysts for Methane-to-Methanol Conversion. *ACS Catal.* **14**, 2992–3005 (2024).
12. Tran, K. & Ulissi, Z. W. Active learning across intermetallics to guide discovery of electrocatalysts for CO<sub>2</sub> reduction and H<sub>2</sub> evolution. *Nat. Catal.* **1**, 696–703 (2018).
13. Wang, Y. *et al.* Descriptor-Based Microkinetic Modeling and Catalyst Screening for CO Hydrogenation. *ACS Catal.* **11**, 14545–14560 (2021).
14. Bunting, R. J., Wodaczek, F., Torabi, T. & Cheng, B. Reactivity of Single-Atom Alloy Nanoparticles: Modeling the Dehydrogenation of Propane. *J. Am. Chem. Soc.* **145**, 14894–14902 (2023).
15. Hook, A. & Celik, F. E. Predicting Selectivity for Ethane Dehydrogenation and Coke Formation Pathways over Model Pt–M Surface Alloys with *ab Initio* and Scaling Methods. *J. Phys. Chem. C* **121**, 17882–17892 (2017).
16. Seemakurthi, R. R. *et al.* Identification of a Selectivity Descriptor for Propane Dehydrogenation through Density Functional and Microkinetic Analysis on Pure Pd and Pd Alloys. *ACS Catal.* **11**, 9588–9604 (2021).
17. Hansen, M. H., Nørskov, J. K. & Bligaard, T. First principles micro-kinetic model of catalytic non-oxidative dehydrogenation of ethane over close-packed metallic facets. *J. Catal.* **374**, 161–170 (2019).
18. Li, X.-T., Chen, L., Shang, C. & Liu, Z.-P. In Situ Surface Structures of PdAg Catalyst and Their Influence on Acetylene Semihydrogenation Revealed by Machine Learning and Experiment. *J. Am. Chem. Soc.* **143**, 6281–6292 (2021).
19. Pablo-García, S. *et al.* Fast evaluation of the adsorption energy of organic molecules on metals via graph neural networks. *Nat. Comput. Sci.* **3**, 433–442

- (2023).
20. Gu, G. H., Lee, M., Jung, Y. & Vlachos, D. G. Automated exploitation of the big configuration space of large adsorbates on transition metals reveals chemistry feasibility. *Nat. Commun.* **13**, 2087 (2022).
  21. Performance of 10,000 hours of operation in Chiyoda's demo plant. *Chiyoda Corporation website* (2024).
  22. Dehydrogenation demonstration begins for green heavy vehicles. *Nanyang Technological University website* (2024).
  23. Okada, Y., Sasaki, E., Watanabe, E., Hyodo, S. & Nishijima, H. Development of dehydrogenation catalyst for hydrogen generation in organic chemical hydride method. *Int. J. Hydrog. Energy* **31**, 1348–1356 (2006).
  24. van Trimont, P. A., Marin, G. B. & Froment, G. F. Activities and selectivities for reforming reactions on unsulfided and sulfided commercial platinum and platinum-rhenium catalysts. *Appl. Catal.* **17**, 161–173 (1985).
  25. Al-ShaikhAli, A. H., Jedidi, A., Cavallo, L. & Takanabe, K. Non-precious bimetallic catalysts for selective dehydrogenation of an organic chemical hydride system. *Chem. Commun.* **51**, 12931–12934 (2015).
  26. Murata, K. *et al.* Structure–Property Relationships of Pt–Sn Nanoparticles Supported on Al<sub>2</sub>O<sub>3</sub> for the Dehydrogenation of Methylcyclohexane. *ACS Appl. Nano Mater.* **4**, 4532–4541 (2021).
  27. Nakaya, Y., Miyazaki, M., Yamazoe, S., Shimizu, K. & Furukawa, S. Active, Selective, and Durable Catalyst for Alkane Dehydrogenation Based on a Well-Designed Trimetallic Alloy. *ACS Catal.* **10**, 5163–5172 (2020).
  28. Giannozzi, P. *et al.* QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter* **21**, 395502

- (2009).
29. Hjorth Larsen, A. *et al.* The atomic simulation environment—a Python library for working with atoms. *J. Phys. Condens. Matter* **29**, 273002 (2017).
  30. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).
  31. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010).
  32. Vanderbilt, D. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B* **41**, 7892 (1990).
  33. Torres, J. A. G., Jennings, P. C., Hansen, M. H., Boes, J. R. & Bligaard, T. Low-scaling algorithm for nudged elastic band calculations using a surrogate machine learning model. *Phys. Rev. Lett.* **122**, 156001 (2019).
  34. Medford, A. J. *et al.* Thermochemistry and micro-kinetic analysis of methanol synthesis on ZnO (0001). *J. Catal.* **309**, 397–407 (2014).
  35. Nørskov, J. K., Bligaard, T., Rossmeisl, J. & Christensen, C. H. Towards the computational design of solid catalysts. *Nat. Chem.* **1**, 37–46 (2009).
  36. Choksi, T. S., Roling, L. T., Streibel, V. & Abild-Pedersen, F. Predicting Adsorption Properties of Catalytic Descriptors on Bimetallic Nanoalloys with Site-Specific Precision. *J. Phys. Chem. Lett.* **10**, 1852–1859 (2019).
  37. Prodan, E. & Kohn, W. Nearsightedness of electronic matter. *Proc. Natl. Acad. Sci.* **102**, 11635–11638 (2005).
  38. Nørskov, J. K. *et al.* The nature of the active site in heterogeneous metal catalysis. *Chem. Soc. Rev.* **37**, 2163–2171 (2008).

39. Vorotnikov, V. & Vlachos, D. G. Group Additivity and Modified Linear Scaling Relations for Estimating Surface Thermochemistry on Transition Metal Surfaces: Application to Furanics. *J. Phys. Chem. C* **119**, 10417–10426 (2015).
40. Halldin Stenlid, J., Streibel, V., Choksi, T. S. & Abild-Pedersen, F. Assessing catalytic rates of bimetallic nanoparticles with active-site specificity: A case study using NO decomposition. *Chem Catal.* **3**, (2023).
41. Saini, S., Halldin Stenlid, J., Deo, S., Plessow, P. N. & Abild-Pedersen, F. A First-Principles Approach to Modeling Surface Site Stabilities on Multimetallic Catalysts. *ACS Catal.* **14**, 874–885 (2024).
42. Campbell, C. T. The Energetics of Supported Metal Nanoparticles: Relationships to Sintering Rates and Catalytic Activity. *Acc. Chem. Res.* **46**, 1712–1719 (2013).
43. Gao, W. *et al.* Determining the adsorption energies of small molecules with the intrinsic properties of adsorbates and substrates. *Nat. Commun.* **11**, 1196 (2020).
44. Ma, X., Li, Z., Achenie, L. E. K. & Xin, H. Machine-Learning-Augmented Chemisorption Model for CO<sub>2</sub> Electroreduction Catalyst Screening. *J. Phys. Chem. Lett.* **6**, 3528–3533 (2015).
45. Xin, H., Holewinski, A. & Linic, S. Predictive Structure–Reactivity Models for Rapid Screening of Pt-Based Multimetallic Electrocatalysts for the Oxygen Reduction Reaction. *ACS Catal.* **2**, 12–16 (2012).
46. Calle-Vallejo, F. The ABC of Generalized Coordination Numbers and Their Use as a Descriptor in Electrocatalysis. *Adv. Sci.* **10**, 2207644 (2023).
47. Kittel, C., McEuen, P. & Sons, J. W. &. *Introduction to Solid State Physics*. (John Wiley & Sons, 2015).
48. Pauling, L. *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry*. (Cornell University Press,

- 1960).
49. Liu, X., Cai, C., Zhao, W., Peng, H.-J. & Wang, T. Machine Learning-Assisted Screening of Stepped Alloy Surfaces for C1 Catalysis. *ACS Catal.* **12**, 4252–4260 (2022).
  50. Sun, Z. *et al.* Machine learning accelerated calculation and design of electrocatalysts for CO<sub>2</sub> reduction. *SmartMat* **3**, 68–83 (2022).
  51. Kayode, G. O. & Montemore, M. M. Latent Variable Machine Learning Framework for Catalysis: General Models, Transfer Learning, and Interpretability. *JACS Au* **4**, 80–91 (2024).
  52. Li, H., Jiao, Y., Davey, K. & Qiao, S.-Z. Data-Driven Machine Learning for Understanding Surface Structures of Heterogeneous Catalysts. *Angew. Chem. Int. Ed.* **62**, e202216383 (2023).
  53. Unke, O. T. *et al.* Machine Learning Force Fields. *Chem. Rev.* **121**, 10142–10186 (2021).
  54. Wang, L. *et al.* Tunable intrinsic strain in two-dimensional transition metal electrocatalysts. *Science* **363**, 870–874 (2019).
  55. Peng, G. & Mavrikakis, M. Adsorbate Diffusion on Transition Metal Nanoparticles. *Nano Lett.* **15**, 629–634 (2015).
  56. Michaelides, A. *et al.* Identification of General Linear Relationships between Activation Energies and Enthalpy Changes for Dissociation Reactions at Surfaces. *J. Am. Chem. Soc.* **125**, 3704–3705 (2003).
  57. Ferrin, P. *et al.* Modeling Ethanol Decomposition on Transition Metals: A Combined Application of Scaling and Brønsted–Evans–Polanyi Relations. *J. Am. Chem. Soc.* **131**, 5809–5815 (2009).
  58. Latimer, A. A. *et al.* Understanding trends in C–H bond activation in heterogeneous

- catalysis. *Nat. Mater.* **16**, 225–229 (2017).
59. Chaudhary, N., Hensley, A., Collinge, G., Wang, Y. & McEwen, J.-S. Coverage-Dependent Adsorption of Phenol on Pt(111) from First Principles. *J. Phys. Chem. C* **124**, 356–362 (2020).
60. Cardwell, N., Hensley, A. J. R., Wang, Y. & McEwen, J.-S. Capturing the Coverage Dependence of Aromatics' Adsorption through Mean-Field Models. *J. Phys. Chem. A* **127**, 10693–10700 (2023).
61. Bhandari, S., Rangarajan, S. & Mavrikakis, M. Combining Computational Modeling with Reaction Kinetics Experiments for Elucidating the In Situ Nature of the Active Site in Catalysis. *Acc. Chem. Res.* **53**, 1893–1904 (2020).
62. Ghanekar, P. G., Deshpande, S. & Greeley, J. Adsorbate chemical environment-based machine learning framework for heterogeneous catalysis. *Nat. Commun.* **13**, 5788 (2022).
63. Ren, G.-Q. *et al.* Effect of group IB metals on the dehydrogenation of propane to propylene over anti-sintering Pt/MgAl<sub>2</sub>O<sub>4</sub>. *J. Catal.* **366**, 115–126 (2018).
64. Marcinkowski, M. D. *et al.* Pt/Cu single-atom alloys as coke-resistant catalysts for efficient C–H activation. *Nat. Chem.* **10**, 325–332 (2018).
65. Dai, Y. *et al.* Recent progress in heterogeneous metal and metal oxide catalysts for direct dehydrogenation of ethane and propane. *Chem. Soc. Rev.* **50**, 5590–5630 (2021).
66. Sinfelt, J. H., Hurwitz, H. & Shulman, R. A. KINETICS OF METHYLCYCLOHEXANE DEHYDROGENATION OVER PT—Al<sub>2</sub>O<sub>3</sub>. *J. Phys. Chem.* **64**, 1559–1562 (1960).
67. Jossens, L. W. & Petersen, E. E. Fouling of a platinum reforming catalyst accompanying the dehydrogenation of methyl cyclohexane. *J. Catal.* **73**, 377–386

- (1982).
68. Zhang, X. *et al.* Study of the carbon cycle of a hydrogen supply system over a supported Pt catalyst: methylcyclohexane–toluene–hydrogen cycle. *Catal. Sci. Technol.* **10**, 1171–1181 (2020).
69. Sebastian, O. *et al.* Stable and Selective Dehydrogenation of Methylcyclohexane using Supported Catalytically Active Liquid Metal Solutions – Ga<sub>52</sub>Pt/SiO<sub>2</sub> SCALMS. *ChemCatChem* **12**, 4533–4537 (2020).
70. Sekine, Y. & Higo, T. Recent Trends on the Dehydrogenation Catalysis of Liquid Organic Hydrogen Carrier (LOHC): A Review. *Top. Catal.* **64**, 470–480 (2021).
71. Kester, P. M., Iglesia, E. & Gounder, R. Parallel Alkane Dehydrogenation Routes on Brønsted Acid and Reaction-Derived Carbonaceous Active Sites in Zeolites. *J. Phys. Chem. C* **124**, 15839–15855 (2020).
72. Cheula, R., Maestri, M. & Mpourmpakis, G. Modeling Morphology and Catalytic Activity of Nanoparticle Ensembles Under Reaction Conditions. *ACS Catal.* **10**, 6149–6158 (2020).
73. Andersen, M., Medford, A. J., Nørskov, J. K. & Reuter, K. Scaling-Relation-Based Analysis of Bifunctional Catalysis: The Case for Homogeneous Bimetallic Alloys. *ACS Catal.* **7**, 3960–3967 (2017).
74. Guisnet, M. & Magnoux, P. Organic chemistry of coke formation. *Catal. Deactiv.* **212**, 83–96 (2001).
75. Fischer, A. F. & Iglesia, E. The nature of “hydrogen spillover”: Site proximity effects and gaseous intermediates in hydrogenation reactions mediated by inhibitor-scavenging mechanisms. *J. Catal.* **420**, 68–88 (2023).