The Elusive Nature of Aromatic Carbocation Intermediates in Confined Catalytic Environments

Chintu Das[1], Princy Jarngal[2], Fabian Berger*[3], Abhishek Khetan[2], and Giovanni Maria Piccini*[4]

[1]Institute of Technical and Macromolecular Chemistry, RWTH Aachen University, Aachen 52074, Germany
[2]Multiscale Modeling of Heterogeneous Catalysis in Energy Systems, The Integrated Fuel & Chemical Science Center, RWTH Aachen University, Aachen 52062, Germany
[3]Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom
[4]Department of Chemical and Geological Sciences (DSCG), University of Modena and Reggio Emilia (UNIMORE), Modena 41125, Italy

Contents

Supplementary Note 1: System preparation

A periodic model of β-zeolite was employed for all simulations. The initial crystallographic structure was obtained from the IZA database[1]. One of the Si^{4+} atoms located at the T7 site—situated at the intersection of the straight and sinusoidal channels—was substituted with Al^{3+} to generate the Brønsted acid site (BAS). The resulting framework negative charge was compensated by placing a proton on one of the four oxygen atoms coordinated to the Al center. The reactant molecules (phenol and cyclohexene) were then positioned in proximity to the BAS.

The system was first relaxed by energy minimization. Following minimization, the structure was equilibrated in the NVT ensemble to reach the reaction temperature of 573 K. The system was then subjected to pressure relaxation in the NPT ensemble at 1 bar and 573 K using isotropic cell scaling. The resulting equilibrated cell parameters (12.5 Å × 12.5 Å × 26.1 Å, all angles fixed at 90°) was used for all subsequent simulations performed in the NVT ensemble considering that the inherent rigidity of the material ensures that small changes in cell dimensions have minimal impact on the calculated free-energy differences among the intermediates.

Supplementary Note 2: Development and Validation of the MLIP

The construction of a reliable MLIP requires a training dataset that captures (meta)stable minima, i.e., reactants, intermediates, and products, as well as high-energy and transition states along the reaction pathway, while remaining compact and, ideally, unbiased. We adopt an active learning (AL) framework to iteratively refine the model during WTMetaD simulations. Following the data-efficient active learning (DEAL) strategy,[2] we combine a query-by-committee (QBC) approach[3,4] with smooth overlap of atomic positions (SOAP)[5] similarity analysis to guide the selection of new training structures and avoid redundancy.[6]

Active learning procedure

To enrich the training dataset with more representative configurations from the relevant phase space, it is essential to strike a balance between versatility and computational efficiency. While a diverse dataset ensures better generalization of the potential, an excessively large dataset can make the training process computationally demanding.

In the first step, we adopted QBC approach, where an ensemble of four MACE models was trained on different random subsets of the available data. The standard deviation of atomic force predicted by these ensemble of models was used as an uncertainty metric

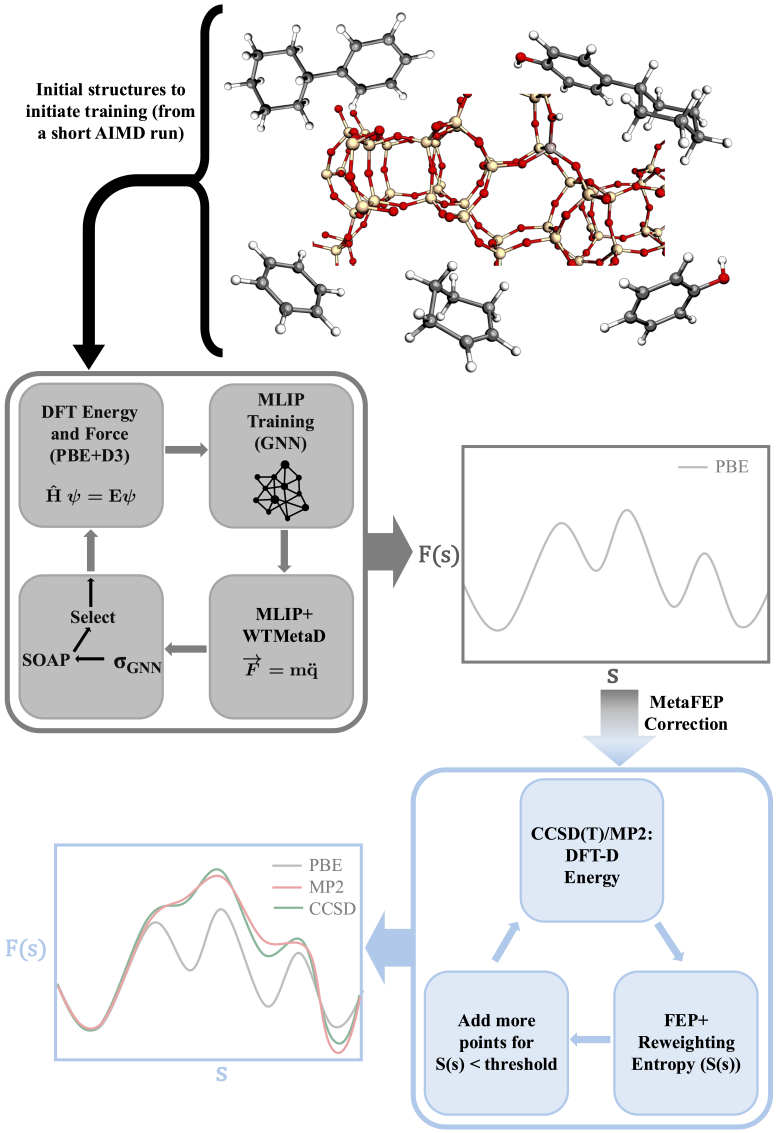
$$\sigma_{\text{atom}} = \sqrt{\frac{1}{4}\sum_{i=1}^{4} \left\|\mathbf{F}_i - \overline{\mathbf{F}}\right\|^2}, \qquad \overline{\mathbf{F}} = \frac{1}{4}\sum_{i=1}^{4} \mathbf{F}_i.$$

Configurations exhibiting large prediction uncertainty (above a chosen threshold) were identified as regions of configuration space insufficiently described by the current model and were selected for further screening.

The resulting pool of high-uncertainty configurations was further refined using the Smooth Overlap of Atomic Positions (SOAP) descriptor to quantify structural similarity[5]. For each new configuration with SOAP vector p_0, a kernel function $k(p_0, p_i)$ was evaluated against all SOAP vectors p_i in the existing training set. The similarity vector was defined as

$$\mathbf{K} = \left(k(p_0, p_1)^\zeta,\, k(p_0, p_2)^\zeta,\, \ldots\right)^T,$$

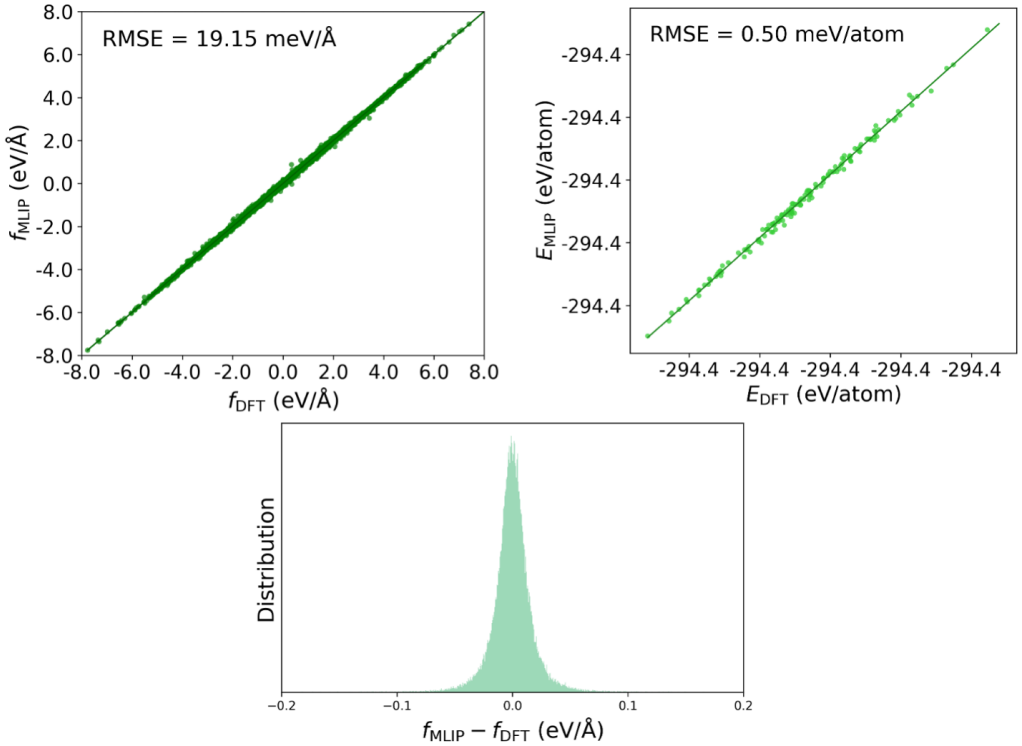
where ζ is the sensitivity exponent. A configuration was retained only if $\max(\mathbf{K}) < \tau$, with $\tau = 0.9990$. SOAP hyperparameters were set to $r_{\text{cut}} = 4$ Å, $n_{\max} = 6$, $l_{\max} = 6$, and $\zeta = 4$.[7] Since the alkylation reaction proceeds at the BAS of the zeolite framework, where distinct local environments are formed during bond-making and bond-breaking events, SOAP screening is particularly valuable for capturing these subtle local structural variations. This QBC+SOAP combined strategy ensures systematic and efficient exploration of phase space, resulting in a compact yet diverse training dataset well-suited for reactive simulations.



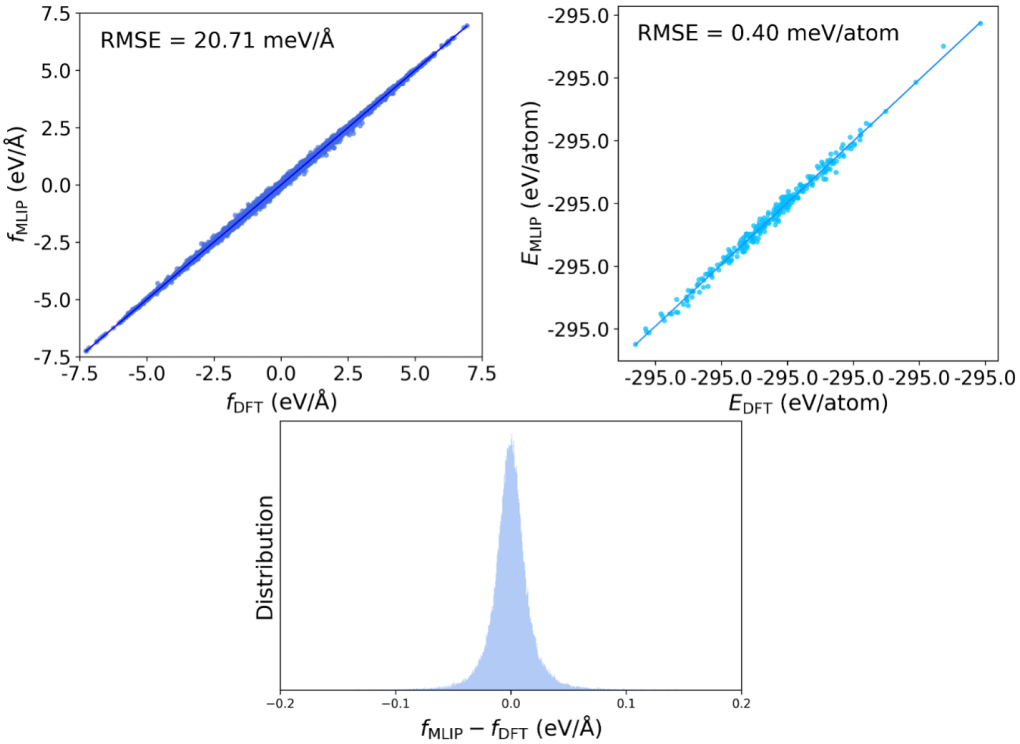
Supplementary Figure 1: Schematic illustration of the workflow for transforming PBE-level free energy surfaces (FES) to CCSD(T)/MP2-level accuracy. An MLIP trained on PBE+D3 energies and forces yields an inaccurate description of the stability of reaction intermediates. MetaFEP is employed to go beyond the PBE description and recover an accurate FES approaching the gold-standard CCSD(T) level.

The initial MLIP is trained on 1,441 DFT-labeled configurations extracted from a short AIMD WT-MetaD trajectory of the phenol alkylation reaction, which captured a single complete transition from reactants to products including all intermediates. These configurations are pre-screened using SOAP similarity to ensure structural diversity. Because this initial dataset contained only limited sampling of the transition-state region, the resulting MLIP was stable for only a few hundred picoseconds of exploration in MLIP–WTMetaD simulations. To systematically improve the model, we used these short exploratory trajectories to extract new, unique configurations (again identified with SOAP), added their DFT energies and forces to the training set, and retrained the model. Two such refinement cycles were sufficient to obtain a MLIP stable enough for nanosecond-scale WTMetaD simulations. Then the MLIP was used to run long WTMetaD simulation to explore the phase space properly and used QBC+SOAP strategy to further refine the training data set. Two cycles of QBC+SOAP refinement was performed to reach the final model which has been used for all the further simulation and analysis. The final model, which is trained on 7881 configurations, exhibits strong generalization, with force RMSEs of 20.7 meV/Å on the phenol alkylation test set and 19.1 meV/Å on the benzene alkylation

test set, as shown in Supplementary Figures 2 and 3.



Supplementary Figure 2. Evaluation of the final Model on test set of benzene alkylation (a) Correlation between MACE-predicted and DFT reference force components; (b) Correlation between MACE-predicted and DFT reference energy ; (c) distribution of force errors; Reference configurations were extracted from the final MLIP–WTMetaD trajectory.



Supplementary Figure 3. Evaluation of the final Model on test set of phenol alkylation (a) Correlation between MACE-predicted and DFT reference force components; (b) Correlation between MACE-predicted and DFT reference energy ; (c) distribution of force errors; Reference configurations were extracted from the final MLIP–WTMetaD trajectory.

Supplementary Note 3: Free Energy Calculations

Well-Tempered Metadynamics

Metadynamics[8] is an enhanced sampling technique in which a history-dependent bias potential $V(\mathbf{s}, t)$, composed of repulsive Gaussian functions, is periodically deposited along a set of collective variables (CVs). This bias discourages the system from revisiting already explored regions of the FES, thereby accelerating sampling across metastable states. At time t, the bias potential can be expressed as:

$$V(\mathbf{s}, t) = \sum_{t'=0,\tau_G,2\tau_G,\ldots<t} w \exp\left[-\sum_{i=1}^{N_s} \frac{\left(s_i(\mathbf{R}) - s_i(\mathbf{R}(t'))\right)^2}{2\sigma_i^2}\right], \tag{1}$$

where w is the Gaussian height, σ_i is the Gaussian width along the i-th CV, and τ_G is the deposition stride.

To address the issue of uncontrolled filling and poor convergence in conventional Metadynamics, the WTMetaD scheme was introduced. In WTMetaD, the Gaussian height decreases adaptively as the bias builds up, according to:

$$w(t) = w_0 \exp\left[-\frac{V(\mathbf{s}, t)}{k_B \Delta T}\right], \tag{2}$$

where ΔT is a fictitious temperature and the bias factor is defined as $\gamma = \frac{T+\Delta T}{T}$. This tempering ensures smooth convergence of the bias potential to the underlying free energy surface while maintaining efficient sampling of rare events.

In the present study, WTMetaD simulations were performed at $T = 573$ K with $w_0 = 2$ kJ mol^{-1}, $\gamma = 20$, $\sigma = 0.1$, and a deposition stride of $\tau_G = 200$ MD steps. These parameters ensured efficient exploration of the FES while avoiding excessive distortion of the underlying potential energy surface.

Free energy surface

The FES along a chosen CV can be derived from the marginal probability distribution $P(s)$ according to

$$F(s) = -\frac{1}{\beta} \ln P(s), \tag{3}$$

where $\beta = 1/(k_{\mathrm{B}}T)$. In the presence of a time-dependent bias potential $V(s, t)$, the unbiased probability distribution $P(s)$ can be recovered from the biased distribution $P_V(s)$ as

$$P(s) \propto e^{+\beta V(s)} P_V(s). \tag{4}$$

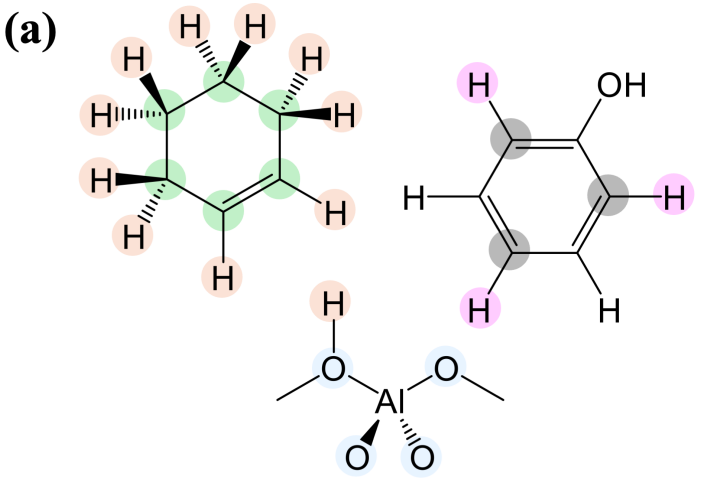
Supplementary Note 4: Collective Variable(CV)

Supplementary Figure 4 illustrates the construction of the CV for benzene alkylation, where $\mathrm{CN}(\alpha;\beta)$ denotes the coordination number between species α and β, defined through a smooth switching function as follows:

$$\mathrm{CN}(\alpha;\beta) = \sum_{i\in\alpha}\sum_{j\in\beta} \frac{1-\left(\frac{r_{ij}}{r_0}\right)^N}{1-\left(\frac{r_{ij}}{r_0}\right)^M}. \tag{5}$$

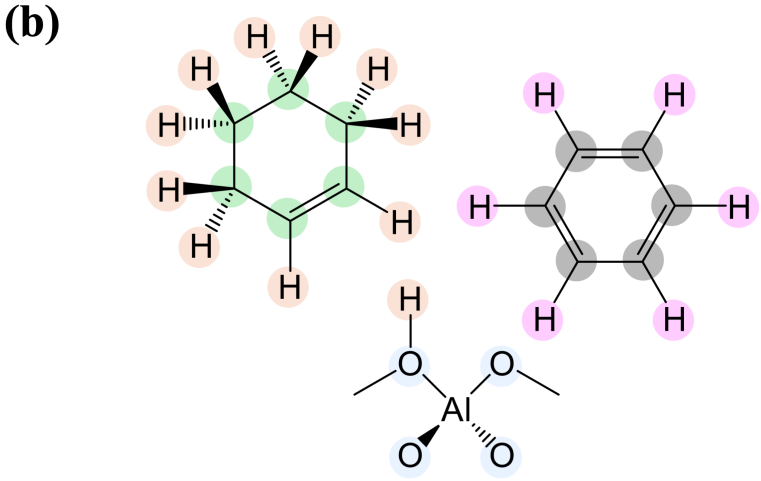
Here, r_{ij} is the distance between atoms i and j belonging to species α and β, respectively. The parameter r_0 represents the cutoff distance that determines when a bond is considered formed, while N and M are the switching function exponents controlling the smoothness of the transition between bonded and non-bonded states.

CV definition for 2D sampling



$$CV1 = 0.5\times[CN(C_{hex};H) - CN(O_{BAS};H)] + CN(C_{hex};O_{BAS})$$

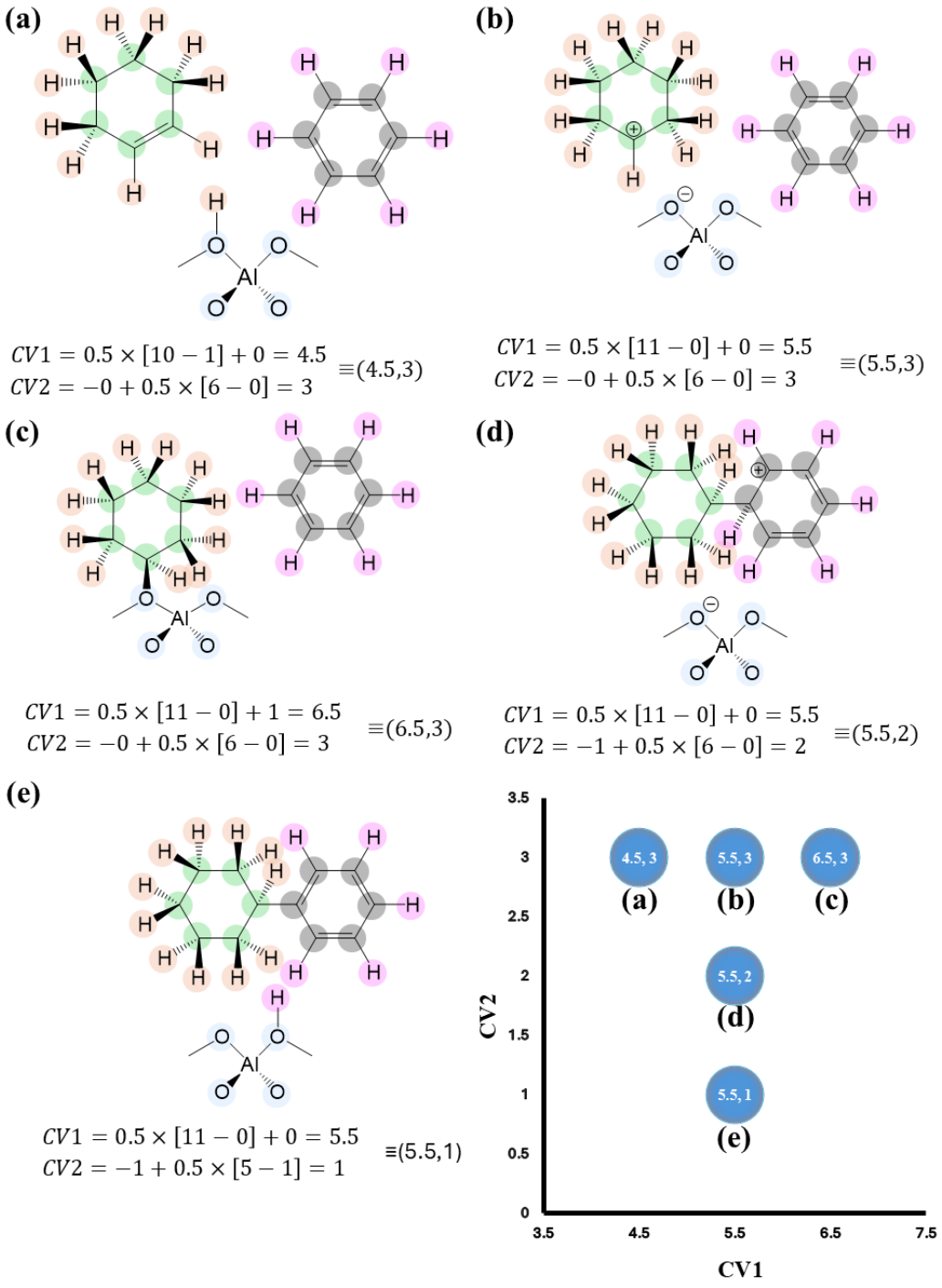
$$CV2 = -CN(C_{hex};C_{ph}) + 0.5\times[CN(C_{ph};H) - CN(O_{BAS};H)]$$



$$CV1 = 0.5\times[CN(C_{hex};H) - CN(O_{BAS};H)] + CN(C_{hex};O_{BAS})$$

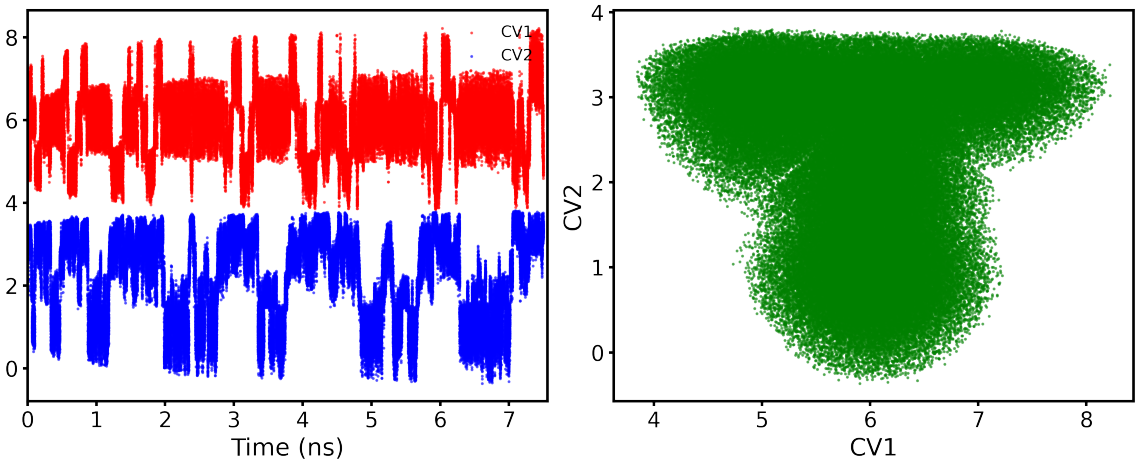
$$CV2 = -CN(C_{hex};C_{benz}) + 0.5\times[CN(C_{benz};H) - CN(O_{BAS};H)]$$

Supplementary Figure 4: Definition of CV1 and CV2 for (a) benzene alkylation and (b) Phenol Alkylation

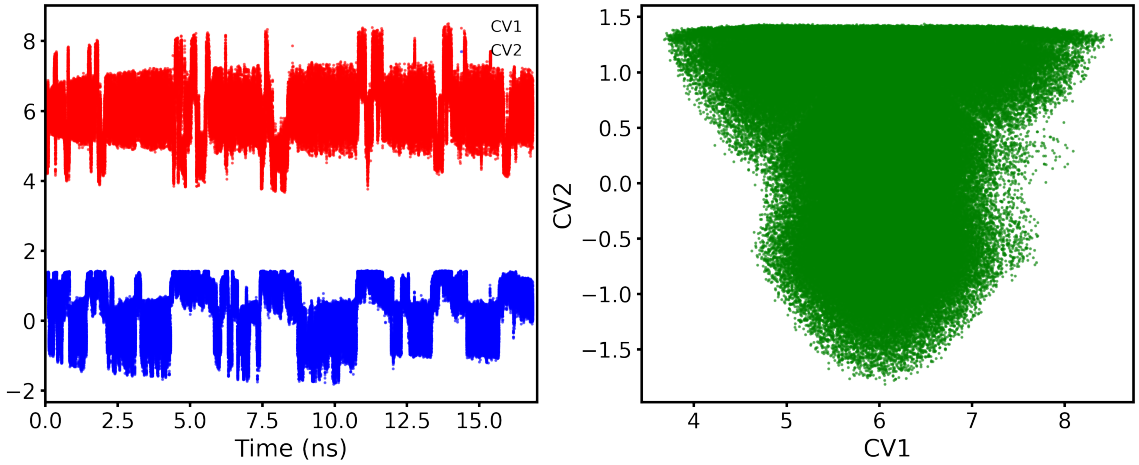


Supplementary Figure 5: Representative configurations illustrating the progression of benzene alkylation and their corresponding theoretical CV1 and CV2 values. (a) Reactant state, (b) cyclohexenium ion, (c) surface alkoxy species, (d) Wheland intermediate, and (e) alkylated product, and. This example demonstrates how the CVs evolve along the reaction coordinate, distinguishing distinct metastable states.

Exploration of phase space(2D)

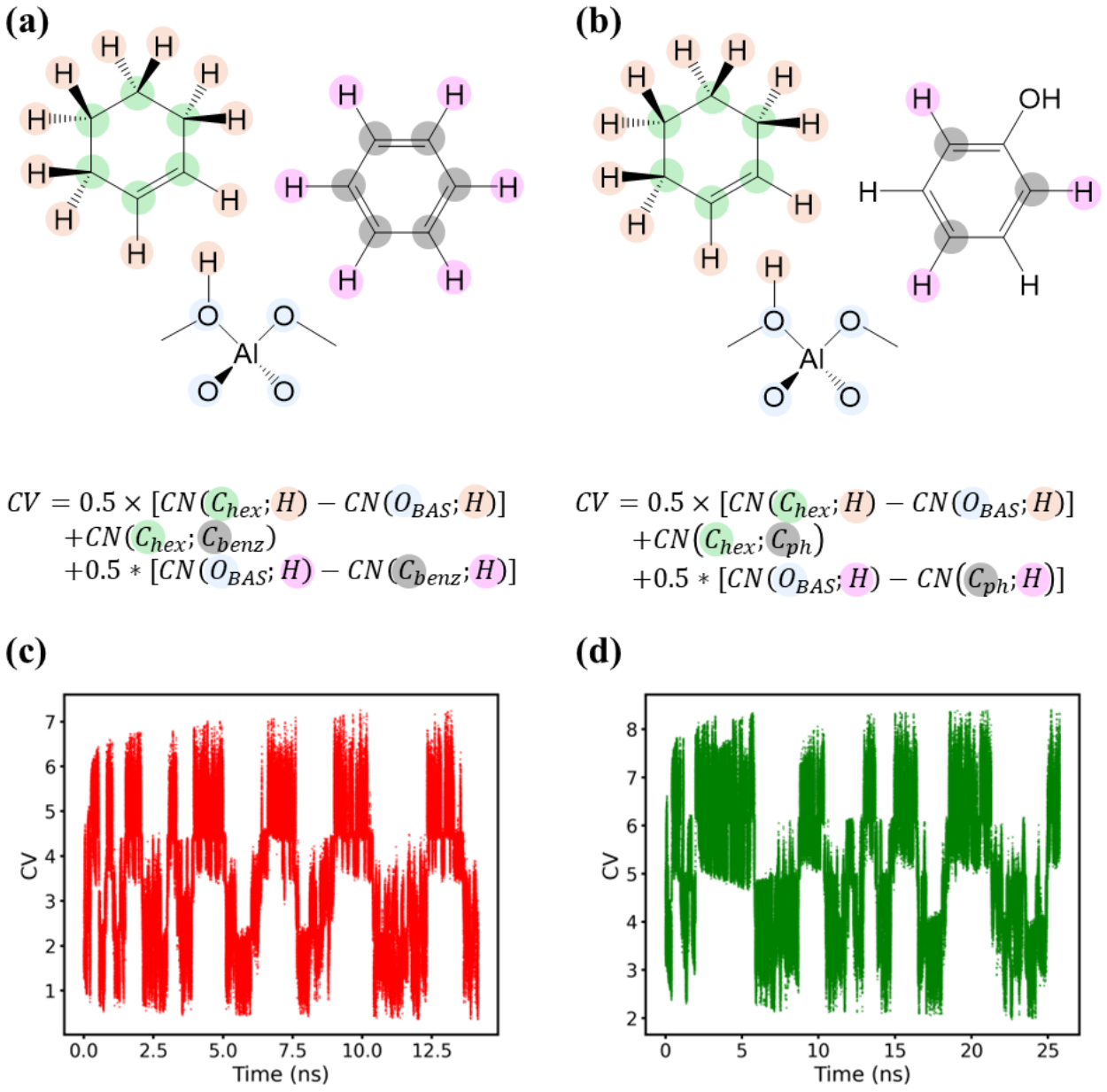


Supplementary Figure 6: Exploration of phase space along biased CVs for Benzene alkylation. Left side plot shows exploration of individual CV and right hand side plot shows the sampling in the 2D phase space



Supplementary Figure 7: Exploration of phase space along biased CVs for phenol alkylation. Left side plot shows exploration of individual CV and right hand side plot shows the sampling in the 2D phase space

CV definition for 1D sampling and exploration in phase space



Supplementary Figure 8: Definition of 1D CV for biasing main reaction pathway to apply MetaFEP. (a)for Benzene Alkylation, (b) for Phenol Alkylation. And corresponding exploration exploration of CV during MLIP-WTMetaD simulation (c) for Benzene Alkylation and (d) for Phenol Alkylation

Harmonic Restrained applied during WTMetaD simulation

A few harmonic restraints needed to apply to maintain sampling within the desired region of phase space. The biases applied from these walls are also considered along with metadynamics bias during the reweighting of FES.

Supplementary Table 1 Harmonic restraints applied in Benzene Alkylation (2D).

CV	**Upper Value**	κ	**Reason**
min[dist(C_{Ben};O_{BAS})]	6.5	200	Prevent phenol diffusing too far from BAS
min[dist(C_{Hex};O_{BAS})]	6.5	200	Prevent cyclohexene diffusing too far from BAS
CN[(C_{Hex};H_{Ben})]	0.4	200	Prevent undesired proton transfer from Benzene to Cyclohexene
CN[(C_{Ben};H_{Hex})]	0.4	800	Prevent undesired proton transfer from Cyclohexene to Benzene
dist(C_{Hex};C_{Hex})*	1.8	200	Prevent breaking of C–C bond to keep cyclic structure intact
CN[(O_{BAS};$H_{Hex,Ben}$)]	1.1	800	Prevent BAS from abstracting more than one proton
CN[(O_{nonBAS};$H_{Hex,Ben}$)]	0.2	400	Prevent proton from going to non-BAS oxygen
CV	**Lower Value**	κ	**Reason**
CN[(C_{Hex};H_{Hex})]*	1.0	400	Prevent undesired proton transfer
CN[(C_{Ben};H_{Ben},C_{Hex})]*	0.8	400	Prevent formation aryl carbanion
N5†	-0.4	800	To keep well defined sampling

Supplementary Table 2 Harmonic restrained applied in Benzene Alkylation (1D).

CV	**Upper Value**	κ	**Reason**
min[dist(C_{Ben};O_{BAS})]	6.5	200	Prevent Benzene diffusing too far from BAS
min[dist(C_{Hex};O_{BAS})]	6.5	200	Prevent cyclohexene diffusing too far from BAS
CN[(C_{Hex};H_{Ben})]	0.4	200	Prevent undesired proton transfer from Benzene to Cyclohexene
CN[(C_{Ben};H_{Hex})]	0.4	800	Prevent undesired proton transfer from Cyclohexene to Benzene
dist(C_{Hex};C_{Hex})*	1.8	200	Prevent breaking of C-C bond to keep cyclic structure intact
CN[(O_{BAS};$H_{Hex,Ben}$)]	1.1	800	Prevent BAS from abstracting more than one proton
CN[(O_{nonBAS};$H_{Hex,Ben}$)]	0.2	400	Prevent proton from going to non-BAS oxygen
CV	**Lower Value**	κ	**Reason**
min[dist(C_{Hex};O_{BAS})]	2.0	200	Prevent formation of Surface alkoxy species
CN[(C_{Hex};H_{Hex})]*	1.0	400	Prevent undesired proton transfer
CN[(C_{Ben};H_{Ben},C_{Hex})]*	0.8	400	Prevent formation aryl carbanion
N5†	-0.4	800	To keep well defined sampling

* means it contains six separate parameters and walls are applied on each of them.
† N5 = s_p− CN[(C_{Hex};C_{Phe})] where s_p is a global acid base CV. More details of s_p can be found in

Supplementary Table 3 Harmonic restrained applied in Phenol Alkylation (2D).

CV	**Upper Value**	κ	**Reason**
min[dist(C_{Phe};O_{BAS})]	6.5	200	Prevent phenol diffusing too far from BAS
min[dist(C_{Hex};O_{BAS})]	6.5	200	Prevent cyclohexene diffusing too far from BAS
CN[(C_{Hex};H_{Phe})]	0.4	200	Prevent undesired proton transfer from Phenol to Cyclohexene
CN[(C_{Phe};H_{Hex})]	0.4	800	Prevent undesired proton transfer from Cyclohexene to Phenol
CN[O_{Phe};$H_{Hex,Phe}$)]	0.2	400	Prevent formation of oxonium ion
dist(C_{Phe};O_{Phe})	1.5	200	Prevent breaking of C-O bond of phenol
CN[C_{Phe-m};C_{Hex})]	0.5	400	Prevent formation of meta product
CN[C_{Phe-m};$H_{Phe,Hex}$)]$^{\#}$	0.6	400	Prevent protonation of meta C of phenol
dist(C_{Hex};C_{Hex})*	1.8	200	Prevent breaking of C-C bond to keep cyclic structure intact
CN[(O_{BAS};$H_{Hex,Phe}$)]	1.1	800	Prevent BAS from abstracting more than one proton
CN[(O_{nonBAS};$H_{Hex,Phe}$)]	0.2	400	Prevent proton from going to non-BAS oxygen
CV	**Lower Value**	κ	**Reason**
dist(O_{Phe};C_{Hex})	2.5	400	Prevent formation O-Alkylated product
CN[O_{Phe};H_{Phe})]	0.9	1200	Prevent formation of phenoxide
CN[(C_{Hex};H_{Hex})]*	1.0	400	Prevent undesired proton transfer
CN[(C_{Phe};H_{Phe},C_{Hex})]$^{\#}$	0.8	400	Prevent formation aryl carbanion
N5†	-0.4	800	To keep well defined sampling

Supplementary Table 4 Harmonic restrained applied in Phenol Alkylation (1D).

CV	**Upper Value**	κ	**Reason**
min[dist(C_{Phe};O_{BAS})]	6.5	200	Prevent phenol diffusing too far from BAS
min[dist(C_{Hex};O_{BAS})]	6.5	200	Prevent cyclohexene diffusing too far from BAS
CN[(C_{Hex};H_{Phe})]	0.4	200	Prevent undesired proton transfer from Phenol to Cyclohexene
CN[(C_{Phe};H_{Hex})]	0.4	800	Prevent undesired proton transfer from Cyclohexene to Phenol
CN[O_{Phe};$H_{Hex,Phe}$)]	0.2	400	Prevent formation of oxonium ion
dist(C_{Phe};O_{Phe})	1.5	200	Prevent breaking of C-O bond of phenol
CN[C_{Phe-m};C_{Hex})]	0.5	400	Prevent formation of meta product
CN[C_{Phe-m};$H_{Phe,Hex}$)]$^{\#}$	0.6	400	Prevent protonation of meta C of phenol
dist(C_{Hex};C_{Hex})*	1.8	200	Prevent breaking of C-C bond to keep cyclic structure intact
CN[(O_{BAS};$H_{Hex,Phe}$)]	1.1	800	Prevent BAS from abstracting more than one proton
CN[(O_{nonBAS};$H_{Hex,Phe}$)]	0.2	400	Prevent proton from going to non-BAS oxygen
CV	**Lower Value**	κ	**Reason**
dist(O_{Phe};C_{Hex})	2.5	400	Prevent formation O-Alkylated product
CN[O_{Phe};H_{Phe})]	0.9	1200	Prevent formation of phenoxide
min[dist(C_{Hex};O_{BAS})]	2.0	200	Prevent formation of Surface alkoxy species
CN[(C_{Hex};H_{Hex})]*	1.0	400	Prevent undesired proton transfer
CN[(C_{Phe};H_{Phe},C_{Hex})]$^{\#}$	0.8	400	Prevent formation aryl carbanion
N5†	-0.4	800	To keep well defined sampling

* means it contains six separate parameters and walls are applied on each of them.
$^{\#}$ means it contains three separate parameters and walls are applied on each of them.
† N5 = s_p− CN[(C_{Hex};C_{Phe})] where s_p is a global acid base CV. More details of s_p can be found in ref[9]

Supplementary Note 5: MetaFEP

The FES obtained from MLIP-based WTMetaD simulations correspond to the target DFT level of theory (PBE+D3). To achieve higher accuracy, these FES were perturbatively corrected using the MetaFEP[10] approach, which combines Metadynamics sampling with Free Energy Perturbation (FEP). From the converged portion of each WTMetaD trajectory, a representative subset of configurations were extracted, and single point energies were evaluated at a higher level of theory. The corresponding higher-level free energy, $F_{\mathrm{HL}}(s)$, is then obtained from the lower-level free energy, $F_{\mathrm{LL}}(s)$, through the relationship:

$$F_{\mathrm{HL}}(s) = F_{\mathrm{LL}}(s) + \Delta F_{\mathrm{FEP}}(s), \tag{6}$$

where the perturbative correction term is given by

$$\Delta F_{\mathrm{FEP}}(s) = \frac{1}{\beta} \ln \frac{\sum_i^N w_i^{\mathrm{M}}(\mathbf{R}_i) w_i^{\mathrm{P}}(\mathbf{R}_i)\, \delta(s - s(\mathbf{R}_i))}{\sum_i^N w_i^{\mathrm{M}}(\mathbf{R}_i)\, \delta(s - s(\mathbf{R}_i))}. \tag{7}$$

Here, $w_i^{\mathrm{M}}(\mathbf{R}_i) = \exp[\beta(V(s(\mathbf{R}_i), t) - c(t))]$ represents the metadynamics reweighting factor for the ith configuration, while $w_i^{\mathrm{P}}(\mathbf{R}_i) = \exp[\beta(U_{\mathrm{HL}}(\mathbf{R}_i) - U_{\mathrm{LL}}(\mathbf{R}_i))]$ denotes the perturbative weight based on the energy difference between the high-level (U_{HL}) and low-level (U_{LL}) Hamiltonians.

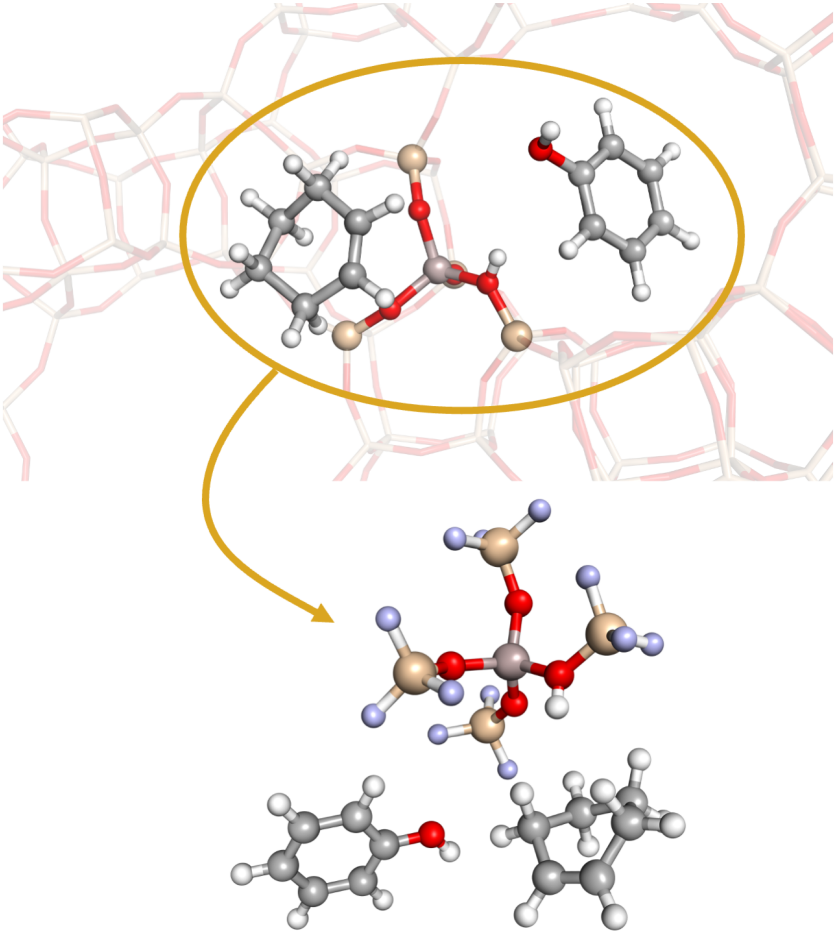
To further assess the quality of the FES obtained from MetaFEP—which ultimately depends on the representativeness of the configurations sampled for the perturbative correction—we employed the concept of reweighting entropy, $\mathcal{S}(s)$, following the approach of Li *et al.*[11]. The reweighting entropy provides a quantitative measure of the effective sampling quality within each CV bin and is defined as

$$\mathcal{S}(s) = -\frac{1}{\ln N_s} \sum_{i \in B_s} \mathcal{P}_i(s) \ln \mathcal{P}_i(s), \tag{8}$$

where

$$\mathcal{P}_i(s) = \frac{w_i^{\mathrm{M}}(\mathbf{R}_i)\, w_i^{\mathrm{P}}(\mathbf{R}_i)\, \delta\big(s - s(\mathbf{R}_i)\big)}{\sum_{j=1}^{N} w_j^{\mathrm{M}}(\mathbf{R}_j)\, w_j^{\mathrm{P}}(\mathbf{R}_j)\, \delta\big(s - s(\mathbf{R}_j)\big)}. \tag{9}$$

The quantity $S(s)$ provides a quantitative assessment of how well each region of the CV space is sampled and thus guides the selection of configurations for MetaFEP. Higher values of $\mathcal{S}(s)$ indicate more uniform and statistically reliable sampling within a given CV bin, while low values suggest that only a few configurations dominate the reweighting, leading to poorer convergence of the FEP correction. Although, $\mathcal{S}(s)$ heavily relies on the overlap of the sampled Hamiltonian and target Hamiltonian but this measure can be exploited iteratively to improve the sampling efficiency and optimize the MetaFEP procedure. Our protocol began by selecting an initial set of N uniformly distributed configurations, for which single point energies at a higher level of theory were calculated. The reweighting entropy distribution, $S(s)$, was then evaluated to assess whether the chosen sampling was sufficient. Regions of the CV space exhibiting $S(s)$ values below a certain threshold were identified, and additional configurations were selectively incorporated from these poorly sampled regions. In practice, the number of additional configurations to take from these regions were determined by a scaling scheme (mentioned in Supplementary Note 7), which gradually increases sampling in low $S(s)$ regions while avoiding oversampling in well-converged regions. By iteratively refining the configuration set in this way, we ensured that the final ensemble was both compact and statistically robust for accurate free energy perturbation.



Supplementary Figure 9: Schematic illustration of the QM:QM scheme. A selected region of the zeolite framework and the reacting substrates (shown in ball-and-stick representation) is treated at a higher level of theory, while the remaining periodic framework (shown as transparent sticks) is described at a lower level of theory. For clarity the linker H atoms are shown in light blue

Supplementary Note 6: Reweighting Entropy-Based Reference Set Improvement for MetaFEP

For the MetaFEP correction, we focus exclusively on the main C-alkylation pathway and omit the formation of surface alkoxy species. This enables the definition of a single collective variable (CV) (see Supplementary Fig. 8) to describe the reaction, which reduces the number of required high-level calculations. PBE+D3 MLIP-WTMetaD simulations using this 1D CV are then employed to explore the main alkylation pathway more thoroughly. The resulting 1D bias reconstructs the FES more smoothly and improves statistical convergence, while retaining the correct shape of the FES compared to the 2D surfaces, with deviations of 0–10 kJ mol^{-1}. Once the MLIP WTMetaD converges, we start with 1000 uniformly distributed configurations from the converged region of the MLIP WTMetaD trajectories. Single point energies are obtained using the domain-based local pair natural orbital (DLPNO) approximation (tightPNO settings) with MP2 and CCSD(T).[12,13] Further, we extrapolate the MP2 and CCSD(T) energies to the complete basis set (CBS) limit using the cc-pVXZ (X = T, Q) basis sets[14,15] in a two-point extrapolation scheme,[16,17] assuming an exponential form for the Hartree Fock (HF) energy,[18,19] with a coefficient in the exponent of 1.5,[17] and a power form for the correlation energy,[20,21] with a X^{-3} behavior.[16] Specifically, we employ the following equations:

$$E_{\mathrm{HF}}^{X} = E_{\mathrm{HF}}^{\mathrm{CBS}} + b_{\mathrm{HF}}\, e^{-1.5\,X} \tag{10}$$

and

$$E_{\mathrm{cor}}^{X} = E_{\mathrm{cor}}^{\mathrm{CBS}} + b_{\mathrm{cor}}\, X^{-3}, \tag{11}$$

where E_{HF}^{X} and E_{cor}^{X} are the Hartree–Fock and correlation energies, respectively, obtained with a given cc-pVXZ basis set. Here, X denotes the cardinal number of the basis set, e.g., $X = 2$ for double-ζ and $X = 3$ for triple-ζ. The quantities $E_{\mathrm{HF}}^{\mathrm{CBS}}$ and $E_{\mathrm{cor}}^{\mathrm{CBS}}$ are the Hartree–Fock and correlation energies, respectively, extrapolated to the complete basis set (CBS) limit. The system- and structure-specific parameters b_{HF} and b_{cor} are determined from two single point calculations at different values of X.

Using the MP2-CBS energies, we computed the reweighting entropy distribution, $S(s)$. As shown in Supplementary Figure 9 and Supplementary Figure 10, the initial $S(s)$ profiles exhibited maxima around 0.6, which we adopted as the threshold for further refinement. Configurations were therefore added selectively from bins with $S(s) < 0.6$ using the scaling scheme described below

$$n_{\mathrm{points}}(s) = \begin{cases} 0, & \mathcal{S}(s) \geq \text{threshold} \\ \max\left(\text{min\_points},\, \text{max\_points} \cdot \left(1 - \frac{\mathcal{S}(s)}{\text{threshold}}\right)\right), & \mathcal{S}(s) < \text{threshold} \end{cases}$$

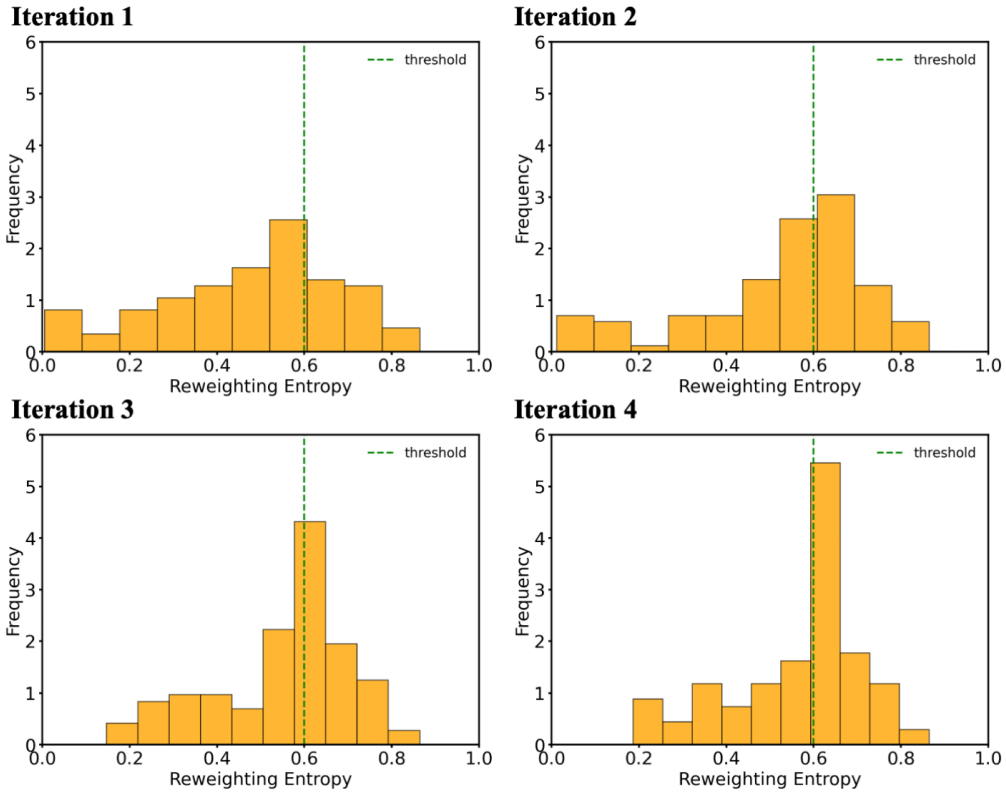
Here, max_points and min_points represent the maximum and minimum number of configurations to select from a bin, respectively, and threshold is the reweighting entropy value below which additional points are sampled. This iterative procedure ensures that the reference set covers the relevant CV space adequately while maintaining computational efficiency.

After convergence of MLIP-WTMetaD, we initially selected 1000 uniformly distributed configurations from the converged region of the trajectories. Single-point energies were then evaluated at the MP2 level with DZVP and TZVP basis sets in the above mentioned QM:QM scheme, and subsequently combined to estimate CBS energies as outlined above. Using the MP2-CBS energies, we computed the reweighting entropy distribution, $S(s)$. As shown in Supplementary Figure 10 and Supplementary Figure 11, the initial $S(s)$ profiles exhibited maxima around 0.6, which we adopted as the threshold for further refinement. Configurations were therefore added selectively from bins with $S(s) < 0.6$ using the scaling scheme described in the methodology, and their MP2-DZVP and MP2-TZVP energies were evaluated to update the dataset. After each augmentation, the $S(s)$ distribution was recalculated to monitor convergence. For benzene alkylation, the fraction of bins with $S(s) > 0.6$ increased from 29% in the initial sampling to 57% after three refinement cycles, while for phenol it increased from 39% to 70% in three refinement cycles (Supplementary Table 5, Supplementary Figure 10 and Supplementary Figure 11). These reductions demonstrate that the adaptive sampling scheme systematically

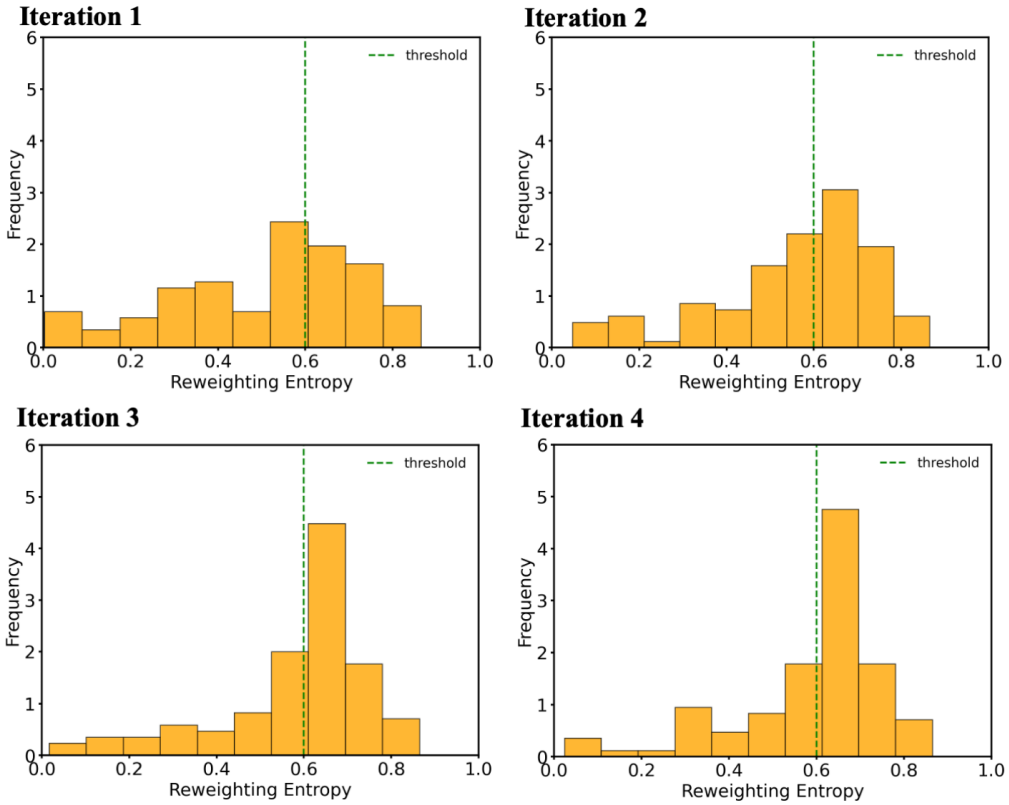
improves the reweighting entropy distribution, thereby yielding statistically more reliable ensembles for MetaFEP analysis. The final refined configuration sets were then employed to construct the perturbed FES, correcting the MLIP (PBE+D3) trajectories to the MP2-CBS level and CCSD(T)-CBS level.

Supplementary Table 5 MetaFEP sampled points summary.

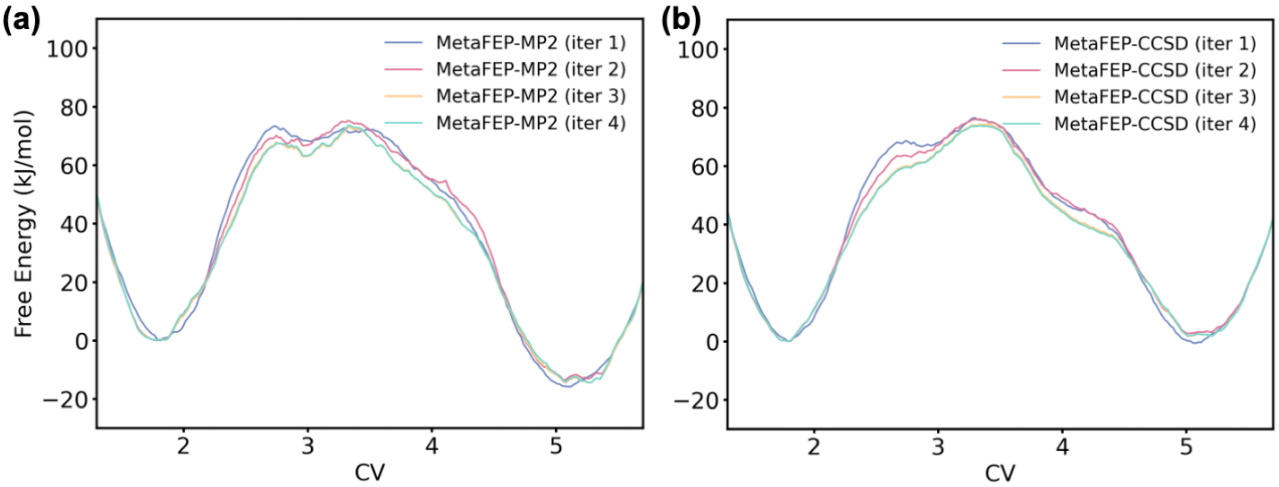
	Benzene alkylation		**Phenol alkylation**	
Iteration	**Total points**	**Above threshold (%)**	**Total points**	**Above threshold (%)**
1	1000	29	1000	39
2	3406	49	2722	54
3	5572	52	4144	64
4	6949	57	5307	70



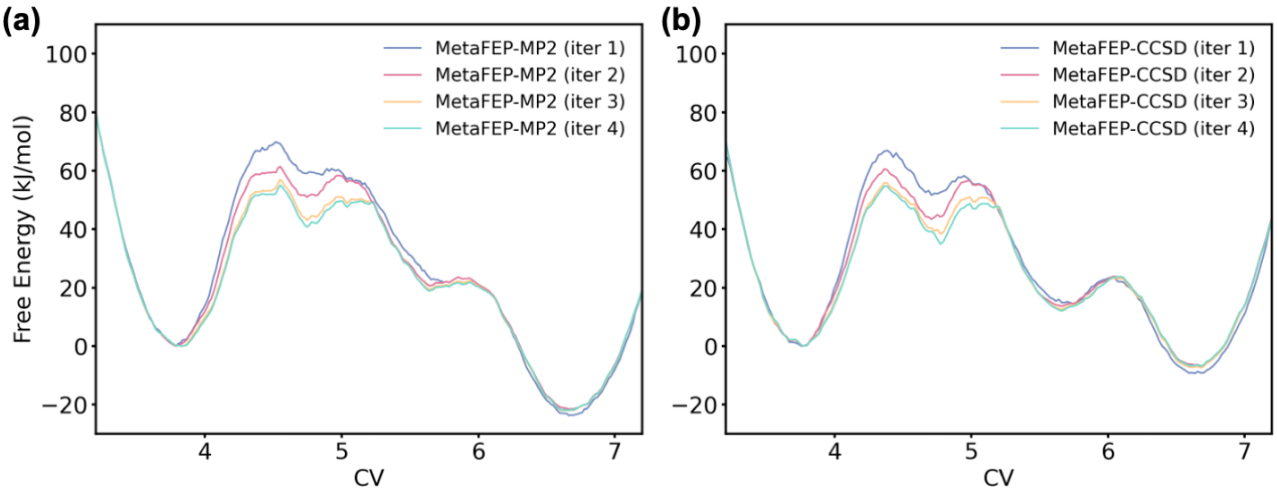
Supplementary Figure:10 Improvement of reweighting entropy for benzene alkylation



Supplementary Figure 11: Improvement of reweighting entropy for phenol alkylation



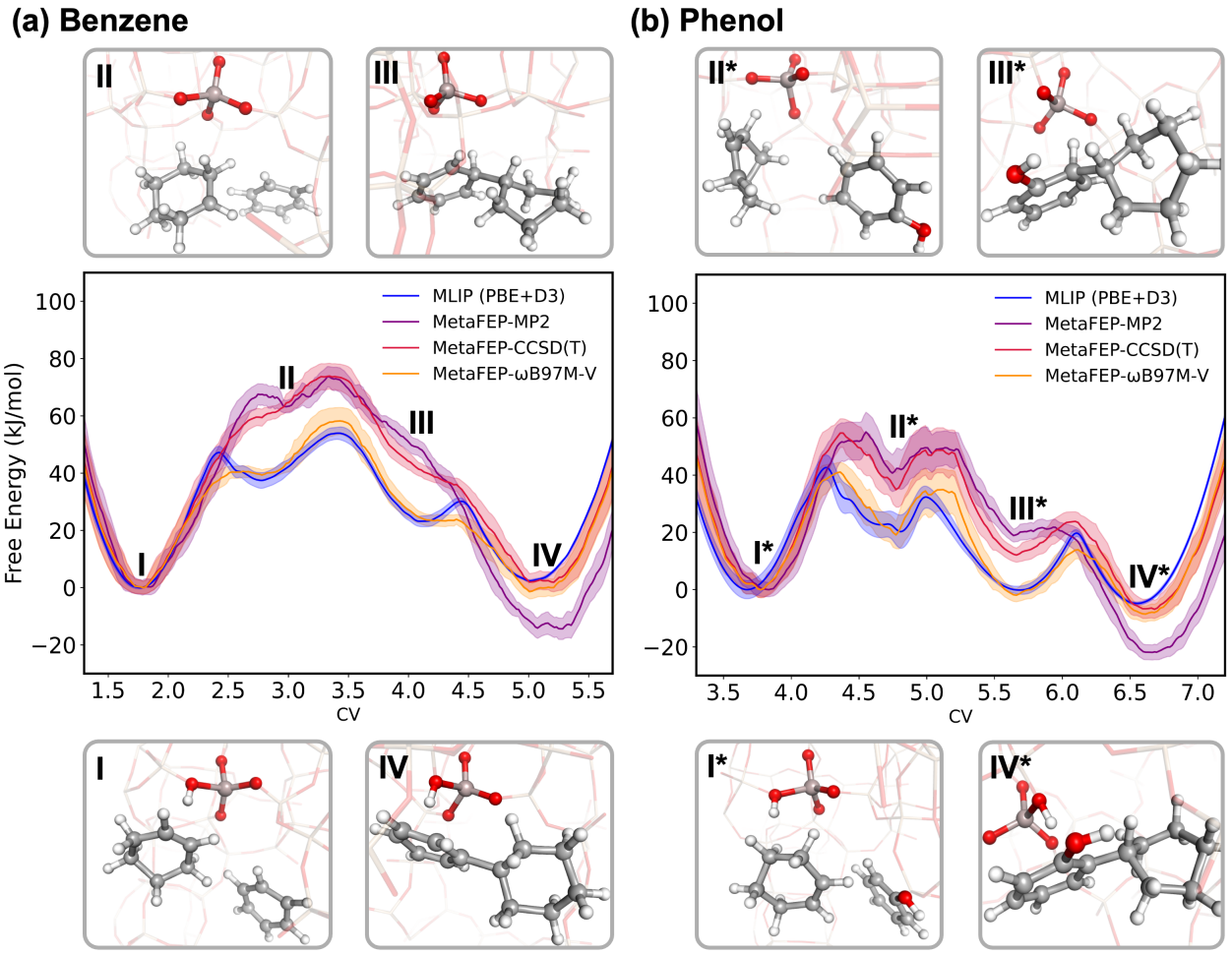
Supplementary Figure 12: (a) Convergence of the FES obtained from MetaFEP for benzene alkylation with the progressive addition of reference configurations selected based on reweighting entropy at the MP2 level. (b) Corresponding evolution of the FES at the CCSD(T) level.



Supplementary Figure 13: (a) Convergence of the FES obtained from MetaFEP for phenol alkylation with the progressive addition of reference configurations selected based on reweighting entropy at the MP2 level. (b) Corresponding evolution of the FES at the CCSD(T) level.

Supplementary Note 7: Comparison of PBE, ωB97M-V, MP2 and CCSD(T) FES

Supplementary Fig. 14 compares the ωB97M-V-, MP2-, and CCSD(T)-quality FESs with the PBE+D3 FES for the benzene and phenol reactions. For both alkylation reactions, ωB97M-V does not significantly improve over PBE+D3, which is surprising given its generally strong performance across diverse benchmarks. While the intermediate stabilities remain similar to PBE+D3, the barriers for the initial protonation and the final deprotonation are even slightly lower. In the case of benzene, this converts



Supplementary Figure 14:FES comparision between MLIP(PBE+D3) and MetFEP (ωB97M-V,MP2 and CCSD(T)) for a) Benzene alkylation and b) Phenol Alkylation at 573K

metastable intermediates into transient species. The formation of such transient species is consistent with the MP2 and CCSD(T) results. However, in contrast to ωB97M-V, the two post-HF methods yield significantly higher barriers and a pronounced destabilization of the charged intermediates relative to PBE+D3 (about +30 kJ mol^{-1} for the carbenium ion), an expected effect given the known limitations of GGA functionals. The MP2 and CCSD(T) FESs are well aligned along the reactive part of the reaction coordinate, although MP2 overstabilizes the formation of the final products relative to the CCSD(T) reference. Consequently, for the overall reaction free energy, comparing products and reactants, both PBE+D3 and, in particular, ωB97M-V agree better with CCSD(T) than MP2.

A notable difference between the benzene and phenol free energy surfaces concerns the stability of the two cationic intermediates. For benzene alkylation, the carbenium and Wheland intermediates appear as shallow minima at the PBE+D3 level but become destabilized and thus transient at the MP2 and CCSD(T) levels. In contrast, for phenol alkylation the cationic species remain shallow intermediates with MP2 and CCSD(T), reflecting the deeper FES wells already obtained with the

PBE+D3 quality MLIPs. Going beyond the GGA functional PBE+D3, and also beyond the range-separated meta-hybrid functional ωB97M-V, to MP2 and CCSD(T) is therefore required to correctly capture the differences between benzene and phenol.

More broadly, it underscores the importance of reaching chemically accurate post-HF-quality FESs to reveal such subtle effects and to describe reaction steps even qualitatively correctly.

Supplementary Table 6 Gas-Phase reaction energy at different level of theory

Level of Theory	**Benzene Alkylation**	**Phenol Alkylation**
MLIP	-4.63	-41.08
PBE	-78.38	-101.58
MP2	-96.08	-121.0
CCSD(T)	-80.92	-105.16

References

[1] Baerlocher, C., Brouwer, D., Marler, B. & McCusker, L. B. Database of zeolite structures. `https://www.iza-structure.org/databases/`.

[2] Perego, S. & Bonati, L. Data efficient machine learning potentials for modeling catalytic reactivity via active learning and enhanced sampling. *npj Computational Materials* **10**, 291 (2024).

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

[4] Vandermause, J. *et al.* On-the-fly active learning of interpretable bayesian force fields for atomistic rare events. *npj Computational Materials* **6**, 20 (2020).

[5] Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Physical Review B—Condensed Matter and Materials Physics* **87**, 184115 (2013).

[6] Zhang, H., Juraskova, V. & Duarte, F. Modelling chemical processes in explicit solvents with machine learning potentials. *Nature Communications* **15**, 6114 (2024).

[7] Himanen, L. *et al.* DScribe: Library of descriptors for machine learning in materials science. *Computer Physics Communications* **247**, 106949 (2020). URL `https://doi.org/10.1016/j.cpc.2019.106949`.

[8] Laio, A. & Parrinello, M. Escaping free-energy minima. *Proceedings of the national academy of sciences* **99**, 12562–12566 (2002).

[9] Grifoni, E., Piccini, G. & Parrinello, M. Microscopic description of acid–base equilibrium. *Proceedings of the National Academy of Sciences* **116**, 4054–4057 (2019).

[10] Piccini, G. & Parrinello, M. Accurate quantum chemical free energies at affordable cost. *The journal of physical chemistry letters* **10**, 3727–3731 (2019).

[11] Wang, M. *et al.* Efficient strategy for the calculation of solvation free energies in water and chloroform at the quantum mechanical/molecular mechanical level. *Journal of Chemical Information and Modeling* **57**, 2476–2489 (2017).

[12] Riplinger, C., Pinski, P., Becker, U., Valeev, E. F. & Neese, F. Sparse maps—a systematic infrastructure for reduced-scaling electronic structure methods. ii. linear scaling domain based pair natural orbital coupled cluster theory. *The Journal of chemical physics* **144** (2016).

[13] Pavošević, F., Pinski, P., Riplinger, C., Neese, F. & Valeev, E. F. Sparsemaps—a systematic infrastructure for reduced-scaling electronic structure methods. iv. linear-scaling second-order explicitly correlated energy with pair natural orbitals. *The Journal of Chemical Physics* **144** (2016).

[14] Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *The Journal of chemical physics* **90**, 1007–1023 (1989).

[15] Woon, D. E. & Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. iii. the atoms aluminum through argon. *The Journal of chemical physics* **98**, 1358–1371 (1993).

[16] Helgaker, T., Klopper, W., Koch, H. & Noga, J. Basis-set convergence of correlated calculations on water. *The Journal of chemical physics* **106**, 9639–9646 (1997).

[17] Jensen, F. Estimating the hartree—fock limit from finite basis set calculations. *Theoretical Chemistry Accounts* **113**, 267–273 (2005).

[18] Feller, D. Application of systematic sequences of wave functions to the water dimer. *The Journal of chemical physics* **96**, 6104–6114 (1992).

[19] Feller, D. The use of systematic sequences of wave functions for estimating the complete basis set, full configuration interaction limit in water. *The Journal of chemical physics* **98**, 7059–7071 (1993).

[20] Schwartz, C. Importance of angular correlations between atomic electrons. *Physical Review* **126**, 1015 (1962).

[21] Schwartz, C. Estimating convergence rates of variational calculations. *Methods in Computational Physics* **2**, 241–266 (1963).