Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Reference-Quality Free Energy Barriers in Catalysis from Machine Learning Thermodynamic Perturbation Theory

J. Rey, C. Chizallet, D. Rocca, T. Bučko*, M. Badawi**

SI. METHODOLOGY

A. Structural model

In our calculations, a structural model of chabazite (CHA, symmetry group R$\bar{3}$m) was used with the supercell defined by lattice vectors $\boldsymbol{a}'_1$, $\boldsymbol{a}'_2$ and $\boldsymbol{a}'_3$ obtained from those of the relaxed rhombohedral primitive cell (lattice constants $a = 9.336$ Å and $\alpha = 94.6°$) via the following transformation of the lattice vectors (see Fig. S1): $\boldsymbol{a}'_1 = \boldsymbol{a}_2 + \boldsymbol{a}_3$, $\boldsymbol{a}'_1 = \boldsymbol{a}_2 + \boldsymbol{a}_3$, and $\boldsymbol{a}'_1 = \boldsymbol{a}_2 + \boldsymbol{a}_3$. The shortest interatomic separation between the atoms in hydrocarbon and the atoms in its periodically repeated images in the supercell was at least 5.5 Å. In the cracking reaction starting from acid zeolite, the Brønsted acid site was initially located on the framework oxygen atom O1, which is, according to experiment [S1], one of the two most populated proton sittings in CHA (Fig. S1). For further details on the structural model we refer to the original publications reporting the production method data [S2–S4].

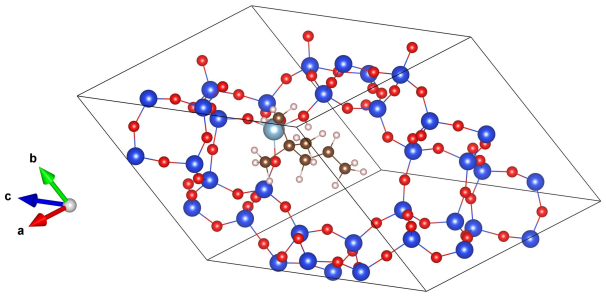

FIG. S1. Snapshot of the supercell of acid chabazite with the π-complex, reactant for the cracking reaction. Color code: Si in blue, O in red, Al in light blue, C in brown.

B. Choice of training configurations

In the MLPT calculations with the RPA target method, 62 configurations evenly distributed along the molecular dynamics trajectories were selected for each stationary state of the reactions under study and their electronic energies $\tilde{V}(\boldsymbol{q})$ were computed at the RPA level of theory. The huge computational cost and memory requirement of these calculations (around one month of calculation was required for each stationary state, on one fat node with 3 TB RAM) prevented us to consider even larger training sets. Training sets of increasing size (12, 25, 40, 50, 62 configurations) were chosen to train several machine learning models reproducing the values of ΔV. The remaining configurations were used as a test set to estimate the error of the ML prediction.[S5] For the largest training set of 62 configurations, a leave one out procedure was used for the error estimation.[S6] Mean errors and root mean square errors on $\tilde{V}(\boldsymbol{q})$, used as metrics in our error evaluations, are compiled in Tables S1 to S3 and S14 to S16. It is important to note that this test set allows to estimate the systematic error Δ_ϵ on ΔA_{TS} and ΔA_R due to the random error of the ML model, as shown by Bučko et al.,[S5] and to ensure that the convergence of the MLPT correction with respect to the size of the training set is achieved.

As the MLPT correction of free energies showed some variability with the particular selection of the training sets of 12 to 50 configurations chosen among the total of 62 configurations, 100 random choices were performed and the MLPT corrections were averaged, showing an excellent convergence of the free energy correction with respect to the size of the training set (see later Tables S6 to S13). Standard errors were estimated from the standard deviation of these series of results. Our best estimate, with estimation of the systematic error of MLPT, was obtained with a leave out one strategy[S6] on the largest training set.

For the other, computationally less expensive, levels of theory, 225 configurations evenly distributed along the molecular dynamics trajectories were selected. This set was split into training and test sets of 200 and 25 configurations, respectively.

C. Error sources on the free energy of activation

Statistical uncertainty
The first source of error on the free energy of activation is the statistical uncertainty on the free energy gradient in production method (PBE+D2) simulations. Since the free energy was computed by integrating free energy gradients along the reaction coordinate, the error tends to accumulate along the transformation path. In our previous work,[S2, S4] where the production method calculations were reported, the standard error in free energy barriers was estimated to be around 4 kJ/mol when the confidence level of 95% was used (3.6 and 2.2 kJ/mol respectively for the forward and reverse reaction of the isomerization reaction[S2] and 4.2 kJ/mol for the cracking reaction[S4]).

Systematic error due to ML The second source of error is the systematic error related to the ML predictions. This systematic error Δ_ϵ on the free energy of reactant or transition state can be estimated using the data from the test sets and correction formula obtained from the second order cumulant approximation of free energy.[S5]. Δ_ϵ is estimated to be around -2 kJ/mol for each stationary state and since the reactant and transition state contributions to the free energy of activation energy are of opposite sign (see eq. 1 in the main text), their contributions tend to cancel to a large extend. The corrected values ΔA_{corr} of the free energy determined at the RPA level, defined as follows:

$$\Delta A_{corr} = \Delta A - \Delta_\epsilon \tag{S1}$$

are reported for stationary state of all three reactions considered in this work in Tables S6-S13.

Systematic error of free energy perturbation theory
The use of thermodynamic perturbation theory (PT) itself can introduce a significant systematic error if the configurational spaces sampled in the simulations at the production and target levels do not have a sufficient overlap.[S7] An index denoted I_w has been introduced as indicator of this problem.[S8] The I_w gives the fraction of configurations contributing to one half of the total statistical weight, after the reweighting used in eq. 2 and 3 in the main text. I_w varies between 0 (no overlap) to 0.5 (perfect overlap). For the PBE+D2 production and RPA target methods, the I_w index values are around 0.1 (see Tables S6 to S8, S11 and S12), which indicates that the use of PT is reliable.[S8] For GGA (PBE+MBD, PBE+MBD-FI) and rev-vdW-DF2 functionals, the configuration spaces overlaps are also significant ($I_w \approx 0.2$), but for hybrid functionals (HSE06 and B3LYP), a too low value of I_w ($\leq 1.3 \times 10^{-3}$) implies that the large corrections predicted by the application of MLPT are unreliable (see Tables S1 to S3, S14 and S15). A possible solution to this problem is application of a Monte Carlo resampling based on MLPT,[S8] which we, however, do not perform here.

SII. ISOMERIZATION REACTION BETWEEN TERTIARY DIBRANCHED AND TRIBRANCHED CATIONS

A. Effect of the level of theory on the barriers

In this section we provide information about the effect of the level of theory on the free energy barrier computed by MLPT and on the electronic energy barrier computed in the static approach.

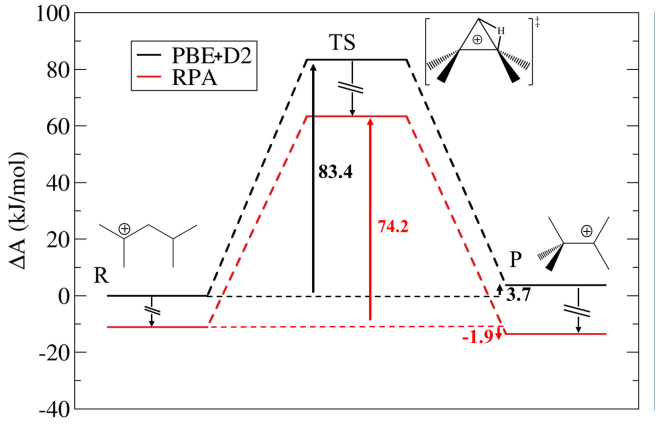

FIG. S2. Free energy diagram computed by MLPT at the PBE+D2 and RPA levels of theory for the isomerization reaction from a tertiary dibranched to a tertiary tribranched cation.

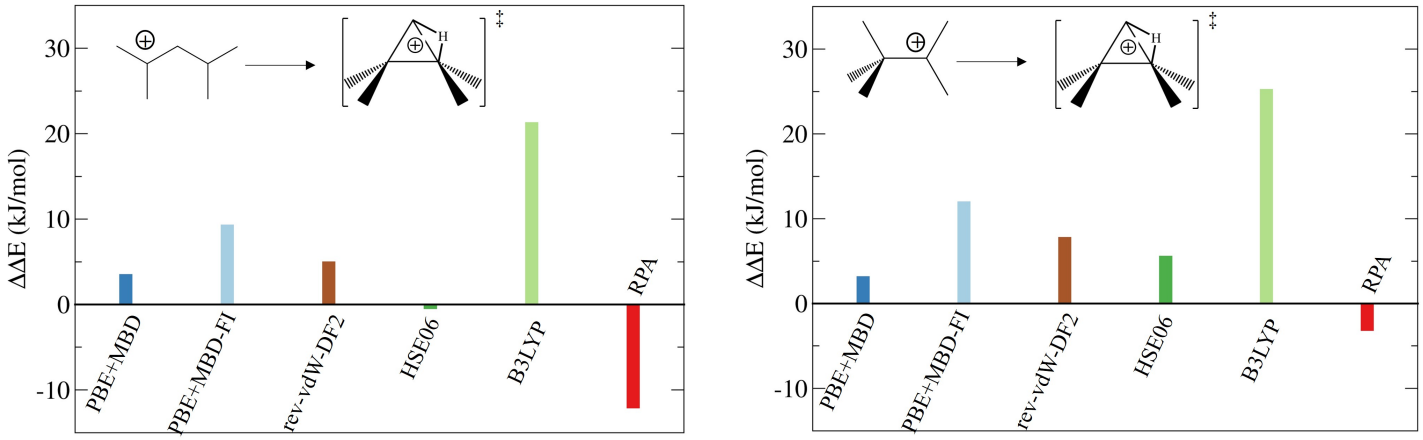

FIG. S3. Effect of the level of theory on the electronic energy barrier of the isomerization reaction between tertiary dibranched and tribranched cations computed in the static approach (forward step on the left, reverse step on the right). The difference $\Delta\Delta E$ is computed with respect to the electronic contribution to the activation energy obtained at the PBE+D2 level.

TABLE S1. MLPT results at different levels of theory (LOT) for the dibranched reactant of the isomerization reaction. The mean error (ME), mean absolute error (MAE), root mean square error (RMSE), the MLPT result ΔA, the correction to this result $\Delta\epsilon$ and the I_w index are provided.

LOT	ME (kJ/mol)	MAE (kJ/mol)	RMSE (kJ/mol)	ΔA (kJ/mol)	$\Delta\epsilon$ (kJ/mol)	I_w
PBE+MBD	0.11	0.64	0.76	-128.23	-0.06	0.19
PBE+MBD-FI	-0.06	0.63	0.80	600.14	-0.02	0.08
rev-vdW-DF2	0.06	0.69	0.89	10279.61	-0.07	0.19
HSE06	0.24	2.07	2.57	-13778.59	-2.99	0.00
B3LYP	0.24	2.40	3.03	-12542.33	-4.08	0.00

TABLE S2. MLPT results at different levels of theory (LOT) for the transition state of the isomerization reaction. The mean error (ME), mean absolute error (MAE), root mean square error (RMSE), the MLPT result ΔA, the correction to this result $\Delta\epsilon$ and the I_w index are provided.

LOT	ME (kJ/mol)	MAE (kJ/mol)	RMSE (kJ/mol)	ΔA (kJ/mol)	$\Delta\epsilon$ (kJ/mol)	I_w
PBE+MBD	0.02	0.73	0.88	-124.74	-0.18	0.20
PBE+MBD-FI	0.14	0.65	0.74	606.00	-0.18	0.09
rev-vdW-DF2	0.09	0.69	0.83	10285.75	0.03	0.22
HSE06	0.21	2.51	3.18	-13776.78	-5.85	0.00
B3LYP	0.17	2.77	3.58	-12520.44	-5.60	0.00

TABLE S3. MLPT results at different levels of theory (LOT) for the tribranched product of the isomerization reaction. The mean error (ME), mean absolute error (MAE), root mean square error (RMSE), the MLPT result ΔA, the correction to this result $\Delta\epsilon$ and the I_w index are provided.

LOT	ME (kJ/mol)	MAE (kJ/mol)	RMSE (kJ/mol)	ΔA (kJ/mol)	$\Delta\epsilon$ (kJ/mol)	I_w
PBE+MBD	-0.08	0.52	0.67	-125.31	-0.16	0.21
PBE+MBD-FI	-0.01	0.45	0.56	606.00	0.11	0.08
rev-vdW-DF2	-0.09	0.64	0.78	10280.19	-0.12	0.22
HSE06	-0.07	1.88	2.21	-13772.73	-4.68	0.00
B3LYP	-0.23	2.31	2.66	-12531.02	-3.11	0.00

TABLE S4. MLPT results at different levels of theory (LOT) for the free energy barrier of the forward isomerisation reaction (from the dibranched reactant to the transition state). The MLPT correction on the barrier, $\Delta\Delta A = \Delta A(TS) - \Delta A(R)$ and its value after correction are reported, as well as the corrected free energy barrier. The free energy barrier computed at the PBE+D2 level was 83.40 kJ/mol.

LOT	$\Delta\Delta A$ (kJ/mol)	$\Delta\Delta A_{corr}$ (kJ/mol)	ΔA (kJ/mol)
PBE+MBD	3.49	3.37	86.77
PBE+MBD-FI	5.86	5.70	89.10
rev-vdW-DF2	6.14	6.24	89.64
HSE06	1.81	-1.05	82.35
B3LYP	21.90	20.37	103.77

TABLE S5. MLPT results at different levels of theory (LOT) for the free energy barrier of the backward isomerisation reaction (from the tribranched product to the transition state). The MLPT correction on the barrier, $\Delta\Delta A = \Delta A(TS) - \Delta A(P)$ and its value after correction are reported, as well as the corrected free energy barrier. The free energy barrier computed at the PBE+D2 level was 79.70 kJ/mol.

LOT	$\Delta\Delta A$ (kJ/mol)	$\Delta\Delta A_{corr}$ (kJ/mol)	ΔA (kJ/mol)
PBE+MBD	0.57	0.55	80.25
PBE+MBD-FI	0.00	-0.28	79.42
rev-vdW-DF2	5.56	5.72	85.42
HSE06	-4.05	-5.22	74.48
B3LYP	10.58	8.09	87.79

B. MLPT results at the RPA level of theory

TABLE S6. MLPT results at the RPA level of theory for the reactant of the isomerization reaction. As a function of N_{train}, the number of configurations in the training set, are reported: the MLPT result ΔA and its standard deviation $\sigma(\Delta A)$, the corrected MLPT result ΔA_{corr}, and its standard deviation $\sigma(\Delta A_{corr})$, the I_w index and its standard deviation $\sigma(I_w)$.

N_{train}	ΔA (kJ/mol)	$\sigma(\Delta A)$ (kJ/mol)	ΔA_{corr} (kJ/mol)	$\sigma(\Delta A_{corr})$ (kJ/mol)	I_w	$\sigma(I_w)$
12	-124814.95	1.70	-124817.38	0.49	0.23	0.04
25	-124815.35	0.84	-124817.30	0.64	0.18	0.03
40	-124815.47	0.52	-124817.34	0.89	0.17	0.01
50	-124815.56	0.28	-124815.56	1.49	0.16	0.01
62	-124815.68	-	-124815.68	-	0.15	-

TABLE S7. MLPT results at the RPA level of theory for the transition state of the isomerization reaction.

N_{train}	ΔA (kJ/mol)	$\sigma(\Delta A)$ (kJ/mol)	ΔA_{corr} (kJ/mol)	$\sigma(\Delta A_{corr})$ (kJ/mol)	I_w	$\sigma(I_w)$
12	-124824.17	1.57	-3.02	1.95	0.21	0.04
25	-124824.49	1.05	-2.43	1.73	0.18	0.02
40	-124824.60	0.55	-2.00	1.54	0.17	0.01
50	-124824.53	0.33	-1.97	1.81	0.16	0.01
62	-124824.57	-	-2.12	-	0.16	-

TABLE S8. MLPT results at the RPA level of theory for the product of the isomerization reaction.

N_{train}	ΔA (kJ/mol)	$\sigma(\Delta A)$ (kJ/mol)	ΔA_{corr} (kJ/mol)	$\sigma(\Delta A_{corr})$ (kJ/mol)	I_w	$\sigma(I_w)$
12	-124821.44	1.74	-124821.44	0.55	0.21	0.05
25	-124821.89	0.87	-124821.89	0.66	0.17	0.03
40	-124821.83	0.48	-124821.83	1.02	0.16	0.01
50	-124821.85	0.34	-124821.85	1.51	0.15	0.01
62	-124821.77	-	-124821.77	-	0.15	-

TABLE S9. MLPT results at the RPA level of theory for the free energy barrier of the isomerization reaction (forward reaction). The MLPT correction on the barrier, $\Delta\Delta A = \Delta A(TS) - \Delta A(R)$ and its standard deviation σ are reported as a function of N_{train}, the number of configurations in the training set. The corrected free energy barrier ΔA is also given. The free energy barrier computed at the PBE+D2 level was 83.40 kJ/mol.

N_{train}	$\Delta\Delta A$ (kJ/mol)	$\sigma(\Delta\Delta A)$ (kJ/mol)	$\Delta\Delta A_{corr}$ (kJ/mol)	$\sigma(\Delta\Delta A_{corr})$ (kJ/mol)	ΔA (kJ/mol)
12	-9.22	2.31	-9.81	2.38	73.59
25	-9.13	1.34	-9.62	1.50	73.78
40	-9.13	0.75	-9.27	1.27	74.13
50	-8.97	0.43	-9.03	1.57	74.37
62	-8.89	-	-9.17	-	74.23

TABLE S10. MLPT results at the RPA level of theory for the free energy barrier of the isomerization reaction (backward reaction). The MLPT correction on the barrier, $\Delta\Delta A = \Delta A(TS) - \Delta A(P)$ and its standard deviation σ are reported as a function of N_{train}, the number of configurations in the training set. The corrected free energy barrier ΔA is also given. The free energy barrier computed at the PBE+D2 level was 79.70 kJ/mol.

N_{train}	$\Delta\Delta A$ (kJ/mol)	$\sigma(\Delta\Delta A)$ (kJ/mol)	$\Delta\Delta A_{corr}$ (kJ/mol)	$\sigma(\Delta\Delta A_{corr})$ (kJ/mol)	ΔA (kJ/mol)
12	-2.73	2.34	-3.44	0.77	76.26
25	-2.60	1.36	-3.51	1.07	76.19
40	-2.78	0.73	-3.38	1.49	76.32
50	-2.68	0.48	-3.39	2.15	76.31
62	-2.79	-	-3.55	-	76.15

SIII. MLPT RESULTS FOR THE TYPE B1 CRACKING REACTION

A. MLPT results at the RPA level of theory

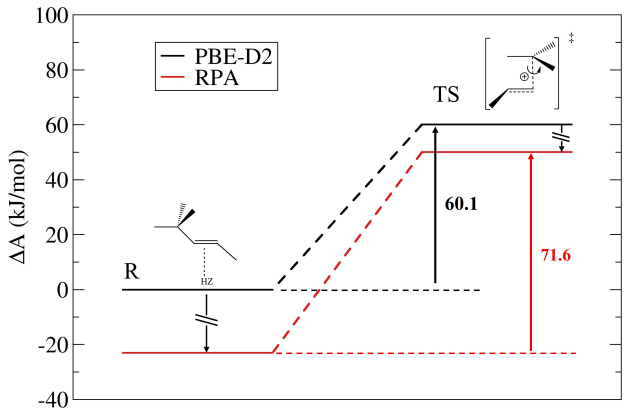

FIG. S4. Free energy diagram computed by MLPT at the RPA level of theory for the type B_1 β-scission

TABLE S11. MLPT results at the RPA level of theory for the reactant of the cracking reaction. As a function of N_{train}, the number of configurations in the training set, are reported: the MLPT result ΔA and its standard deviation $\sigma(\Delta A)$, the corrected MLPT result ΔA_{corr}, and its standard deviation $\sigma(\Delta A_{corr})$, the I_w index and its standard deviation $\sigma(I_w)$.

N_{train}	ΔA (kJ/mol)	$\sigma(\Delta A)$ (kJ/mol)	ΔA_{corr} (kJ/mol)	$\sigma(\Delta A_{corr})$ (kJ/mol)	I_w	$\sigma(I_w)$
12	-124834.55	1.44	-124837.83	0.50	0.16	0.04
25	-124835.50	1.17	-124837.78	0.93	0.12	0.02
40	-124836.04	0.65	-124837.91	1.20	0.11	0.01
50	-124836.27	0.47	-124836.27	1.92	0.11	0.01
62	-124836.49	-	-124836.49	-	0.11	-

TABLE S12. MLPT results at the RPA level of theory for the transition state of the cracking reaction.

N_{train}	ΔA (kJ/mol)	$\sigma(\Delta A)$ (kJ/mol)	ΔA_{corr} (kJ/mol)	$\sigma(\Delta A_{corr})$ (kJ/mol)	I_w	$\sigma(I_w)$
12	-124822.05	2.07	-124827.29	0.88	0.16	0.04
25	-124822.63	1.25	-124826.84	1.34	0.12	0.03
40	-124823.12	0.74	-124826.76	1.79	0.10	0.02
50	-124823.73	0.49	-124823.73	2.02	0.08	0.01
62	-124823.45	-	-124823.45	-	0.08	-

TABLE S13. MLPT results at the RPA level of theory for the free energy barrier of the cracking reaction. The MLPT correction on the barrier, $\Delta\Delta A = \Delta A(TS) - \Delta A(R)$ and its standard deviation σ are reported as a function of N_{train}, the number of configurations in the training set. The corrected free energy barrier ΔA is also given. The free energy barrier computed at the PBE+D2 level was 60.10 kJ/mol.

N_{train}	$\Delta\Delta A$ (kJ/mol)	$\sigma(\Delta\Delta A)$ (kJ/mol)	$\Delta\Delta A_{corr}$ (kJ/mol)	$\sigma(\Delta\Delta A_{corr})$ (kJ/mol)	ΔA (kJ/mol)
12	12.51	2.53	10.54	1.01	70.64
25	12.87	1.72	10.94	1.64	71.04
40	12.92	0.99	11.15	2.15	71.25
50	12.53	0.68	11.99	2.79	72.09
62	13.04	-	11.51	-	71.61

B. Effect of the level of theory on the barriers of the type B1 cracking reaction

TABLE S14. MLPT results at different levels of theory (LOT) for the reactant of the cracking reaction. The mean error (ME), mean absolute error (MAE), root mean square error (RMSE), the MLPT result ΔA, the correction to this result $\Delta\epsilon$ and the I_w index are provided.

LOT	ME (kJ/mol)	MAE (kJ/mol)	RMSE (kJ/mol)	ΔA (kJ/mol)	$\Delta\epsilon$ (kJ/mol)	I_w
PBE+MBD	0.12	0.57	0.75	-128.13	-0.14	0.20
PBE+MBD-FI	0.10	0.63	0.73	-15.92	0.10	0.23
rev-vdW-DF2	0.07	0.70	0.87	10278.77	-0.31	0.18
HSE06	-0.37	2.18	2.89	-13793.51	-1.55	0.00
B3LYP	-0.42	2.43	3.15	-12551.00	-3.04	0.00

TABLE S15. MLPT results at different levels of theory (LOT) for the transition state of the cracking reaction. The mean error (ME), mean absolute error (MAE), root mean square error (RMSE), the MLPT result ΔA, the correction to this result $\Delta\epsilon$ and the I_w index are provided.

LOT	ME (kJ/mol)	MAE (kJ/mol)	RMSE (kJ/mol)	ΔA (kJ/mol)	$\Delta\epsilon$ (kJ/mol)	I_w
PBE+MBD	-0.30	0.68	0.82	-124.15	-0.38	0.19
PBE+MBD-FI	-0.17	0.67	0.81	-14.02	-0.30	0.21
rev-vdW-DF2	-0.30	0.81	0.98	10279.96	-0.77	0.22
HSE06	-0.33	2.21	2.89	-13765.77	-1.55	0.00
B3LYP	-0.46	2.78	3.78	-12542.57	-4.54	0.00

TABLE S16. MLPT results at different levels of theory (LOT) for the free energy barrier of the cracking reaction. The MLPT correction on the barrier, $\Delta\Delta A = \Delta A(TS) - \Delta A(R)$ and its value after correction are reported. The corrected free energy barrier ΔA is also given. The free energy barrier computed at the PBE+D2 level was 60.10 kJ/mol.

LOT	$\Delta\Delta A$ (eV)	$\Delta\Delta A_{corr}$ (eV)	ΔA (kJ/mol)
PBE+MBD	3.98	3.74	63.84
PBE+MBD-FI	1.90	1.50	61.60
rev-vdW-DF2	1.18	0.72	60.82
HSE06	27.74	27.75	87.85
B3LYP	8.43	6.92	67.02

SIV. EXTRACTION OF RELEVANT KINETIC DATA FROM PREVIOUS WORK

In ref.[S9], experiments were performed to investigate the zeolite-catalyzed isomerization and cracking of alkenes containing 7 carbon atoms (through the bifunctional hydroisomerization of n-heptane). A single-event kinetic model was built on the basis of previous ab initio calculations,[S2–S4] and on additional static calculations made on purpose, all at the PBE+D2 level. In the initial kinetic model, the type B isomerization free energy barrier transforming dibranched tertiary cations into the tribranched tertiary cation was extracted from a series of static calculations, and set at 82.7 kJ/mol at 500 K. These static calculations in gas phase were consistent with the data computed by AIMD on some examples. The reverse free energy barrier (tribranched tertiary cation to dibranched tertiary cations) was set to 87.5 kJ/mol. The B1 cracking free energy barrier initially set in the model, enclosing the formation of the secondary carbenium ion (elusive species) from the corresponding π-complex, was determined by AIMD, being equal to 60 kJ/mol (about 45 kJ/mol from the π-complex to the secondary cation, and 15 kJ/mol from the secondary cation to the cracking transition state).

To reproduce accurately the isomerization over cracking selectivity, it turned out that the difference between the two free energy barriers (type B isomerization minus B1 cracking) needed to be adjusted. The choice was made to keep the isomerization barrier unchanged, and to increase the cracking barrier by 15 kJ/mol, to reach 75 kJ/mol. Other adjustments were also made to adapt the value of the activity, see ref.[S9]. This led to a very satisfactory results with respect to experiments, in terms of description of the isomerization/cracking selectivity. However, one must notice that other choices could have been made to lead to a similar agreement. One may have chosen to decrease by 15 kJ/mol the isomerization barriers, keeping the cracking barrier unchanged, and adjusting some other parameters to get the right activity. The isomerization/cracking selectivity is indeed monitored by the free energy barrier difference between the two kinds of events, independently of their absolute value. Thus, the reference data we can extract from experimental selectivities is the following: the free energy barrier differences between type B isomerization and B1 cracking are 7.7 and 12.5 kJ/mol for the forward and backward isomerization reactions respectively. Having more information about the barriers of isomerization and cracking independently would require additional experiments to confirm the evaluation of the number of acid sites and adsorption data of C7 alkenes on Brønsted acid sites of the zeolite.

Notably, while a single B1 cracking step exists in n-heptane hydrocracking, the reaction network consists of many type B isomerization steps. Given the computational requirements (in particular, for the PBE+D2 AIMD part of the work), we have selected one representative reaction of type B isomerization in the present work. We note that the selected reaction is the only elementary step in which the carbenium corresponding to the 2,2,3-trimethylbutane product is formed (experimentally, alkanes are detected after deprotonation of the carbenium and hydrogenation of alkenes), whereas all other skeletons of carbenium ions are generated in multiple elementary steps. Thus, the chosen isomerization reaction is the best one to extract the least correlated kinetic data.

SV. MLPT CALCULATION WORKFLOW AND SPEED UP WITH RESPECT TO EXPLICIT RPA MD CALCULATIONS

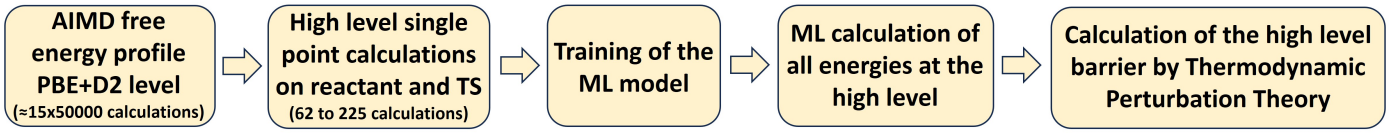

FIG. S5. Workflow of a typical MLPT calculation presented in this work.

As performing the full MD with RPA is not feasible, we rely on estimates based on the following facts:

- At the PBE+D2 level, 1000 MD steps on the system under consideration basically takes 2 hours and 9 minutes on 36 cores, ie 0.0774 CPU h per step (close to 10^{-1} CPU h).
- A typical RPA single point calculation represents 1008 CPU h (in practice, it was performed in 35000 s on 112 cores), hence 1.3×10^4 longer time that in PBE+D2.
- A blue moon profile for the determination of the free energy barrier with 16 points (50 ps each, timestep of 1 fs, meaning 50000 steps) plus 2 free runs for the reactant and the product (100 ps each, meaning 100000 steps) requires 10^6 calculation steps, thus 10^5 CPU hours at the PBE+D2 level, and close to 1.3×10^9 CPU hours with RPA.
- Applying the MLPT method at the RPA level requires to perform first the AIMD at the PBE+D2 level (estimate: 10^5 CPU hours) plus about 200 RPA single points (herein, we made 62 single points for the reactant, 62 for the product, and 62 for the transition state, considering isomerization reactions). The latter takes about 2.10^5 CPU h. The low-level production MD plus the RPA calculations thus correspond to about 3.10^5 CPU h (the other steps of the MLPT calculation are negligible in terms of computational time).
- Thus, the full RPA calculation would take about 4.10^3 more time than the requirements of the MLPT approach.

[S1] L. Smith, A. Davidson, A. Cheetham, A neutron diffraction and infrared spectroscopy study of the acid form of the aluminosilicate zeolite, chabazite (H-SSZ-13), Catal. Lett. 49 (3/4) (1997) 143–146. doi:10.1023/A:1019097019846.

[S2] J. Rey, A. Gomez, P. Raybaud, C. Chizallet, T. Bučko, On the origin of the difference between type A and type B skeletal isomerization of alkenes catalyzed by zeolites: The crucial input of ab initio molecular dynamics, J. Catal. 373 (2019) 361–373. doi:10.1016/j.jcat.2019.04.014.

[S3] J. Rey, P. Raybaud, C. Chizallet, T. Bučko, Competition of Secondary versus Tertiary Carbenium Routes for the Type B Isomerization of Alkenes over Acid Zeolites Quantified by Ab Initio Molecular Dynamics Simulations, ACS Catal. 9 (11) (2019) 9813–9828. doi:10.1021/acscatal.9b02856.

[S4] J. Rey, C. Bignaud, P. Raybaud, T. Bučko, C. Chizallet, Dynamic Features of Transition States for β-Scission Reactions of Alkenes over Acid Zeolites Revealed by AIMD Simulations, Angew. Chem. Int. Ed. 132 (43) (2020) 19100–19104. doi:10.1002/ange.202006065.

[S5] T. Bučko, M. Gešvandtnerová, D. Rocca, Ab Initio Calculations of Free Energy of Activation at Multiple Electronic Structure Levels Made Affordable: An Effective Combination of Perturbation Theory and Machine Learning, J. Chem. Theory Comput. 16 (10) (2020) 6049–6060. doi:10.1021/acs.jctc.0c00486.

[S6] K. M. Jablonka, D. Ongari, S. M. Moosavi, B. Smit, Big-Data Science in Porous Materials: Materials Genomics and Machine Learning, Chem. Rev. 120 (16) (2020) 8066–8129. doi:10.1021/acs.chemrev.0c00004.

[S7] C. Chipot, A. Pohorille (Eds.), Free Energy Calculations, Springer Series in Chemical Physics, Springer Berlin Heidelberg, 2007.

[S8] B. Herzog, M. Chagas da Silva, B. Casier, M. Badawi, F. Pascale, T. Bučko, S. Lebègue, D. Rocca, Assessing the Accuracy of Machine Learning Thermodynamic Perturbation Theory: Density Functional Theory and Beyond, J. Chem. Theory Comput. 18 (3) (2022) 1382–1394. doi:10.1021/acs.jctc.1c01034.

[S9] J.-M. Schweitzer, J. Rey, C. Bignaud, T. Bučko, P. Raybaud, M. Moscovici-Mirande, F. Portejoie, C. James, C. Bouchy, C. Chizallet, Multiscale modeling as a tool for the prediction of catalytic performances: The case of n-heptane hydroconversion in a large-pore zeolite, ACS Catalysis 12 (2) (2022) 1068–1081. doi:10.1021/acscatal.1c04707.
URL https://doi.org/10.1021/acscatal.1c04707