

Cite this: *Catal. Sci. Technol.*, 2024,
14, 5314

Machine learning thermodynamic perturbation theory offers accurate activation free energies at the RPA level for alkene isomerization in zeolites†

Jérôme Rey,^a Michael Badawi,^{ib}*^{ab} Dario Rocca,^a
Céline Chizallet^{ib}*^c and Tomáš Bučko^{ib}*^{de}

The determination of accurate free energy barriers for reactions catalyzed by proton-exchanged zeolites by quantum chemistry approaches is a challenge. While *ab initio* molecular dynamics is often required to sample correctly the various states described by the system, the level of theory also has a crucial impact. In the present work, we report the determination of accurate barriers for a type B isomerization of a monobranched C7 alkene (4-methyl-hex-1-ene) into a dibranched tertiary cation inside a protonated chabazite zeolite. This is done by using the Machine Learning Thermodynamic Perturbation Theory (MLPT) at the Random Phase Approximation (RPA) level, on the basis of blue-moon sampling dynamic data obtained at the Generalized Gradient Approximation (GGA) level (PBE+D2). The comparison of PBE+D2 and RPA profiles shows that the former overstabilizes cationic intermediates with respect to neutral ones. The transition state of the isomerization is a non-classical edge protonated cyclopropane, the stabilization of which is lower than that of the π -complex when PBE+D2 is replaced by RPA, but higher than that of the classical tertiary carbenium. Consequently, the backward isomerization barrier is decreased. Applying the MLPT approach to recompute the free energy barriers with various dispersion correction schemes to the PBE energies shows that none of the schemes is sufficient to improve both the forward and backward barriers with respect to the RPA reference. These data complement previously determined alkene cracking barriers [Rey *et al.*, *Angew. Chem., Int. Ed.*, 2024, **63**, e202312392], thanks to which it is possible to compare the presently determined barriers with reference experimental data [Schweitzer *et al.*, *ACS Catal.*, 2022, **12**, 1068–1081]. The agreement with experiments is significantly improved at the RPA with respect to GGA. Chemical accuracy is approached (maximum deviation of 6.4 kJ mol⁻¹), opening the door to predictive kinetic modelling starting from first principles approaches.

Received 29th April 2024,
Accepted 16th July 2024

DOI: 10.1039/d4cy00548a

rsc.li/catalysis

1 Introduction

The determination of accurate free energy barriers is a key challenge in computational catalysis. Indeed, as it can be deduced from the Eyring equation,¹ errors made on free energy barriers translate exponentially on the rate constant, itself being directly linked to the rate of elementary steps

through rate equations. Reaching chemical accuracy, *i.e.*, predicting barriers with a deviation with respect to experiments lower than 4.2 kJ mol⁻¹, is thus more than ever needed. For example, at a reaction temperature of 500 K, the chemical accuracy implies that the ratio of experimental *versus* computed rate constants must not exceed 2.8.

Achieving this goal requires improving both the level of theory and the quality of the sampling of states to quantify the stability of reactants, transition states and products. In the current common practice, which is determined by the available computer power, performing *ab initio* molecular dynamics (AIMD) for simulation cells relevant for the modelling of heterogeneous catalysts at higher levels of theory than density functional theory (DFT) is not feasible. In the present study, our goal is to quantify accurate barriers making use of the Random Phase Approximation^{2,3} (RPA) level. RPA was indeed shown to perform remarkably well for a series of properties of solids such as lattice parameters,^{4,5} for adsorption energies of various molecules on surface

^a Laboratoire de Physique et Chimie Théoriques (LPCT), CNRS, Université de Lorraine, F-54000 Nancy, France^b Laboratoire Lorrain de Chimie Moléculaire (L2CM), CNRS, Université de Lorraine, F-54000 Nancy, France. E-mail: michael.badawi@univ-lorraine.fr^c IFP Energies Nouvelles, Rond-Point de l'Échangeur de Solaize, BP3, 69360 Solaize, France. E-mail: celine.chizallet@ifpen.fr^d Department of Physical and Theoretical Chemistry, Faculty of Natural Sciences, Comenius University Bratislava, Ilkovičova 6, SK-84215 Bratislava, Slovakia^e Institute of Inorganic Chemistry, Slovak Academy of Sciences, Dúbravská Cesta 9, SK-84236 Bratislava, Slovakia. E-mail: tomas.bucko@uniba.sk† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d4cy00548a>

(including metals^{6,7} and zeolites^{8,9}), and for barriers of simple reactions.^{10–13} Very recently, we quantified free energy barriers for a set of zeolite catalyzed reactions¹⁴ at the RPA level combined to AIMD, thanks to Machine Learning Thermodynamic Perturbation Theory (MLPT).^{15–18} The breakthrough result of this study is that it demonstrated that it is possible to perform sampling of realistic systems at a very high level of theory in a reasonable time (a few weeks), which is just a small fraction of the estimated computational time that would be needed to perform AIMD at the RPA level of the same length in a naive way (a millennium). An unprecedented accuracy was reached when comparing the computed barriers with experimental ones, extracted from a complex reaction network thanks to kinetic modelling.¹⁹

Zeolite-catalyzed reactions play a major role in a large array of mature and developing catalytic processes, belonging to the fields of refining, petrochemistry, pollution abatement, biomass conversion and plastic recycling, among others.^{20–29} Understanding the mechanisms at play and being able to accurately predict the rate constants of important elementary steps for such reactions is thus highly important.³⁰ For a large array of zeolite-catalyzed reactions, reliance on high levels of theory^{31–34} (beyond DFT) and explicit sampling of configuration space *via* AIMD^{35–41} has proven essential to properly capture the nature and stability of intermediates. As mentioned above, effective combination of these two requirements became achievable only very recently,^{14,42,43} thanks to development of machine learning approaches.

Many of the previously cited works address hydrocarbon transformation reactions on Brønsted acidic zeolite, in which main active sites are bridging Si(OH)–Al groups. The reactions involve both neutral (adsorbed alkanes and alkenes) and ionic (carbonium or carbenium ions, obtained by protonation of the hydrocarbon molecules) species. The variation of the free energy estimate of these species as a function of the level of theory (GGA *versus* higher level methods) appeared to depend strongly on the charge state of these species.^{31,34,42,44–46} It is crucial to quantify the intensity of this effect, also accounting for the impact of the charge state on the dynamic behavior of the confined intermediates. This requirement holds true for the family of reactions that we have considered in our previous work,¹⁴ namely the isomerization and cracking of alkenes containing seven carbon atoms.

By the MLPT approach, we have previously addressed type B isomerization reactions (which induce a change in the branching degree of the hydrocarbon skeleton) linking di- and tri-branched tertiary carbenium ions, and a type B₁ cracking reaction transforming a secondary cation into a tertiary carbenium plus propene (Fig. 1a and c). In the present work, we extend this analysis by considering a type B isomerization of a monobranched alkene (4-methyl-hex-1-ene) into a dibranched tertiary cation, passing through a secondary carbenium ion (Fig. 1b). Previous AIMD study conducted at the PBE+D2 level showed that the latter was only an elusive reaction intermediate.³⁵ This additional case study is relevant in a practical context, as the bifunctional

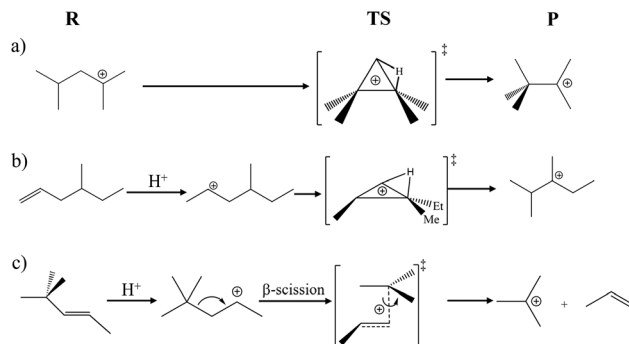


Fig. 1 Reactants (R), transition states (TS) and products (P) of the reactions compared in the present work: a) type B isomerization between dibranched and tribranched tertiary carbenium ions, b) type B isomerization between a mono-branched π -complex and a dibranched tertiary carbenium ion, c) type B₁ β -scission cracking reaction. All reactions are catalyzed by proton-exchanged zeolite chabazite.

hydroisomerization and cracking of linear alkanes pass through the gradual increase of the degree of branching of intermediates (alkenes and carbenium ions, obtained after dehydrogenation of alkanes on a metallic phase).^{19,30,47} We show in the present work that MLPT at the RPA level is a precious tool to approach chemical accuracy, whereas adjusting the dispersion correction scheme brought to the PBE energies does not lead to substantial improvement. A fundamental interest can also be found in the combination of neutral and ionic intermediate states in a same isomerization reaction, whereas in our previous work,¹⁴ a neutral intermediate was invoked as the reactant of a cracking reaction. Thanks to MLPT, a better rationalization can be made between the level of theory and the stability of important species in the reaction network, taking into account their dynamic behavior.

2 Methods

All the simulations have been performed using the periodic DFT program VASP,^{48–50} utilizing the projector-augmented wave (PAW) method implemented by Blöchl and Kresse *et al.*^{51,52} The Kohn–Sham equations were solved self-consistently with a convergence criterion of 10^{-7} eV per cell. The basis set with plane waves with kinetic energy of up to 400 eV was used and the integrations over the Brillouin zone were realized using a single k -point at the position Γ . A Gaussian smearing with a smearing parameter of 0.05 eV was applied. The simulation cell used to represent the chabazite structure is shown in Fig. 2. The proton initially lies onto the framework oxygen atom O1, which is, according to experiment,⁵³ one of the two most populated proton sittings in CHA.

The MLPT calculations were performed using the NVT Born–Oppenheimer AIMD free energy calculations at 500 K (Andersen thermostat, collision frequency per atom of 0.01 fs⁻¹) at the PBE⁵⁴ level with D2 (ref. 55) dispersion correction described in our previous work.³⁶ In the following, we shall refer to the PBE+D2 method used to produce the AIMD data



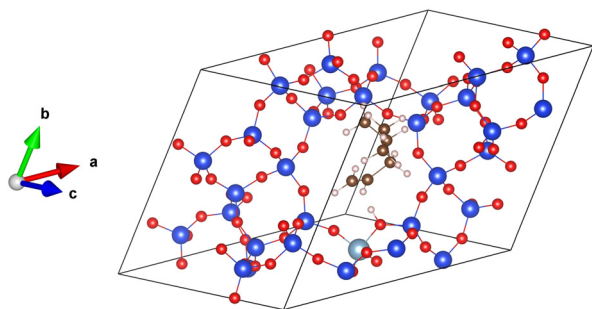


Fig. 2 The supercell of acid chabazite with the π -complex, reactant for the isomerization reaction. Color code: Si in blue, O in red, Al in light blue, C in brown, H in white.

as the production method. The free energy of activation (ΔA^\ddagger) was determined thanks to simulations in the Blue Moon ensemble,^{56,57} as reported in ref. 35. A reaction coordinate based on distances between relevant atoms (C and H from the hydrocarbon moiety, H and O from the zeolite acid site) was chosen and shown to lead to reliable results. In the present work, using the computed free energy of activation and ensembles of reactant and transition state structures obtained with the production method, the free energies of activation at various target method levels ($\Delta \tilde{A}^\ddagger$), involving PBE+D3(0) and PBE+D3(BJ),^{58,59} PBE+D4,^{60–62} PBE+dDsC,⁶³ PBE+TS,^{64,65} PBE+TS/HI,^{66,67} PBE+MBD,^{68–70} PBE+MBD/FI,⁷¹ and RPA,^{2,3,72–74} were determined *via*

$$\Delta \tilde{A}^\ddagger = \Delta A^\ddagger + \Delta A_{\text{TS}} - \Delta A_{\text{R}}, \quad (1)$$

with the terms associated with transition state (TS) and reactant (R) being defined as

$$\Delta A_{\text{TS}} = -k_{\text{B}}T \ln \langle \exp[-\Delta V(\mathbf{q})/k_{\text{B}}T] \rangle_{\xi^*} \quad (2)$$

and

$$\Delta A_{\text{R}} = -k_{\text{B}}T \ln \langle \exp[-\Delta V(\mathbf{q})/k_{\text{B}}T] \rangle, \quad (3)$$

respectively. The angular brackets $\langle \dots \rangle$ and $\langle \dots \rangle_{\xi^*}$ represent the NVT ensemble average over free R, and that over the configurations with the reaction coordinate ($\xi(\mathbf{q})$) fixed at the value characteristic for TS (ξ^*), respectively. k_{B} is the Boltzmann constant, T is thermodynamic temperature, and $\Delta V(\mathbf{q})$ is the difference in potential energy computed using the target and production methods for a configuration with Cartesian coordinates \mathbf{q} . Note that eqn (1) can be used to express also $\Delta \tilde{A}^\ddagger$ of the reverse reaction mode, but the term ΔA_{R} must be replaced by ΔA_{P} corresponding to contribution of product (P), which is defined analogously to eqn (3).

As an auxiliary quantity, the internal energy of activation can be defined as follows:^{15,57}

$$\Delta \tilde{U}^\ddagger = \tilde{U}_{\text{TS}} - \tilde{U}_{\text{R}}, \quad (4)$$

with

$$\tilde{U}_{\text{TS}} = \frac{\langle \tilde{V}(\mathbf{q}) Z^{-1/2} \exp[-\Delta V(\mathbf{q})/k_{\text{B}}T] \rangle_{\xi^*}}{\langle Z^{-1/2} \exp[-\Delta V(\mathbf{q})/k_{\text{B}}T] \rangle_{\xi^*}}, \quad (5)$$

and

$$\tilde{U}_{\text{R}} = \frac{\langle \tilde{V}(\mathbf{q}) \exp[-\Delta V(\mathbf{q})/k_{\text{B}}T] \rangle}{\langle \exp[-\Delta V(\mathbf{q})/k_{\text{B}}T] \rangle}, \quad (6)$$

where $Z = \sum_{i=1}^N \sum_{v=x,y,z} \frac{1}{m_i} \left(\frac{\partial \xi}{\partial q_{i,v}} \right)^2$ with N and m_i being the number of atoms in the system and the mass of an atom i .

In order to compute the terms ΔA_i and \tilde{U}_i , up to 225 configurations from production method AIMD of the state i were selected and the target method potential energies were computed in single point calculations. The differences $\Delta V(\mathbf{q})$ were then determined using the production method energies available from AIMD and these data were employed to train a Δ -ML model that was subsequently used to predict $\Delta V(\mathbf{q})$ for a representative ensemble of configurations of a state i (every 10th point of the original production method trajectory). As in our previous work,¹⁴ the kernel ridge regression⁷⁵ with the RMatch kernel⁷⁶ and the SOAP (smooth overlap of atomic positions) descriptors⁷⁷ were used for this purpose. Compared to other ML-based methods used to study chemical reactions, our approach has advantage that it does not require building a global model for whole reaction process but, instead, it uses two separate models specifically trained for R and TS.

The results for ΔA_{TS} and ΔA_{R} obtained as described above can be affected by the systematic error due to imperfect Δ -ML model. As shown in ref. 16, this source of error can be significantly reduced by applying the correction:

$$\Delta A_i \leftarrow \Delta A_i + \Delta_{\varepsilon} \quad (7)$$

where

$$\Delta_{\varepsilon} = \langle \varepsilon(\mathbf{q}) \rangle - \frac{\langle \varepsilon(\mathbf{q}) \rangle - \langle \varepsilon \rangle^2}{2k_{\text{B}}T} - \frac{\langle \varepsilon(\mathbf{q}) \Delta V(\mathbf{q}) \rangle - \langle \varepsilon \rangle \langle \Delta V(\mathbf{q}) \rangle}{k_{\text{B}}T} \quad (8)$$

and $\varepsilon(\mathbf{q})$ is the difference between actual and predicted $\Delta V(\mathbf{q})$. Note that the terms on the right hand side of eqn (8) are closely related to the mean error (ME), root mean square error (RMSE) and covariance of error with $\Delta V(\mathbf{q})$ (COVE), which are often used as measures of quality of the ML model and can be determined using a reasonably large independent test set with known exact labels (here $\Delta V(\mathbf{q})$). Thus, our correction scheme can be integrated into the usual training-validation workflow conventionally used in the ML-based calculations. In the context of present work, the test set used to determine Δ_{ε} is termed “correction set”. Due to enormous computational demands of RPA calculations, we sought for a computationally more effective solution that allows us to estimate $\varepsilon(\mathbf{q})$ without the need of defining an independent correction set. For this purpose, we test in this work the use



of the leave one out (LOO) strategy,⁷⁸ in which, out the N_{train} labelled data points, $N_{\text{train}} - 1$ points are used as a training set to prepare a model which is subsequently used to make prediction for the remaining “left-out” point. The whole process is repeated N_{train} times such that N_{train} independent predictions (subsequently used as correction set) are made using the same number of models that are all very similar (sharing all but one training points) with the model validated (trained on all N_{train} points). Obviously, the prerequisite for this strategy to yield realistic estimates of Δ_{e} is a sufficiently high N_{train} enabling a reasonable estimates of ME, RMSE, and COVE on one hand, and ensuring near identity of models trained on $N_{\text{train}} - 1$ and N_{train} data points on the other. We test this approach on the example of a time-effective DFT calculation in Sec. 3.1.

Finally, let us remark that if the ΔA_{TS} and ΔA_{R} terms are determined using a Δ -ML model based on the same method and values of hyperparameters, which is the case in this study, the corresponding Δ_{e} values are likely to take very similar values and, consequently, their contribution to $\Delta\tilde{A}^{\ddagger}$ determined *via* eqn (1) is likely to vanish. Indeed, $\Delta\tilde{A}^{\ddagger}$ was shown^{16,79} to converge much faster than the individual ΔA_{TS} and ΔA_{R} terms.

3 Results and discussion

3.1 Convergence of the MLPT results

In this section we examine the convergence of the ΔA_{R} , ΔA_{TS} , and $\Delta\tilde{A}^{\ddagger} - \Delta A^{\ddagger} = \Delta A_{\text{TS}} - \Delta A_{\text{R}}$ terms for the forward reaction mode with respect to the training set size used in the MLPT calculations. As a demonstration example, PBE+dDsC target method is used but similar trends are observed also for other methods considered here. For both states, R and TS, 225 configurations were first selected from production parts of the MD-generated trajectories in such a way that their separation was 400 MD steps (*i.e.*, more than assumed correlation time of ~ 100 fs). Out of these configurations, $N_{\text{train}} = 10, 25, 62, 113,$ and 225 structures were used as a

Table 1 Terms ΔA_{R} , ΔA_{TS} , and $\Delta\tilde{A}^{\ddagger} - \Delta A^{\ddagger} = \Delta A_{\text{TS}} - \Delta A_{\text{R}}$ (all in kJ mol^{-1}) computed using MLPT for the PBE+D2 production and PBE+dDsC target methods for the forward step of the type B isomerization reaction from the mono-branched π -complex to the dibranched tertiary carbenium ion. Results obtained using different sizes of training set (N_{train}) and different methods for computing correction Δ_{e} (no correction (None), independent correction set (225- N_{train}), or leave one out procedure (LOO))

Quantity	Correction	N_{train}				
		10	25	62	113	225
ΔA_{R}	None	-44.38	-44.18	-43.58	-43.48	-43.62
	225- N_{train}	-43.62	-43.53	-43.66	-43.72	—
	LOO	-45.96	-44.74	-43.86	-43.62	-43.70
ΔA_{TS}	None	-39.52	-39.17	-39.57	-39.70	-39.74
	225- N_{train}	-40.22	-40.02	-39.90	-39.99	—
	LOO	-40.36	-39.69	-39.84	-39.87	-39.81
$\Delta\tilde{A}^{\ddagger} - \Delta A^{\ddagger}$	None	4.86	5.01	4.01	3.77	3.88
	225- N_{train}	3.41	3.51	3.76	3.72	—
	LOO	5.60	5.05	4.02	3.75	3.89

training and remaining 225- N_{train} as a correction set for calculation of Δ_{e} *via* eqn (8). Alternatively, the LOO strategy described in Sec. 2 was applied using the training set only.

The numerical results obtained for the ΔA_{R} , ΔA_{TS} , and $\Delta\tilde{A}^{\ddagger}$ terms are listed in Table 1. As evident from the data, the uncorrected values show very rapid convergence with respect to the training set size, and even the model prepared with $N_{\text{train}} = 10$ yields result for $\Delta\tilde{A}^{\ddagger} - \Delta A^{\ddagger}$ that is within 1 kJ mol^{-1} identical to our best estimate obtained with $N_{\text{train}} = 225$. Such an excellent convergence is possible because of the strong linear correlation between the potential energies of the target and production methods, which can be demonstrated for all target methods considered in this work (see Fig. 3).

Despite this strong correlation, the variance in $\Delta V(\mathbf{q})$ is far from being negligible - if this were the case, the calculation of $\Delta\tilde{A}^{\ddagger} - \Delta A^{\ddagger}$ would be trivial, namely $\Delta\tilde{A}^{\ddagger} - \Delta A^{\ddagger} = \Delta V(\mathbf{q}_i)$ for

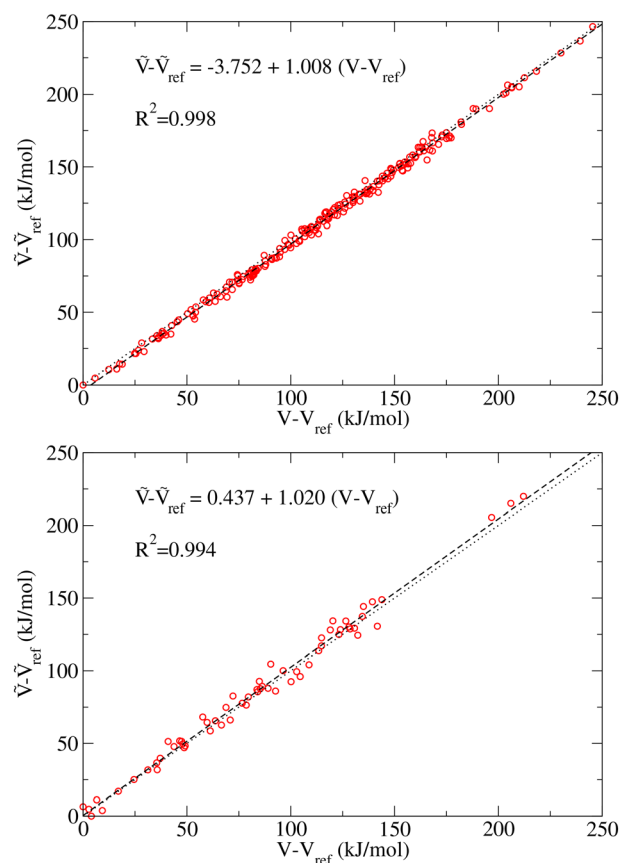


Fig. 3 Correlation between target (\tilde{V}) and production (V) potential energies of reactant of the type B isomerization of mono-branched π -complex into a dibranched tertiary carbenium. Results for the PBE+dDsC (above) and RPA (below) target and PBE+D2 production methods are shown. The dashed lines correspond to the linear fit to data points (represented as red symbols) computed for selected structures generated using the AIMD simulation at the PBE+D2 level (the parameters of the linear fit and the coefficient of determination are shown in the inset of each plot), while the dotted lines represent a perfect agreement between both sets of data. For easier visualization, all potential energies are referenced to the minimal values obtained within the dataset using the given method (\tilde{V}_{ref} and V_{ref} for target and production methods, respectively).



arbitrarily chosen q_i . In the case of the 225 configurations reactant state, for instance, $\Delta V(q)$ varies between -46.88 and -34.61 kJ mol $^{-1}$. Since a similar variance is observed also in the TS data (here $\Delta V(q)$ ranges between -48.84 and -27.98 kJ mol $^{-1}$), the uncertainty in prediction of $\Delta A^\ddagger - \Delta A^\ddagger$ made using a randomly selected single R and TS configurations would be much greater than the $\Delta A^\ddagger - \Delta A^\ddagger$ itself (-3.86 kJ mol $^{-1}$, see Table 1) and even the sign of the free energy difference would be a matter of chance.

The correction Δ_ϵ obtained using 225- N_{train} data points basically eliminates any uncertainty resulting from systematic error of ML in predicted ΔA_{R} , ΔA_{TS} , and ΔA^\ddagger (see Table 1). At first sight, the benefit of the present procedure might not be obvious since, after all, the same number of data points is used in all cases with different N_{train} , the difference being only in their partitioning between the training and correction set. Nevertheless, even this simple example shows that, since the systematic error can be eliminated *a posteriori*, one should actually demand convergence of corrected rather than uncorrected free energy values. This idea is particularly useful in context of our alternative computationally more effective simulation protocol in which the correction set is prepared using the LOO procedure (see Sec. 2) and hence the explicit target method calculations need to be performed only for N_{train} data points. In this case, as expected, the quality of correction for very small training sets is poor but systematically improves by increasing N_{train} (see Table 1) and the values obtained with 62 and 225 datapoints are virtually identical to the results corrected with independent training set (Table 1). Importantly, consistency in corrected and uncorrected results for $\Delta A_{\text{TS}} - \Delta A_{\text{R}}$ obtained using a certain value of N_{train} clearly indicates that the systematic error does not significantly affect predictions of free energy differences. As shown in Table 2, this indeed holds true for all methods discussed in this work when $N_{\text{train}} = 62$ was used. Discussion presented in remaining part of this work is therefore based on the results obtained with this setting and correction estimated from the LOO procedure.

Finally we note that a good overlap between the configuration spaces sampled by target and production methods must be achieved to get reliable MLPT results.⁸⁰ This important requirement can conveniently be measured by the

index I_w ^{16,17} with the limiting values of 0.0 and 0.5, corresponding to zero and perfect overlap, respectively. This is achieved by computing the weighting factors $w_i = e^{-\Delta V(q_i)/k_B T}$ for all N configurations used in MLPT calculation, ordering them in an ascendant order, and identifying the lowest M fulfilling

the condition $\frac{\sum_{i=1}^M w_i}{\sum_{j=1}^N w_j} \geq 0.5$, which then defines I_w as follows:

$$I_w = \frac{(N-M)}{N}. \quad (9)$$

As shown by Herzog *et al.*,¹⁷ the MLPT procedure is reliable if $I_w \geq 0.05$, which was fulfilled for all methods except of PBE+TS/HI applied to the TS structure, see Table 3.

3.2 MLPT results at various levels of theory

In the type B isomerization reaction investigated in this work, a π -complex (reactant) is first transformed into a monobranched secondary cation after proton is transferred from the zeolite. The very elusive secondary cation is subsequently isomerized into a tertiary dibranched cation *via* an edge protonated cyclopropane transition state (Fig. 1b). The corresponding free and internal energy barriers for the forward and backward steps determined by AIMD (PBE+D2) and MLPT (other levels of theory) are reported in Table 4. In addition, the free energy profiles computed at the PBE+D2 (AIMD) and RPA (MLPT) levels of theory are compared in Fig. 4.

The forward step corresponds to the transition from the neutral π -complex to the cationic edge protonated cyclopropane (PCP) transition structure and the corresponding ΔA^\ddagger obtained in Blue Moon ensemble AIMD simulations at the PBE+D2 level is 69.2 kJ mol $^{-1}$.³⁵ The MLPT calculation at the RPA target method level leads to increase of this barrier to 76.3 kJ mol $^{-1}$, which is related to a relative stabilization of the π -complex reactant with respect to the cationic edge protonated transition state. For the backward reaction, which involves transformation from tertiary dibranched cation to PCP, the PBE+D2 free energy of activation is 94.5 kJ mol $^{-1}$. Unlike for the forward reaction,

Table 2 Terms $\Delta A_{\text{TS}} - \Delta A_{\text{R}}$ and $\Delta A_{\text{TS}} - \Delta A_{\text{P}}$ (in kJ mol $^{-1}$) obtained for various methods using MLPT with $N_{\text{train}} = 62$ without and with (indicated with prime) correction Δ_ϵ (see eqn (8)) determined using the LOO procedure

Target method	$\Delta A_{\text{TS}} - \Delta A_{\text{R}}$	$\Delta A'_{\text{TS}} - \Delta A'_{\text{R}}$	$\Delta A_{\text{TS}} - \Delta A_{\text{P}}$	$\Delta A'_{\text{TS}} - \Delta A'_{\text{P}}$
PBE+D3(0)	3.55	3.70	7.56	7.52
PBE+D3(BJ)	6.11	6.16	7.16	7.09
PBE+D4	4.86	5.00	6.35	6.36
PBE+TS	1.70	1.87	3.72	3.57
PBE+TS/HI	-2.97	-2.81	6.08	5.90
PBE+dDsC	4.01	4.02	3.66	3.51
PBE+MBD	1.48	1.54	3.70	3.64
PBE+MBD/FI	-0.85	-0.83	3.98	3.87
RPA	6.92	7.09	-1.62	-2.04

Table 3 The I_w values obtained in MLPT calculations of reactant (R), transition state (TS), and product (P) of the type B isomerization reaction with PBE+D2 production and different target methods

Target method	R	TS	P
PBE+D3(0)	0.16	0.08	0.20
PBE+D3(BJ)	0.23	0.13	0.24
PBE+D4	0.23	0.14	0.23
PBE+TS	0.17	0.17	0.18
PBE+TS/HI	0.06	0.02	0.13
PBE+dDsC	0.27	0.19	0.25
PBE+MBD	0.21	0.18	0.20
PBE+MBD/FI	0.23	0.16	0.28
RPA	0.12	0.07	0.13



Table 4 Forward and backward isomerization free (ΔA^\ddagger) and internal (ΔU^\ddagger) energies of activation (kJ mol^{-1}) determined by Blue Moon ensemble AIMD (PBE+D2) and MLPT (other levels) using various levels of theory. Errors with respect to the RPA reference are shown in parentheses

Method	Forward		Backward	
	ΔA^\ddagger	ΔU^\ddagger	ΔA^\ddagger	ΔU^\ddagger
PBE+D2	69.2 (-7.1)	72.6 (-2.7)	94.5 (2.0)	60.6 (1.8)
PBE+D3(0)	72.9 (-3.4)	77.9 (2.6)	102.0 (9.5)	68.5 (9.7)
PBE+D3(BJ)	75.4 (-0.9)	82.4 (7.1)	101.6 (9.1)	68.4 (9.6)
PBE+D4	74.2 (-2.1)	80.7 (5.4)	100.9 (8.4)	69.5 (9.7)
PBE+TS	71.1 (-5.2)	74.8 (-0.5)	98.1 (5.6)	58.6 (-0.2)
PBE+TS/Hi	66.4 (-9.9)	69.0 (-6.3)	100.4 (7.9)	68.8 (10.0)
PBE+dDsC	73.2 (-3.1)	77.6 (2.3)	98.0 (5.5)	62.3 (3.5)
PBE+MBD	70.7 (-5.6)	76.1 (0.8)	98.1 (5.6)	67.6 (8.8)
PBE+MBD/FI	68.4 (-7.9)	75.3 (0.0)	98.4 (5.9)	65.6 (6.8)
RPA	76.3	75.3	92.5	58.8

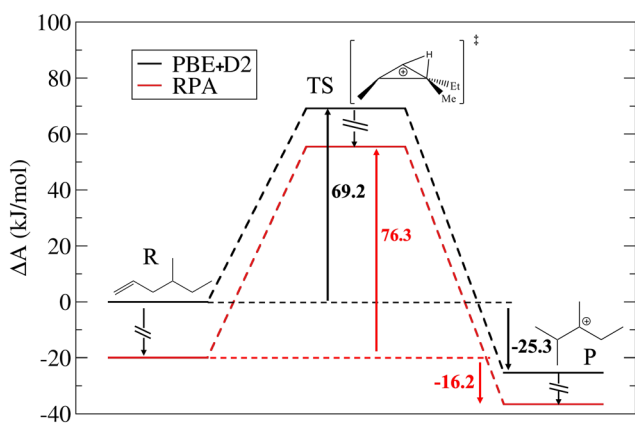


Fig. 4 Comparison of free energy diagrams computed by AIMD PBE+D2 and MLPT RPA levels of theory for the isomerization reaction from the monobranched alkene to the tertiary dibranched cation.

RPA predicts decrease of ΔA^\ddagger by 2.0 kJ mol^{-1} compared to the PBE+D2 result. Thus, in contrast to the type B isomerization linking tertiary cations considered in our previous work¹⁴ (Fig. 1a), the product (tertiary cation) is slightly less stabilized than the cationic PCP transition state by RPA. Computing the free energy of reaction as the difference between ΔA^\ddagger determined for the forward and backward reaction modes, we find that both methods predict that the tertiary carbenium ion is more stable than the π -complex. Nevertheless, the relative stability of the former with respect to the latter is significantly reduced when PBE+D2 is replaced by RPA, whereby the free energy of reaction increases from $-25.3 \text{ kJ mol}^{-1}$ to $-16.2 \text{ kJ mol}^{-1}$.

Since the difference between the ΔA^\ddagger and ΔU^\ddagger terms determined for the $R \rightarrow TS$ transformation (see Table 4) is only 3.4 (PBE+D2) and -1.0 kJ mol^{-1} (RPA), one can deduce that the entropy change linked with this first part of reaction is only modest. In contrast, $\Delta A^\ddagger - \Delta U^\ddagger$ for the $P \rightarrow TS$ transformation is much larger (33.9 and 33.7 kJ mol^{-1} for PBE+D2 and RPA, respectively), indicating that the entropy of product is significantly higher than that of reactant and

transition states. As explained in our previous work,³⁵ the entropy gain in the product state is due mainly to release of frustrated rotations and translations upon breaking relatively strong H-bonds present in reactant. Interestingly, this entropic effect predicted by RPA is about the same as that obtained in PBE+D2 calculations.

Combining the free energy results presented here with the data previously reported for the type B isomerization between tertiary cations and for type B₁ cracking,^{14,37} we predict the general trend according to which all neutral intermediates are more stabilized with respect to ionic species at the RPA than at the PBE+D2 level of theory. The fact that PBE+D2 overestimates the stability of ionic species with respect to higher levels of theory is qualitatively compatible with previous reports made on short-chains alkenes.^{42,44} The isomerization PCP transition states do not follow strictly the trend of tertiary carbenium ions, so that the barriers linking tertiary cations to these TS are lower at the RPA level.

Equipped with the RPA high-level reference results, we can now proceed to evaluation of the quality of approximate methods based on various dispersion correction methods to the PBE functional, which belongs to most popular computationally inexpensive *ab initio* simulation methods used in theoretical catalysis. For the forward reaction mode, the approximate methods predict change in ΔA^\ddagger by -2.8 (PBE+TS/Hi) to 6.2 kJ mol^{-1} (PBE+D3(BJ)) with respect to the PBE+D2 method, whereby the closest agreement with the RPA reference is achieved by the PBE+D3(BJ) followed by the PBE+D4 and PBE+dDsC methods (ΔA^\ddagger of 75.4 , 74.2 , and 73.2 kJ mol^{-1} , respectively). In contrast, the PBE+TS/Hi and PBE+MBD/FI methods predict a qualitatively different trend, *i.e.*, a slight reduction in ΔA^\ddagger (by -2.8 and -0.8 kJ mol^{-1} , respectively). Disappointingly somewhat, all the dispersion correction methods predict incorrect trend for the free energy barrier of the backward reaction mode, whereby the predicted changes range between 3.5 (PBE+dDsC) to 7.5 kJ mol^{-1} (PBE+D3(0)). Importantly, the effect of dispersion correction on the forward and backward barrier is similar in most cases, leading to a very small effect on the free energy of reaction. Exceptions are the PBE+TS/Hi and PBE+MBD/FI methods, which predict an increase in backward and a decrease in forward barriers, leading to a significant decrease in the free energy of reaction (-34.0 and $-30.0 \text{ kJ mol}^{-1}$, respectively) in comparison to the PBE+D2 method ($-25.3 \text{ kJ mol}^{-1}$), which is in striking contrast to the trend observed for the RPA method (increase to $-16.2 \text{ kJ mol}^{-1}$).

Furthermore, comparison of the ΔA^\ddagger and ΔU^\ddagger values presented in Table 4 reveals that although the general entropy trend along the $R \rightarrow TS \rightarrow P$ sequence obtained by the different methods remain qualitatively the same as that predicted by the RPA and PBE+D2 methods, the quantitative differences are non-negligible. Hence, $\Delta A^\ddagger - \Delta U^\ddagger$ for the $R \rightarrow TS$ step ranges between 2.6 (PBE+TS) and 7.0 kJ mol^{-1} (PBE+D3(BJ)), while that for the $P \rightarrow TS$ step is between 30.5 (PBE+MBD) and 39.5 (PBE+TS) . These



results show that a significant portion of the deviation between most levels of theory and RPA can be assigned to entropic factors.

All these results suggest that the improvement in the approximate results cannot be achieved *via* the choice of the method for description of the long-range dispersion interactions but rather through a careful selection of the density functional approximation. In particular, the use of hybrid functionals that incorporate a portion of Hartree–Fock exchange can be expected to reduce some of the errors of semi-local functionals, such as the self-interaction or delocalization errors, possibly leading to improvement of the description of ionic states and reaction barriers. Unfortunately, in line with our previous work,^{14,17} our attempted MLPT calculations with hybrid functionals (namely HSE06 (ref. 81–83) and B3LYP^{84,85}) did not yield reliable results because of poor overlaps between the configuration spaces sampled by the production and target methods ($I_w < 0.001$). As proposed by Herzog *et al.*,¹⁷ this problem can be overcome by performing additional Monte Carlo calculations employing the Δ -ML model trained within MLPT (so called MLMC approach). However, these calculations would require performing additional tens of thousands of explicit calculations of the production method, and hence represent a significant additional computational cost to an already very high cost of the RPA simulations presented in this work. Nevertheless, our qualitative static calculations (see Sec. SI in the ESI†) suggest that hybrid functionals will not lead to reliable results with respect to the RPA reference level.

3.3 Comparison with experimental and kinetic modelling results

With these new data in hand, we can now compare the improvement brought by the RPA MLPT approach with the previous estimates, using as reference data extracted from the kinetic modelling of experimental data (bifunctional hydroisomerization of *n*-heptane) acquired thanks to a high-throughput setup, as reported in ref. 19. The comparison between computational and experimental results is not straightforward as the reaction network consists of forty type B isomerization steps, including the ones (forward and backward) investigated in the present work (termed TS5b in ref. 19). A single-event kinetic model was built on the basis of previous AIMD calculations for the cracking steps³⁷ and on static calculations made on purpose for the whole set of type B isomerization steps.

The agreement between experiments and the *ab initio*-based kinetic model was not optimal at the first try. The difference between the free energy barriers of the type B isomerization and B₁ cracking needed to be adjusted to reproduce accurately the isomerization over cracking selectivity. The choice was made to keep the isomerization barrier unchanged, and to increase the cracking barrier by

Table 5 Free energy barriers (kJ mol⁻¹) at 500 K of the forward and backward type B isomerization step (Isom. B) considered in the present work and for the B₁ cracking step obtained using AIMD at the PBE+D2, MLPT at the RPA level and the kinetic modelling based on experimental data. For the forward isomerization step and the cracking step, the reactant state is considered to be a π -complex

Step	PBE+D2 AIMD	RPA MLPT	Kinetic modelling
Isom. B, forward	69.2	76.3	74.7
Isom. B, backward	94.5	92.5	89.5
B ₁ cracking	60.1	71.6	75.0

15 kJ mol⁻¹, but one may have chosen to decrease by 15 kJ mol⁻¹ the isomerization barriers, keeping the cracking barrier unchanged. A similar improvement of the description of the isomerization *versus* cracking selectivity would have been modelled. 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, we consider as “reference data” the free energy barrier differences between type B isomerization and B₁ cracking, but not the absolute values themselves.

Table 5 reports the free energy barriers for the currently considered type B isomerization and for the B₁ cracking step, by the various computational approaches undertaken (AIMD at the PBE+D2 level, RPA MLPT) and the one used in the optimal kinetic model. From these values, the difference between free energy barriers of type B isomerization and B₁ cracking (kJ mol⁻¹) was determined and reported in Table 6. It can be seen very clearly that the RPA level provides much better results when comparing to the available reference data than the PBE+D2 AIMD calculations. Chemical accuracy (deviation with respect to experiment within 4.2 kJ mol⁻¹) is approached with RPA MLPT (remaining deviation from experimental data of 5.0 and 6.4 kJ mol⁻¹ for the forward and backward isomerization steps, respectively). These results are consistent with those obtained in our previous work¹⁴ for the type B isomerization reaction involving tertiary cations only (Fig. 1a) showing that the improvement achieved by replacing the PBE+D2 by the RPA level of theory is significant regardless of the nature of the reactant under consideration (π -complex or cation).

Table 6 Differences between free energy barriers of type B isomerization and B₁ cracking (kJ mol⁻¹), for the forward and backward isomerization steps, determined by AIMD at the PBE+D2 level (data extracted from ref. 35 and 37), by MLPT at the RPA level (data from ref. 14 and present work), and those used in the optimal kinetic model (from ref. 19). The latter being compared to experimental results is considered as a reference data. Deviation of the computed parameters with respect to the reference data is reported into brackets

Isomerization step	PBE+D2 AIMD	RPA MLPT	Kinetic modelling
Forward	9.1 (9.4)	4.7 (5.0)	-0.3
Backward	34.4 (19.9)	20.9 (6.4)	14.5



4 Conclusions

In the present work, an accurate quantification of the barrier of the Brønsted acidic zeolite-catalyzed type B isomerization of a monobranched C7 alkene (4-methyl-hex-1-ene) into a dibranched tertiary cation was reported at the RPA level and PBE method linked with various dispersion corrections. The combination of AIMD with RPA was made possible thanks to MLPT within a few weeks, whereas it is not feasible to date to perform full AIMD at the RPA level of theory.

The isomerization reaction considered here is of interest because it links a neutral reactant (a π -complex) to a cationic species (a tertiary carbenium ions). Analyzing the respective stabilization of neutral and ionic species when shifting from PBE+D2 to RPA indicates that PBE+D2 overstabilizes cationic species with respect to neutral ones, in agreement with previous reports made on shorter-chain species. The transition state of the isomerization is a non-classical carbenium ion, called an edge protonated cyclopropane. Accordingly, its stabilization is weaker than that of the π -complex, but stronger than that of the classical tertiary carbenium ion, leading to a slight decrease of the backward isomerization barrier.

Using a level of theory as high as RPA appeared to be mandatory to approach chemical accuracy. Deviations with respect to reference experimental data, extracted thanks to kinetic modelling, were as low as 5.0 and 6.4 kJ mol⁻¹ when considering the difference between the B₁ cracking barrier and the forward/backward isomerization barrier respectively. This free energy barrier difference is indeed the relevant kinetic parameter to compare to, as absolute barriers of independent events cannot unambiguously be determined with the current experimental results. Applying the MLPT approach to a variety of dispersion correction schemes to the PBE energies, reveals that such changes are not sufficient to reduce the deviation with respect to reference data.

The use of MLPT has thus the potential of being instrumental in accessing high levels of theory while accounting for dynamic effects in heterogeneous catalysis, a first step toward the predictive kinetic modeling of catalytic reactions starting from *ab initio* calculations.

Data availability

Data for this article, including MLPT codes, production and target levels calculations, are available at Zenodo at <https://doi.org/10.5281/zenodo.12597668>.

Author contributions

JR: investigation, formal analysis, writing – original draft, writing – review & editing, DR: conceptualization, writing – review & editing, MB: conceptualization, resources, supervision, project administration, writing – review & editing, CC: conceptualization, project administration, writing – original draft, writing – review & editing, TB: conceptualization, project administration, investigation, formal analysis, writing – original draft, writing – review & editing.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was granted access to the HPC resources of TGCC under the allocation 2022-A0120810433 and 2023-A0140810433 made by GENCI. Part of the research was obtained using the computational resources procured in the national project National competence centre for high performance computing within the Operational programme Integrated infrastructure (project code: 311070AKF2). J. R., D. R. and M. B. are grateful for the financial support given by the COMETE project (CONCEPTION *in silico* de Matériaux pour l'Environnement et l'Energie), co-funded by the European Union as part of the “FEDER-FSE Lorraine et Massif des Vosges 2014-2020” program. C. C. and M. B. acknowledge the Agence Nationale de la Recherche under France 2030 (contract ANR-22-PEBB-0009), for support in the context of the MAMABIO project (B-BEST PEPR). T. B. acknowledges support from Slovak Research and Development Agency under the Contract No. APVV-20-0127 and the grant VEGA 1/0254/24 from the Ministry of Education Research, Development and Youth of the Slovak Republic.

References

- H. Eyring, *J. Chem. Phys.*, 1935, **3**, 107–115.
- D. Bohm and D. Pines, *Phys. Rev.*, 1953, **92**, 609–625.
- D. Langreth and J. Perdew, *Solid State Commun.*, 1975, **17**, 1425–1429.
- J. Harl and G. Kresse, *Phys. Rev. Lett.*, 2009, **103**, 056401.
- J. Harl, L. Schimka and G. Kresse, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2010, **81**, 115126.
- X. Ren, P. Rinke and M. Scheffler, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2009, **80**, 045402.
- L. Schimka, J. Harl, A. Stroppa, A. Grüneis, M. Marsman, F. Mittendorfer and G. Kresse, *Nat. Mater.*, 2010, **9**, 741–744.
- F. Göttl, A. Grüneis, T. Bučko and J. Hafner, *J. Chem. Phys.*, 2012, **137**, 114111.
- D. Rocca, A. Dixit, M. Badawi, S. Lebègue, T. Gould and T. C. V. Bučko, *Phys. Rev. Mater.*, 2019, **3**, 040801.
- J. Paier, X. Ren, P. Rinke, G. E. Scuseria, A. Grüneis, G. Kresse and M. Scheffler, *New J. Phys.*, 2012, **14**, 043002.
- J. E. Bates, P. D. Mezei, G. I. Csonka, J. Sun and A. Ruzsinszky, *J. Chem. Theory Comput.*, 2017, **13**, 100–109.
- C. Waitt, N. M. Ferrara and H. Eshuis, *J. Chem. Theory Comput.*, 2016, **12**, 5350–5360.
- J. Chedid, N. M. Ferrara and H. Eshuis, *Theor. Chem. Acc.*, 2018, **137**, 158.
- J. Rey, C. Chizallet, D. Rocca, T. Bučko and M. Badawi, *Angew. Chem., Int. Ed.*, 2024, **63**, e202312392.
- B. Chehaibou, M. Badawi, T. Bučko, T. Bazhirov and D. Rocca, *J. Chem. Theory Comput.*, 2019, **15**, 6333–6342.
- T. Bučko, M. Gešvandtnerová and D. Rocca, *J. Chem. Theory Comput.*, 2020, **16**, 6049–6060.



- 17 B. Herzog, M. Chagas da Silva, B. Casier, M. Badawi, F. Pascale, T. Bučko, S. Lebègue and D. Rocca, *J. Chem. Theory Comput.*, 2022, **18**, 1382–1394.
- 18 B. Herzog, A. Gallo, F. Hummel, M. Badawi, T. Bučko, S. Lebègue, A. Grüneis and D. Rocca, *npj Comput. Mater.*, 2024, **10**, 68.
- 19 J.-M. Schweitzer, J. Rey, C. Bignaud, T. Bučko, P. Raybaud, M. Moscovici-Mirande, F. Portejoie, C. James, C. Bouchy and C. Chizallet, *ACS Catal.*, 2022, **12**, 1068–1081.
- 20 A. Corma, *Chem. Rev.*, 1995, **95**, 559–614.
- 21 G. Busca, *Chem. Rev.*, 2007, **107**, 5366–5410.
- 22 W. Vermeiren and J.-P. Gilson, *Top. Catal.*, 2009, **52**, 1131–1161.
- 23 E. T. C. Vogt and B. M. Weckhuysen, *Chem. Soc. Rev.*, 2015, **44**, 7342–7370.
- 24 T. Ennaert, J. van Aelst, J. Dijkmans, R. de Clercq, W. Schutyser, M. Dusselier, D. Verboekend and B. F. Sels, *Chem. Soc. Rev.*, 2016, **45**, 584–611.
- 25 A. J. Martín, C. Mondelli, S. D. Jaydev and J. Pérez-Ramírez, *Chem*, 2021, **7**, 1487–1533.
- 26 A. T. Sandro Brandenberger, O. Kröcher and R. Althoff, *Catal. Rev.: Sci. Eng.*, 2008, **50**, 492–531.
- 27 A. M. Beale, F. Gao, I. Lezcano-Gonzalez, C. H. F. Peden and J. Szanyi, *Chem. Soc. Rev.*, 2015, **44**, 7371–7405.
- 28 C. Bornes, I. C. Santos-Vieira, R. Vieira, L. Mafra, M. M. Simões and J. Rocha, *Catal. Today*, 2023, **419**, 114159.
- 29 Z. Dong, W. Chen, K. Xu, Y. Liu, J. Wu and F. Zhang, *ACS Catal.*, 2022, **12**, 14882–14901.
- 30 C. Chizallet, C. Bouchy, K. Larmier and G. Pirngruber, *Chem. Rev.*, 2023, **123**, 6107–6196.
- 31 C. Tuma, T. Kerber and J. Sauer, *Angew. Chem., Int. Ed.*, 2010, **49**, 4678–4680.
- 32 G. Piccini, M. Alessio and J. Sauer, *Angew. Chem., Int. Ed.*, 2016, **55**, 5235–5237.
- 33 M. Rybicki and J. Sauer, *J. Am. Chem. Soc.*, 2018, **140**, 18151–18161.
- 34 F. Berger, M. Rybicki and J. Sauer, *J. Catal.*, 2021, **395**, 117–128.
- 35 J. Rey, P. Raybaud, C. Chizallet and T. Bučko, *ACS Catal.*, 2019, **9**, 9813–9828.
- 36 J. Rey, A. Gomez, P. Raybaud, C. Chizallet and T. Bučko, *J. Catal.*, 2019, **373**, 361–373.
- 37 J. Rey, C. Bignaud, P. Raybaud, T. Bučko and C. Chizallet, *Angew. Chem.*, 2020, **132**, 19100–19104.
- 38 V. Van Speybroeck, M. Bocus, P. Cnudde and L. Vanduyfhuys, *ACS Catal.*, 2023, **13**, 11455–11493.
- 39 V. Van Speybroeck, K. De Wispelaere, J. Van der Mynsbrugge, M. Vandichel, K. Hemelsoet and M. Waroquier, *Chem. Soc. Rev.*, 2014, **43**, 7326–7357.
- 40 G. Piccini, M.-S. Lee, S. F. Yuk, D. Zhang, G. Collinge, L. Kollias, M.-T. Nguyen, V.-A. Glezakou and R. Rousseau, *Catal. Sci. Technol.*, 2022, **12**, 12–37.
- 41 C. Paolucci, I. Khurana, A. A. Parekh, S. Li, A. J. Shih, H. Li, J. R. D. Iorio, J. D. Albarracin-Caballero, A. Yezerets, J. T. Miller, W. N. Delgass, F. H. Ribeiro, W. F. Schneider and R. Gounder, *Science*, 2017, **357**, 898–903.
- 42 K. De Wispelaere, P. N. Plessow and F. Studt, *ACS Phys. Chem. Au*, 2022, **2**, 399–406.
- 43 F. Berger, M. Rybicki and J. Sauer, *ACS Catal.*, 2023, **13**, 2011–2024.
- 44 Q. Ren, M. Rybicki and J. Sauer, *J. Phys. Chem. C*, 2020, **124**, 10067–10078.
- 45 T. J. Goncalves, P. N. Plessow and F. Studt, *ChemCatChem*, 2019, **11**, 4368–4376.
- 46 F. Berger and J. Sauer, *Angew. Chem.*, 2021, **60**, 3529–3533.
- 47 C. Bouchy, G. Hastoy, E. Guillon and J. A. Martens, *Sci. Technol. Energy Transition*, 2009, **64**, 91–112.
- 48 G. Kresse and J. Hafner, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1993, **47**, 558–561.
- 49 G. Kresse and J. Hafner, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1994, **49**, 14251–14269.
- 50 G. Kresse and J. Furthmüller, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1996, **54**, 11169–11186.
- 51 P. E. Blöchl, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1994, **50**, 17953–17979.
- 52 G. Kresse and D. Joubert, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1999, **59**, 1758–1775.
- 53 L. Smith, A. Davidson and A. Cheetham, *Catal. Lett.*, 1997, **49**, 143–146.
- 54 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
- 55 S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787–1799.
- 56 M. Sprik and G. Ciccotti, *J. Chem. Phys.*, 1998, **109**, 7737–7744.
- 57 E. Carter, G. Ciccotti, J. T. Hynes and R. Kapral, *Chem. Phys. Lett.*, 1989, **156**, 472–477.
- 58 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 59 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.
- 60 E. Caldeweyher, C. Bannwarth and S. Grimme, *J. Chem. Phys.*, 2017, **147**, 034112.
- 61 E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth and S. Grimme, *J. Chem. Phys.*, 2019, **150**, 154122.
- 62 E. Caldeweyher, J.-M. Mewes, S. Ehlert and S. Grimme, *Phys. Chem. Chem. Phys.*, 2020, **22**, 8499–8512.
- 63 S. N. Steinmann and C. Corminboeuf, *J. Chem. Phys.*, 2011, **134**, 044117.
- 64 A. Tkatchenko and M. Scheffler, *Phys. Rev. Lett.*, 2009, **102**, 073005.
- 65 T. Bučko, S. Lebègue, J. Hafner and J. G. Ángyán, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, **87**, 064110.
- 66 T. Bučko, S. Lebègue, J. Hafner and J. G. Ángyán, *J. Chem. Theory Comput.*, 2013, **9**, 4293–4299.
- 67 T. Bučko, S. Lebègue, J. G. Ángyán and J. Hafner, *J. Chem. Phys.*, 2014, **141**, 034114.
- 68 A. Tkatchenko, R. A. DiStasio, R. Car and M. Scheffler, *Phys. Rev. Lett.*, 2012, **108**, 236402.
- 69 A. Ambrosetti, A. M. Reilly, J. DiStasio, A. Robert and A. Tkatchenko, *J. Chem. Phys.*, 2014, **140**, 8.



- 70 T. Bučko, S. Lebègue, T. Gould and J. G. Ángyán, *J. Phys.: Condens. Matter*, 2016, **28**, 045201.
- 71 T. Gould, S. Lebègue, J. G. Ángyán and T. Bučko, *J. Chem. Theory Comput.*, 2016, **12**, 5920–5930.
- 72 M. Gajdoš, K. Hummer, G. Kresse, J. Furthmüller and F. Bechstedt, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2006, **73**, 045112.
- 73 M. Shishkin and G. Kresse, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2006, **74**, 035101.
- 74 M. Kaltak, J. Klimeš and G. Kresse, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2014, **90**, 054115.
- 75 M. Rupp, *Int. J. Quantum Chem.*, 2015, **115**, 1058–1073.
- 76 S. De, A. P. Bartók, G. Csányi and M. Ceriotti, *Phys. Chem. Chem. Phys.*, 2016, **18**, 13754–13769.
- 77 A. P. Bartók, R. Kondor and G. Csányi, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, **87**, 184115.
- 78 K. M. Jablonka, D. Ongari, S. M. Moosavi and B. Smit, *Chem. Rev.*, 2020, **120**, 8066–8129.
- 79 M. Gešvandtnerová, D. Rocca and T. Bučko, *J. Catal.*, 2021, **396**, 166–178.
- 80 A. Pohorille, C. Jarzynski and C. Chipot, *J. Phys. Chem. B*, 2010, **114**, 10235–10253.
- 81 J. Heyd, G. E. Scuseria and M. Ernzerhof, *J. Chem. Phys.*, 2003, **118**, 8207–8215.
- 82 J. Paier, R. Hirschl, M. Marsman and G. Kresse, *J. Chem. Phys.*, 2005, **122**, 234102.
- 83 J. Heyd, G. E. Scuseria and M. Ernzerhof, *J. Chem. Phys.*, 2006, **124**, 219906.
- 84 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
- 85 J. Paier, M. Marsman and G. Kresse, *J. Chem. Phys.*, 2007, **127**, 024103.

