



**HAL**  
open science

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

Jérôme Rey, Céline Chizallet, Dario Rocca, Tomáš Bučko, Michael M Badawi

► **To cite this version:**

Jérôme Rey, Céline Chizallet, Dario Rocca, Tomáš Bučko, Michael M Badawi. Reference-Quality Free Energy Barriers in Catalysis from Machine Learning Thermodynamic Perturbation Theory. *Angewandte Chemie International Edition*, 2024, 63 (6), pp.e202312392. 10.1002/anie.202312392 . hal-04504311

**HAL Id: hal-04504311**

**<https://ifp.hal.science/hal-04504311>**

Submitted on 14 Mar 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

## Catalysis

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

Jérôme Rey, Céline Chizallet,\* Dario Rocca, Tomáš Bučko,\* and Michael Badawi\*

**Abstract:** For the first time, we report calculations of the free energies of activation of cracking and isomerization reactions of alkenes that combine several different electronic structure methods with molecular dynamics simulations. We demonstrate that the use of a high level of theory (here Random Phase Approximation—RPA) is necessary to bridge the gap between experimental and computed values. These transformations, catalyzed by zeolites and proceeding via cationic intermediates and transition states, are building blocks of many chemical transformations for valorization of long chain paraffins originating, e.g., from plastic waste, vegetable oils, Fischer–Tropsch waxes or crude oils. Compared with the free energy barriers computed at the PBE+D2 production level of theory via constrained ab initio molecular dynamics, the barriers computed at the RPA level by the application of Machine Learning thermodynamic Perturbation Theory (MLPT) show a significant decrease for isomerization reaction and an increase of a similar magnitude for cracking, yielding an unprecedented agreement with the results obtained by experiments and kinetic modeling.

applicable solution, especially in catalysis.<sup>[1–5]</sup> First principles calculations indeed provide an access to invaluable microscopic information about the reaction mechanisms, the key intermediates, and tentative values for rate constants.<sup>[6,7]</sup> However, the lack of precision of first principles kinetic constants is still a serious obstacle that limits the straightforward use of theoretical data. According to the Eyring equation, a deviation of +10 kJ/mol on the free energy of activation at 500 K leads to deviations in rate constants by about one order of magnitude, with tremendous consequences on kinetics prediction. With this in mind, it is clear that approaching chemical accuracy (deviation with respect to experiment within 4.2 kJ/mol) becomes an urgent need.

The limitations due to insufficient accuracy of electronic structure methods are particularly serious for the simulations of complex chemical systems. This requires the use of large atomistic models, for which the straightforward combination of high level of theory (such as the Coupled-Cluster CC approaches<sup>[8]</sup> or Random Phase Approximation RPA)<sup>[9,10]</sup> with advanced sampling methods, such as ab initio molecular dynamics (AIMD),<sup>[11–15]</sup> is completely out of reach with the computer power available today. This is in spite of the tremendous progress we have seen in recent years, which allowed us, on the one hand, to apply accurate high-level theory methods to ever larger systems and, on the other hand, to make the advanced AIMD simulations more efficient. The combination of these two aspects<sup>[16–18]</sup> would be particularly necessary in important cases where the dynamical effects are crucial, e.g., when the interaction between species is not dominated by strong covalent bonds or when non-classical transition states are involved.

The acid-catalyzed transformations of hydrocarbons by zeolites are among reactions with strong dynamical effects whose modelling requires a careful choice of the level of

## Introduction

The prediction of chemical and biological processes is an essential but challenging task. In addition to traditional kinetic modeling based on certain pre-selected kinetic schemes, and fitting of rate constants from a set of training experimental data, the construction of kinetic models from ab initio calculations is becoming a realistic and practically

[\*] Dr. J. Rey, Dr. D. Rocca, Prof. M. Badawi  
Laboratoire de Physique et Chimie Théoriques LPCT UMR  
7019-CNRS, Université de Lorraine,  
Vandœuvre-lés-Nancy, France  
E-mail: michael.badawi@univ-lorraine.fr

Dr. C. Chizallet  
IFP Energies nouvelles, Rond-Point de l'Échangeur de Solaize,  
BP3, 69360 Solaize, France  
E-mail: celine.chizallet@ifpen.fr

Dr. T. Bučko  
Department of Physical and Theoretical Chemistry, Faculty of  
Natural Sciences, Comenius University in Bratislava,  
Ilkovičova 6, SK-84215 Bratislava, Slovakia

and  
Institute of Inorganic Chemistry, Slovak Academy of Sciences,  
Dúbravská cesta 9, SK-84236 Bratislava, Slovakia  
E-mail: bucko19@uniba.sk

Prof. M. Badawi  
Laboratoire Lorrain de Chimie Moléculaire L2CM UMR  
7053-CNRS, Université de Lorraine,  
Metz, France  
E-mail: michael.badawi@univ-lorraine.fr

© 2023 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

theory. These transformations are crucial from both a fundamental and an industrial perspective. In particular, alkene isomerization and cracking (often put into practice from alkanes, first dehydrogenated by a metallic phase<sup>[19]</sup>) play a key role in the production of chemicals and fuels from conventional and renewable sources, and are promising reactions for plastics recycling.<sup>[20–24]</sup> These reaction networks are based on carbenium chemistry. Carbenium ions can indeed be obtained by protonation of alkenes, thanks to the bridging Si-OH-Al groups of proton-exchanged zeolites. As it was shown by AIMD,<sup>[14,25–29]</sup> these species are highly mobile in the porosity of zeolites. Therefore, locating the correct reaction intermediates and quantifying their free energies using conventional static calculations fail in such reactions due to the significant contribution of a large ensemble of configurations. On the other hand, usual generalized gradient approximation (GGA) approaches, which we are currently restricted to in AIMD calculations of these systems, have been shown to be unreliable in relevant hydrocarbon transformation reactions in proton-exchanged zeolites.<sup>[17,30–34]</sup> High level static calculations<sup>[30–35]</sup> and first order thermodynamic perturbation theory on molecular dynamics simulations<sup>[17]</sup> showed that the level of theory is crucial when computing enthalpies of adsorption and free energies of activation involving  $\pi$ -complexes and cations in zeolites. Static calculations at the PBE+D2 level of theory reveal a very large relative stabilization of cations compared to bound species but dynamics and entropic effects tend to decrease this difference.<sup>[34,35]</sup> Herein, we address the questions of the dynamical effects and of the quality of the level of theory for electronic calculations via full free energy perturbation theory.

As a prototypical case, we selected the relevant elementary cracking and isomerization reaction steps over the chabazite zeolite,<sup>[27–29]</sup> for which a large set of AIMD data has been previously collected at the GGA level. The so-called type B isomerization, with a non-classical protonated cyclopropane (PCP) ion as a transition state, induces a change in the branching degree (Figure 1a). Type B<sub>1</sub> cracking converts a secondary carbenium ion into an alkene plus a smaller tertiary cation via  $\beta$ -scission. The transition state of this cracking reaction was shown to be highly dynamic, with a free rotation along the breaking C–C bond (Fig-

ure 1b).<sup>[29]</sup> Incorporation of the GGA rate constants in a single-event kinetic model was helpful to construct a robust kinetic model, but adjustment of some inaccurate rate constants turned out to be necessary to reproduce the experimentally observed relative selectivities of cracking and isomerization reactions.<sup>[36]</sup> This adjustment strategy provides us with the reference target rate constant ratio to be obtained theoretically and indicates that the difference between the free energy of activation for type B isomerization reactions minus type B<sub>1</sub> cracking is close to +10 kJ/mol. The latter value was overestimated at the PBE+D2 level of theory.

Herein, we make the combination of a high level of theory with AIMD feasible thanks to a recently developed Machine Learning thermodynamic Perturbation Theory (MLPT),<sup>[37–39]</sup> which is a post-processing approach, that improves description of isomerization and cracking barriers of heptenes catalyzed by a proton-exchanged zeolite to a so far unreached level of accuracy. The target level of theory chosen is RPA, that has provided very accurate results for the quantification of the adsorption energy of various molecules in zeolites.<sup>[40,41]</sup> The accuracy of RPA for barrier heights of reactions catalyzed by porous materials is not well established in the previous literature; however, systematic benchmark on molecular reaction barriers has shown that the RPA is significantly more accurate than semi-local functionals, hybrid functionals, and wave function methods such as MP2.<sup>[42]</sup> Based on our tests, we estimate that obtaining a free energy profile for an elementary step at the RPA level typically requires factor 10<sup>4</sup> longer computational time than GGA, indicating that it would take almost one millennium to obtain a single AIMD-quality rate constant at the RPA level. Since MLPT requires only a limited number of fixed-geometry RPA calculations, the simulation time can be reduced to a few weeks or even a few days depending on the available computational resource. In spite of promising achievements with simpler approaches or systems,<sup>[16,17,43]</sup> this work is, to the best of our knowledge, the first report of activation free energy barriers obtained by combining a high post-DFT level of theory with AIMD for such complex reactions.

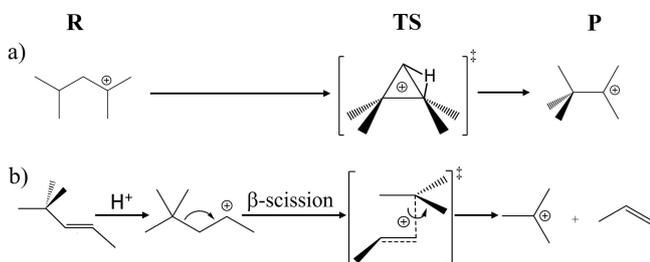
## Methodology

### MLPT approach

As shown in Ref. [38], the free energy of activation at the target level of theory ( $\Delta\tilde{A}^\ddagger$ ) can be determined from that obtained at other, typically less CPU-intensive, production method ( $\Delta A^\ddagger$ ) via free energy perturbation theory<sup>[44]</sup> as follows:

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

where the terms  $\Delta A_{TS}$  and  $\Delta A_R$  are the corrections to free energy of the transition state and reactant, respectively. It is a remarkable property of Eq. (1) that while the calculation



**Figure 1.** Reactants (R), transition states (TS) and products (P) of the two reactions investigated in this work: a) type B isomerization between dibranched and tribranched species, b) type B<sub>1</sub>  $\beta$ -scission cracking reaction. Both reactions are catalyzed by proton-exchanged zeolite chabazite.

of  $\Delta A^\ddagger$  requires lengthy MD simulations performed for several intermediate states along the reaction coordinate  $\zeta(\mathbf{q})$ , the change of this quantity due to changing the Hamiltonian from  $H(\mathbf{p}, \mathbf{q})$  to  $\tilde{H}(\mathbf{p}, \mathbf{q}) = H(\mathbf{p}, \mathbf{q}) + \Delta V(\mathbf{q})$  (with  $\mathbf{p}$  and  $\mathbf{q}$  being, respectively, momenta and Cartesian coordinates of all atoms defining the system) can be computed by considering the contributions from the initial (reactant, R) and final (transition state, TS) points only. These contributions write:

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

and

$$\Delta A_R = -k_B T \ln \langle \exp[-\Delta V(\mathbf{q})/k_B T] \rangle, \quad (3)$$

where  $k_B$  is the Boltzmann constant,  $T$  is temperature,  $\langle \dots \rangle$  and  $\langle \dots \rangle_{\xi^*}$  stand for the NVT ensemble averages of the quantity enclosed that are determined for the reactant and transition state, respectively. In our simulation protocol based on blue moon ensemble approach,<sup>[45,46]</sup> the ensemble of transition states is obtained in constrained MD, in which  $\zeta(\mathbf{q})$  is fixed at the value  $\xi^*$  characteristic for the transition state.

The determination of well converged  $\Delta A_{TS}$  and  $\Delta A_R$  in a straightforward fashion can be very time consuming as it may require as many as tens of thousands of the  $\tilde{V}(\mathbf{q})$  calculations. This would be a serious problem when performing calculations for a CPU intensive high level target method, such as the RPA considered in this work. To this end, the use of the  $\Delta$ -machine learning<sup>[47]</sup> turns out to be very useful. In practice, we only need to perform explicit target method energy calculations for a relatively small number (several tens to few hundreds) of selected configurations (training set) and, afterwards, the machine learning procedure employing the kernel ridge regression<sup>[48]</sup> with the REMatch kernel<sup>[49]</sup> and the SOAP (smooth overlap of atomic positions) descriptors<sup>[50]</sup> is used to predict  $\tilde{V}(\mathbf{q})$  for remaining members of the ensembles representing reactant or transition states. Since the MLPT is required to provide reliable predictions for the initial and final points of the transformation path, but not for any intermediate state, two independent models describing R and TS can be trained. This is a great advantage compared to other ML-based strategies building global models for whole potential energy landscapes.<sup>[51–53]</sup>

However, it should be mentioned that the MLPT can yield reliable results only if the configuration spaces sampled under the action of Hamiltonians of the target and production methods overlap.<sup>[54]</sup> As discussed in a recent work by Herzog et al.,<sup>[39]</sup> this can be reliably measured by a simple metrics  $I_w$ . Importantly, a good overlap for the PBE + D2 production and RPA target methods applied to zeolitic systems was reported.<sup>[39]</sup>

### Simulation details

All calculations presented in this work have been performed using the periodic DFT program VASP,<sup>[55–57]</sup> utilizing the projector-augmented wave (PAW) method implemented by Blochl and Kresse et al.<sup>[58,59]</sup> The Kohn–Sham equations were solved self-consistently with a convergence criterion of  $10^{-7}$  eV/cell. The sampling of the Brillouin zone was restricted to the  $\Gamma$ -point as we used a rather large supercell. The plane wave cutoff was set to 400 eV. A Gaussian smearing of 0.05 eV was applied. The supercell of chabazite used in the calculations is described in Sec. S1.A of the Supporting Information. The ab initio molecular dynamics calculations have been performed using the leap-frog algorithm with integration step of 1 fs, whereby the mass of hydrogen was set to that of tritium. The simulation temperature of 500 K, which was inspired by relevant experimental conditions,<sup>[36]</sup> was fixed using the Andersen thermostat, with a collision frequency of  $0.01 \text{ fs}^{-1}$ .<sup>[60]</sup> We emphasize that all the explicit MD calculations have been performed only at the production method level, while the target methods were used to perform up to 225 single point calculations per state. The latter were subsequently used for the training and testing of the ML model (see Sec. S1.B and C in the Supporting Information for details).

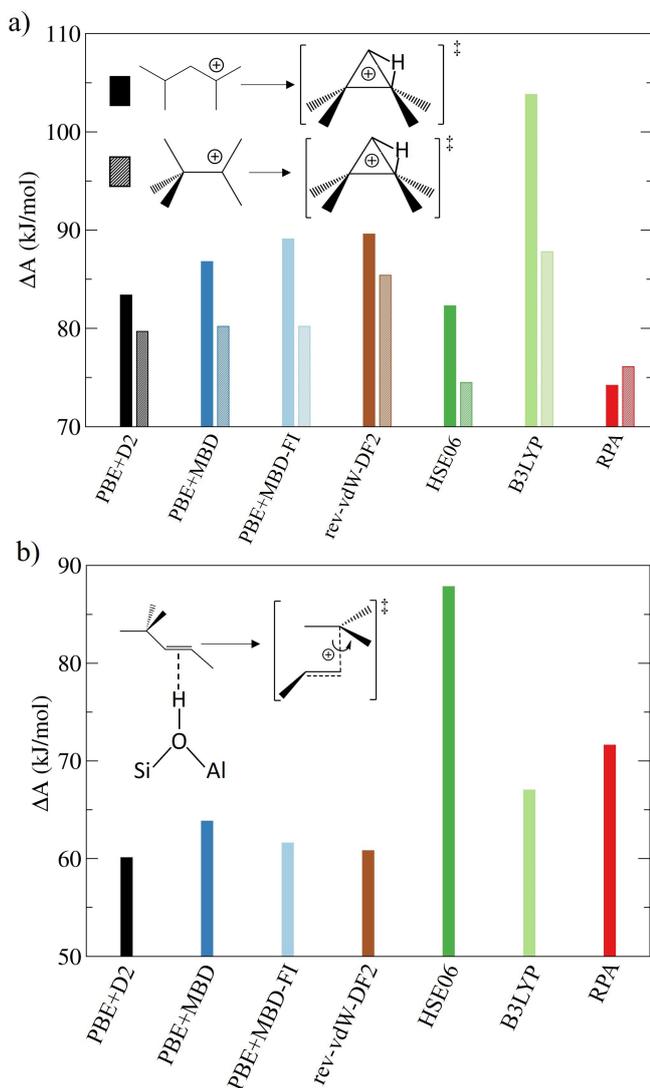
As a production method, the PBE functional<sup>[61]</sup> with the D2 dispersion<sup>[62]</sup> for long range dispersion interactions was used. These calculations were reported in the work previously published by some of us.<sup>[27–29]</sup> In addition to the Random Phase Approximation<sup>[9,10]</sup> (RPA), as implemented in VASP 5.4.4.,<sup>[63–65]</sup> several density function approximations (DFAs) from different rungs of the Jacob's ladder of DFT functionals<sup>[66]</sup> have been selected as target methods. In particular, we considered the generalized gradient DFA with sophisticated treatment of dispersion interaction PBE + MBD<sup>[67]</sup> and PBE + MBD-FI,<sup>[68]</sup> a non local dispersion corrected functional rev-vdW-DF2,<sup>[69]</sup> hybrid functionals HSE06<sup>[70–72]</sup> and B3LYP.<sup>[73,74]</sup>

## Results and Discussion

### Type B isomerization between dibranched and tribranched tertiary cations

The type B isomerization reaction connecting a tertiary dibranched cation (reactant R) and a tertiary tribranched cation (product P) proceeds via a transition state in which an edge protonated cyclopropane is formed (Figure 1a). This particular type B isomerization step was chosen among a large ensemble of possible type B isomerization reactions (see Ref. [36] for an exhaustive list) because of the availability of PBE + D2 AIMD data (Ref. [27]) and because it 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),<sup>[36]</sup> whereas all other skeletons of carbenium ions are generated in multiple elementary steps.

As shown in Figure 2a, the free energy barriers for the forward and reverse reaction steps are significantly affected by the choice of methodology. Compared to the PBE+D2 results (83.4 and 79.7 kJ/mol for the forward and reverse modes, respectively), the PBE+MBD and PBE+MBD-FI methods yield small increase (within 5.7 kJ/mol) in the barrier of forward and negligible change in the barrier of reverse steps. In the case of non local functional rev-vdW-DF2, both barriers are consistently higher by ~6 kJ/mol. Interestingly, the two hybrid functionals considered in this work yield distinctly different results. While the HSE06 method tends to slightly lower both barriers (by 1.1 and 5.4 kJ/mol for forward and reverse modes, respectively), the B3LYP method leads to their significant increase (by 20.4 and 8.1 kJ/mol for forward and reverse modes, respectively).



**Figure 2.** Effect of the level of theory on the free energy barrier computed by MLPT of a) the isomerization reaction between tertiary dibranched and tribranched cations (dark bars: isomerization between tertiary dibranched to tribranched cations; light bars: isomerization between tertiary tribranched to dibranched cations, thus the backward reaction) and b) the type B<sub>1</sub> cracking reaction.

We note, however, that the quality of the MLPT calculations for the the hybrid functionals is compromised because of a poor overlap of the underlying configurational spaces (see the “Improvement of the accuracy at the RPA target level of theory” section below and Tables S1–S3 in the Supporting Information). Finally, the RPA barriers are 9.2 and 3.7 kJ/mol lower compared to the PBE+D2 values (the comparison of the free energy profiles determined at the RPA and PBE+D2 levels are shown in Figure S2 of the Supporting Information). As evident from data presented in Tables S6–S8, these decreases are primarily related to a relative stabilization of the transition state (an edge protonated cyclopropane), and to a lesser extent of the product (a tertiary tribranched cation), with respect to the reactant (a tertiary dibranched cation). These trends are in qualitative agreement with the results of the gas phase calculations of Sandbeck et al. Ref. [75], where a decrease in a barrier of a reaction analogous to the isomerization studied in our work (called ‘3C→1C’ in their work) by 4.1 kJ/mol was reported when DFT PBE was replaced by a high level CCSD(T) method.

The trends discussed above can be estimated from electronic energy differences determined for fixed geometries, e.g., the stationary points identified at the PBE+D2 level. Such an approach is a basis for various hybrid correction schemes proposed in the literature.<sup>[32,35]</sup> Albeit being qualitatively correct in most cases, such an estimation is insufficient for reaching the level of accuracy required in modern quantum chemical calculations, as evident from comparison of Figure 2 and Figure S3. Nevertheless, the qualitative agreement between the MLPT and the static single-point results suggests that the reported free energy changes are primarily due to the changes in exchange and correlation energy and to a lesser extent to geometry changes caused by the choice of the electronic structure method.

### Cracking reaction

The type B<sub>1</sub> cracking reaction starts with the protonation of a  $\pi$ -complex into a short-lived secondary cation, which in turn undergoes the  $\beta$ -scission yielding the tert-butyl cation and propene (see Figure 1b). This reaction is considered as irreversible<sup>[36]</sup> since the probability of recombination of the two product fragments is very low under the relevant operating conditions. Whatever the level of theory, MLPT predicts an increase of the cracking barrier with respect to the PBE+D2 estimate (Figure 2b and section SIII). The extent of the increase, however, strongly depends on the level of theory. Very moderate changes are computed for the dispersion-corrected PBE functionals and with the rev-vdW-DF2 functionals, at most 3.7 kJ/mol. More drastic changes are observed for hybrid functionals, but, in contrast to the isomerization barriers, HSE06 leads to the largest increase (27.8 kJ/mol). Also in this case, the quality of the estimation using the hybrid functionals is poor due to a too small overlap of the configurational spaces (see Tables S14 and S15). The MLPT calculations at the RPA target level

lead to a significant increase in the free energy barrier of activation, from 60.1 to 71.6 kJ/mol. This increase is related to a larger stabilization of the reactant (a neutral  $\pi$ -complex) than the  $\beta$ -scission transition state (a positively charged species, see Figure S4), as evident from the data presented in Tables S11 and S12 which show that the decrease in free energy of reactant is  $\sim 13$  kJ/mol larger than that for TS when PBE+D2 is replaced by RPA. Furthermore, it is evident from a comparison of the data presented in Tables S6 and S11 that RPA stabilizes the neutral reactant of the cracking reaction  $\sim 21$  kJ/mol more strongly than the cationic reactant of the isomerization reaction. At the same time the stabilization of the transition states of both reactions is very similar, whereby the difference in  $A(\text{RPA}) - A(\text{PBE} + \text{D2})$  by MLPT is only  $\sim 1$  kJ mol $^{-1}$  (cf. data in Tables S7 and S12). Further tests involving a larger number of reactions will be needed to determine whether the observed stabilization of neutral states by RPA is a general trend or is specific to the particular reactions studied in this work.

### Comparison with reference calculations in the literature

Thanks to the MLPT approach, we find systematic trends in terms of the respective stabilization of neutral versus charged compounds that agree reasonably well with the previous reports on related systems.

Indeed, using high level hybrid QM:QM static calculations (with PBE+D2 as low-level method for periodic calculations, MP2 as a mid-level, and CCSD(T) as the highest level method for cluster calculations), Berger et al.<sup>[76]</sup> showed that for reactions in zeolites starting with protonation of alkanes, PBE+D2 underestimates the enthalpy of activation by 40 to 65 kJ/mol. Using the same hybrid scheme, Ren et al.<sup>[35]</sup> showed that carbenium ions are over-stabilized by 25 to 50 kJ/mol (in free energy) with respect to  $\pi$ -complexes at the PBE+D2 level compared to reference CCSD(T) cluster calculations. These differences between the semilocal DFT and high level wave-function based calculations are attributed to the self-interaction error of PBE functional that causes over-stabilization of the cations.

To compare the stability of  $\pi$ -complexes and carbocations in acid zeolites, de Wispelaere et al.<sup>[17]</sup> used an approach based on the first-order approximation to free-energy perturbation theory: a high level, wave-function method (CCSD(T)) static correction is computed for a single configuration and applied to the free energy obtained by DFTMD. In this way, the authors found that the stability of the carbenium ion is overestimated by around 50 kJ/mol by DFT-MD.

For reactions starting from neutral  $\pi$ -complexes and passing through a cationic transition state we found the same tendency as these authors, although the magnitude of the changes predicted by MLPT at the RPA level is much smaller (increase of +11.5 kJ/mol for the cracking reaction). One of the likely reasons why this increase of the barrier is less important than in static calculations of Ref. [76] is that the dynamic (entropic) effects stabilize the mobile cationic

species with respect to the bound -complexes.<sup>[25,28,35]</sup> These entropic effects cannot be accurately captured by static calculations, whereas our MLPT method relies on a rigorous application of free energy perturbation theory to the well-converged DFT-MD trajectories of reactant and transition states. On the other hand, a part of the difference between the barrier height reported here and in Ref. [76] is also certainly due to different accuracy of the RPA and CCSD(T) methods, and to the difference in nature in terms of reaction (alkane cracking is studied in Ref. [76]).

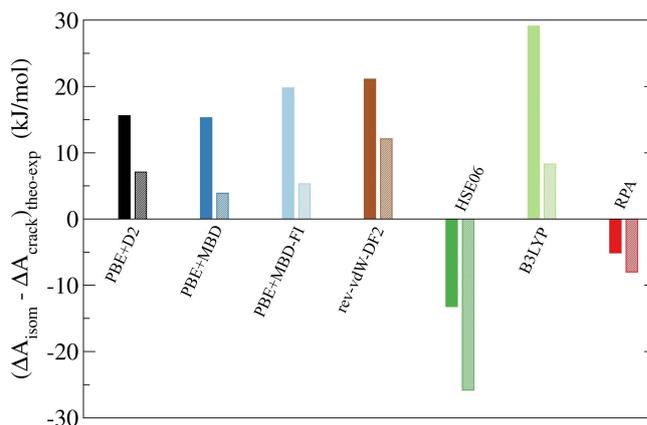
For the isomerization reaction, the free energy barriers separating the classical tertiary cation and the edge PCP decrease slightly at the RPA level of theory. This trend has already been reported for transformations in gas phase of similar cations with 6 carbon atoms studied using static calculations,<sup>[75]</sup> where a small decrease (2 to 4 kJ/mol) was observed when the CCSD(T) level of theory replaced the PBE method. This result can be attributed to the correction of the errors of the PBE functional (self-interaction and over-delocalization errors).

Thus, the evolution of the free energy profiles that we have computed, from the PBE+D2 level to the RPA level, is qualitatively consistent with previous works, although none of these previous works has used the MLPT method, nor has it accounted for the two key reaction steps (isomerization and cracking) in a consistent manner.

### Improvement of the accuracy at the RPA target level of theory

We can now compare our estimates at various levels of theory with experimental observations, to assess the validity of the method proposed and identify the required level of theory to reach chemical accuracy. Figure 3 summarizes our main findings in that respect.

In a previous experimental study, analyzed thanks to ab initio-based kinetic modelling,<sup>[36]</sup> some PBE+D2 rate con-



**Figure 3.** Deviation with respect to experiment<sup>[36]</sup> of the difference of the free energies of activation at the various levels of theory investigated between the isomerization and cracking reactions (dark bars: isomerization between tertiary dibranched to tribranched cations; light bars: isomerization between tertiary tribranched to dibranched cations, thus the backward reaction).

stants had to be adjusted in order to achieve a good agreement between the theoretical and experimentally measured data. In particular, the  $B_1$  cracking barrier was increased by 15 kJ/mol in the kinetic model so as to correctly reproduce the experimental cracking/isomerization selectivity. From this analysis, we extracted reference values for the difference between the type B isomerization and  $B_1$  cracking free energy barriers, as detailed in Supporting Information SIV. Thus, the free energy barrier differences between type B isomerization and  $B_1$  cracking are expected to be 7.7 and 12.5 kJ/mol for the forward and backward isomerization reactions, respectively. These values serve as reference in Figure 3.

As previously discussed, the MLPT prediction obtained with hybrid functionals cannot be trusted because of a too small  $I_w$  index, showing an insufficient overlap between configurational spaces. Our exploratory tests performed with the D3(BJ) method<sup>[77,78]</sup> show that the long-range dispersion corrections have a little impact on this result. For GGAs and the nonlocal rev-vdW-DF2 functionals, the MLPT correction can be considered accurate but at these levels of theory the predicted change in the barrier is too small to improve the agreement with the trend observed in kinetic modeling: a small increase of the free energy barriers of +3/+6 kJ/mol is observed for the isomerization between tertiary cations (see section “Type B isomerization between dibranched and tribranched tertiary cations” above) as well as a small increase (below 5 kJ/mol) for cracking reaction. Thus, there is no significant improvement over the PBE+D2 data as compared to the experiment-based reference.

At the target RPA level of theory, the overlap index  $I_w$  between configurational spaces is  $I_w \approx 0.1$  (see Tables S6 to S8, S11, and S12) which is adequate for an accurate MLPT estimate. In the end, the RPA prediction of the difference of barriers between isomerization and cracking gives the best agreement with experiments, with deviations of 5.1 and 8.0 kJ/mol (forward and backward, Figure 3).

PBE+MBD and PBE+MBD-FI predictions are slightly better (deviation of 3.9 kJ/mol in absolute value for PBE+MBD, 5.2 kJ/mol for PBE+MBD-FI, versus 8 kJ/mol for RPA) for the backward reaction, but far worse (deviation of 15.3 kJ/mol in absolute value for PBE+MBD, 19.8 kJ/mol for PBE+MBD-FI, versus 5.1 kJ/mol for RPA) for the forward reaction. This makes us discard PBE+MBD and PBE+MBD-FI as best levels of theory tested. The fact that RPA deviates more from experiments in the case of the backward reaction than in the case of the forward reaction, may be related to the uncertainty in the experimental data themselves. Indeed, as mentioned previously, the forward reaction is the only elementary step that leads to formation of 2,2,3-trimethylbutane. The backward reaction produces 2,4-dimethylpentane, but other type B reactions yield to the same skeleton in the reaction network,<sup>[36]</sup> making the extraction of uncorrelated intrinsic kinetic parameters more difficult.

We can conclude that for these reactions, a high level of theory such as RPA is required to obtain an accurate estimate of the free energy barrier of activation and a good agreement with experiment. Thanks to the MLPT method

targeting the RPA level of theory, we have significantly improved the accuracy while taking into account dynamic effects, which would not have been feasible without MLPT.

## Conclusion

In the present work, we took up the challenge of calculating ab initio activation free energy barriers of isomerization and cracking of alkenes in protonic zeolites at a reference quality level (RPA) while accounting for dynamic effects using AIMD. This task, unfeasible by conventional computational approaches, has been implemented for the first time using thermodynamic perturbation theory assisted by machine learning. For isomerization steps starting from cations, a decrease of the free energy barrier is observed, while an increase is found for cracking reaction. The magnitude of this effect (correction of 15–21 kJ/mol to the activation barrier differences obtained at the PBE+D2 level) is in quantitative agreement with experimental results and their kinetic modelling (about 15 kJ/mol correction). This unique combination of high level electronic calculations (RPA) and high quality sampling of the potential energy surface (by blue moon sampling at the PBE+D2 level) is now possible. It is shown with such an approach, the RPA level is necessary to approach chemical accuracy, as GGAs, local dispersion corrected functionals, and hybrid functionals fail in this respect. Although the MLPT methodology is presented here for a specific case of hydrocarbon conversion reactions, it can be directly applied to any other type of chemical reaction.

## 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 performed using the computational resources procured in the slovak 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 20142020 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. The authors are grateful to Dr. Mauricio Chagas Da Silva from Université de Lorraine for fruitful discussions.

## Conflict of Interest

The authors declare no conflict of interest.

## Data Availability Statement

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

**Keywords:** ab initio molecular dynamics · catalysis · hydrocracking · machine learning · zeolites

- 
- [1] M. Saliccioli, M. Stamatakis, S. Caratzoulas, D. Vlachos, *Chem. Eng. Sci.* **2011**, *66*, 4319.
- [2] M. Stamatakis, *J. Phys. Condens. Matter* **2015**, *27*, 013001.
- [3] M. K. Sabbe, M.-F. Reyniers, K. Reuter, *Catal. Sci. Technol.* **2012**, *2*, 2010.
- [4] A. Bruix, J. T. Margraf, M. Andersen, K. Reuter, *Nat. Catal.* **2019**, *2*, 659.
- [5] C. Chizallet, *Top. Catal.* **2022**, *65*, 69.
- [6] W. Thiel, *Angew. Chem. Int. Ed.* **2014**, *53*, 8605.
- [7] B. W. J. Chen, L. Xu, M. Mavrikakis, *Chem. Rev.* **2021**, *121*, 1007.
- [8] R. J. Bartlett, M. Musiał, *Rev. Mod. Phys.* **2007**, *79*, 291.
- [9] D. Bohm, D. Pines, *Phys. Rev.* **1953**, *92*, 609.
- [10] D. Langreth, J. Perdew, *Solid State Commun.* **1975**, *17*, 1425.
- [11] A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562.
- [12] P. Vidossich, A. Lledós, G. Ujaque, *Acc. Chem. Res.* **2016**, *49*, 1271.
- [13] M. Bocus, L. Vanduyfhuys, F. De Proft, B. M. Weckhuysen, V. Van Speybroeck, *JACS Au* **2022**, *2*, 502.
- [14] V. Van Speybroeck, K. De Wispelaere, J. Van der Mynsbrugge, M. Vandichel, K. Hemelsoet, M. Waroquier, *Chem. Soc. Rev.* **2014**, *43*, 7326.
- [15] G. Piccini, M.-S. Lee, S. F. Yuk, D. Zhang, G. Collinge, L. Kollias, M.-T. Nguyen, V.-A. Glezakou, R. Rousseau, *Catal. Sci. Technol.* **2022**, *12*, 12.
- [16] F. Berger, M. Rybicki, J. Sauer, *ACS Catal.* **2023**, *13*, 2011.
- [17] K. De Wispelaere, P. N. Plessow, F. Studt, *ACS Phys. Chem. Au* **2022**, *2*, 399.
- [18] V. Van Speybroeck, M. Bocus, P. Cnudde, L. Vanduyfhuys, *ACS Catal.* **2023**, *13*, 11455.
- [19] K. Cheng, L. C. J. Smulders, L. I. van der Wal, J. Oenema, J. D. Meeldijk, N. L. Visser, G. Sunley, T. Roberts, Z. Xu, E. Doskocil, H. Yoshida, Y. Zheng, J. Zeevi, P. E. de Jongh, K. P. de Jong, *Science* **2022**, *377*, 204.
- [20] C. Bouchy, G. Hastoy, E. Guillon, J. A. Martens, *Oil Gas Sci. Technol., Rev. IFP* **2009**, *64*, 91.
- [21] J. A. Martens, D. Verboekend, K. Thomas, G. Vanbutsele, J.-P. Gilson, J. Pérez-Ramírez, *ChemSusChem* **2013**, *6*, 421.
- [22] J. Zecevic, G. Vanbutsele, K. P. de Jong, J. A. Martens, *Nature* **2015**, *528*, 245.
- [23] A. J. Martín, C. Mondelli, S. D. Jaydev, J. Pérez-Ramírez, *Chem* **2021**, *7*, 1487.
- [24] C. Chizallet, C. Bouchy, K. Larmier, G. Pirngruber, *Chem. Rev.* **2023**, *123*, 6107.
- [25] P. Cnudde, K. De Wispelaere, J. Van der Mynsbrugge, M. Waroquier, V. Van Speybroeck, *J. Catal.* **2017**, *345*, 53.
- [26] P. Cnudde, K. De Wispelaere, L. Vanduyfhuys, R. Demuyneck, J. Van der Mynsbrugge, M. Waroquier, V. Van Speybroeck, *ACS Catal.* **2018**, *8*, 9579.
- [27] J. Rey, A. Gomez, P. Raybaud, C. Chizallet, T. Bučko, *J. Catal.* **2019**, *373*, 361.
- [28] J. Rey, P. Raybaud, C. Chizallet, T. Bučko, *ACS Catal.* **2019**, *9*, 9813.
- [29] J. Rey, C. Bignaud, P. Raybaud, T. Bučko, C. Chizallet, *Angew. Chem. Int. Ed.* **2020**, *132*, 19100.
- [30] C. Tuma, T. Kerber, J. Sauer, *Angew. Chem. Int. Ed.* **2010**, *49*, 4678.
- [31] G. Piccini, M. Alessio, J. Sauer, *Angew. Chem. Int. Ed.* **2016**, *55*, 5235.
- [32] M. Rybicki, J. Sauer, *J. Am. Chem. Soc.* **2018**, *140*, 18151.
- [33] T. J. Goncalves, P. N. Plessow, F. Studt, *ChemCatChem* **2019**, *11*, 4368.
- [34] F. Berger, J. Sauer, *Angew. Chem. Int. Ed.* **2021**, *60*, 3529.
- [35] Q. Ren, M. Rybicki, J. Sauer, *J. Phys. Chem. C* **2020**, *124*, 10067.
- [36] J.-M. Schweitzer, J. Rey, C. Bignaud, T. Bučko, P. Raybaud, M. Moscovici-Mirande, F. Portejoie, C. James, C. Bouchy, C. Chizallet, *ACS Catal.* **2022**, *12*, 1068.
- [37] B. Chehaibou, M. Badawi, T. Buko, T. Bazhurov, D. Rocca, *J. Chem. Theory Comput.* **2019**, *15*, 6333.
- [38] T. Bučko, M. Gešvandtnerová, D. Rocca, *J. Chem. Theory Comput.* **2020**, *16*, 6049.
- [39] B. Herzog, M. Chagas da Silva, B. Casier, M. Badawi, F. Pascale, T. Bučko, S. Lebégue, D. Rocca, *J. Chem. Theory Comput.* **2022**, *18*, 1382.
- [40] D. Rocca, A. Dixit, M. Badawi, S. Lebégue, T. Gould, T. Bučko, *Phys. Rev. Mater.* **2019**, *3*, 040801.
- [41] F. Göttl, A. Grüneis, T. Bučko, J. Hafner, *J. Chem. Phys.* **2012**, *137*, 114111.
- [42] X. Ren, P. Rinke, G. E. Scuseria, M. Scheffler, *Phys. Rev. B* **2013**, *88*, 035120.
- [43] G. Piccini, M. Parrinello, *J. Phys. Chem. Lett.* **2019**, *10*, 3727.
- [44] R. W. Zwanzig, *J. Chem. Phys.* **1954**, *22*, 1420.
- [45] E. A. Carter, G. Ciccotti, J. T. Hynes, R. Kapral, *Chem. Phys. Lett.* **1989**, *156*, 472.
- [46] G. Ciccotti, M. Sprik, *J. Chem. Phys.* **1998**, *109*, 7737.
- [47] R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. von Lilienfeld, *J. Chem. Theory Comput.* **2015**, *11*, 2087.
- [48] M. Rupp, *Int. J. Quantum Chem.* **2015**, *115*, 1058.
- [49] S. De, A. P. Bartók, G. Csányi, M. Ceriotti, *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754.
- [50] A. P. Bartók, R. Kondor, G. Csányi, *Phys. Rev. B* **2013**, *87*, 184115.
- [51] L. Shen, J. Wu, W. Yang, *J. Chem. Theory Comput.* **2016**, *12*, 4934.
- [52] L. Shen, W. Yang, *J. Chem. Theory Comput.* **2018**, *14*, 1442.
- [53] P. Li, X. Jia, X. Pan, Y. Shao, Y. Mei, *J. Chem. Theory Comput.* **2018**, *14*, 5583.
- [54] A. Pohorille, C. Jarzynski, C. Chipot, *J. Phys. Chem. B* **2010**, *114*, 10235.
- [55] G. Kresse, J. Hafner, *Phys. Rev. B* **1993**, *47*, 558.
- [56] G. Kresse, J. Hafner, *Phys. Rev. B* **1994**, *49*, 14251.
- [57] G. Kresse, J. Furthmüller, *Phys. Rev. B* **1996**, *54*, 11169.
- [58] P. E. Blöchl, *Phys. Rev. B* **1994**, *50*, 17953.
- [59] G. Kresse, D. Joubert, *Phys. Rev. B* **1999**, *59*, 1758, no. 3.
- [60] H. C. Andersen, *J. Chem. Phys.* **1980**, *72*, 2384.
- [61] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865.
- [62] S. Grimme, *J. Comput. Chem.* **2006**, *27*, 1787.
- [63] M. Gajdo, K. Hummer, G. Kresse, J. Furthmüller, F. Bechstedt, *Phys. Rev. B* **2006**, *73*, 045112.
- [64] M. Shishkin, G. Kresse, *Phys. Rev. B* **2006**, *74*, 035101.
- [65] M. Kaltak, J. Klime, G. Kresse, *Phys. Rev. B* **2014**, *90*, 054115.
- [66] J. P. Perdew, K. Schmidt, *AIP Conf. Proc.* **2001**, *577*, 1.
- [67] T. Bučko, S. Lebégue, T. Gould, J. G. Ángyán, *J. Phys. Condens. Matter* **2016**, *28*, 045201.

- [68] T. Gould, S. Lebégue, J. G. Ángyán, T. Bučko, *J. Chem. Theory Comput.* **2016**, *12*, 5920.
- [69] I. Hamada, *Phys. Rev. B* **2014**, *89*, 121103.
- [70] J. Heyd, G. E. Scuseria, M. Ernzerhof, *J. Chem. Phys.* **2003**, *118*, 8207.
- [71] J. Heyd, G. E. Scuseria, M. Ernzerhof, *J. Chem. Phys.* **2006**, *124*, 219906.
- [72] J. Paier, R. Hirschl, M. Marsman, G. Kresse, *J. Chem. Phys.* **2005**, *122*, 234102.
- [73] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623.
- [74] J. Paier, M. Marsman, G. Kresse, *J. Chem. Phys.* **2007**, *127*, 024103.
- [75] D. J. S. Sandbeck, D. J. Markewich, A. L. L. East, *J. Org. Chem.* **2016**, *81*, 1410.
- [76] F. Berger, M. Rybicki, J. Sauer, *J. Catal.* **2021**, 395, 117.
- [77] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
- [78] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456.

Manuscript received: August 23, 2023

Accepted manuscript online: December 6, 2023

Version of record online: December 28, 2023