Unraveling the nature of adsorbed isobutene in H–SSZ–13 with operando simulations at the top of Jacob's ladder

Massimo Bocus,[+,a] Sander Vandenhaute,[+,a] Veronique Van Speybroeck*[a]

Unraveling the nature of adsorbed olefins in zeolites is crucial to understand numerous zeolite-catalyzed processes. A well-grounded theoretical description critically depends on both an accurate determination of the potential energy surface (PES) and a reliable account of entropic effects at operating conditions. Herein, we propose a transfer learning approach to perform random phase approximation (RPA) quality enhanced sampling molecular dynamics simulations, thereby approaching chemical accuracy on both the determination and exploration of the PES. The proposed methodology is used to investigate isobutene adsorption in H–SSZ–13 as prototypical system to estimate the relative stability of physisorbed olefins, carbenium ions and surface alkoxide species (SAS) in Brønsted-acidic zeolites. We show that the *tert*-butyl carbenium ion formation is highly endothermic and no entropic stabilization is observed compared to the physisorbed complex within H–SSZ–13. Hence, its predicted concentration and lifetime are negligible, making a direct experimental observation unlikely. Yet, it remains a shallow minimum on the free energy surface over the whole considered temperature range (273–873 K), being therefore a short-lived reaction intermediate rather than a transition state species.

Introduction

Zeolites are microporous, crystalline aluminosilicate materials. The charge imbalance created by the defective substitution of Si^{4+} with Al^{3+} can be compensated by the presence of an acidic proton (the Brønsted acid site, or BAS for short) on one of the oxygen atoms in the Al first coordination sphere. The BAS acidity, the strong confinement effects of the micro-sized pores and their remarkable stability have made Brønsted-acidic zeolites pivotal catalysts in the traditional petrochemical industry as well as in novel processes for the conversion of biomass-derived sustainable feedstocks[1,2]. Most acid-catalyzed reactions in zeolites are triggered by an initial transfer of the BAS from the framework to the reacting substrate, potentially creating a positively charged intermediate. Therefore, understanding the chemical nature and reactivity of the latter is of crucial interest to obtain a comprehensive mechanistic view of the chemical process under investigation[3].

[a] Dr. M. Bocus, S. Vandenhaute, Prof. Dr. V. Van Speybroeck*
Center for Molecular Modeling, Ghent University, Technologiepark 46, 9052 Zwijnaarde (Belgium)
E-mail: veronique.vanspeybroeck@ugent.be

[+] These authors contributed equally.

In this context, a particularly challenging case study is the adsorption of olefins in zeolites (Figure 1 a). When unsaturated hydrocarbons approach the BAS, the double C=C bond interacts with the electrophilic proton forming a so-called π-complex. A complete proton transfer from the zeolite to the molecule creates a positively-charged carbenium ion, which can then chemisorb on one of the oxygen atoms in the first coordination sphere of the Al site to form a surface alkoxide species (SAS). The persistent nature of each of these adsorbed species at operating conditions has been the topic of extensive theoretical and experimental studies[3]. It is generally accepted that their relative stability is affected by a series of factors, including temperature, zeolite topology and substitution degree of the double bond which, in general, create a subtle balance between enthalpic and entropic contributions. Enthalpically, SAS would be preferred but their stability may decrease due to steric hindrance in small concave environments[4,5]. Carbenium ions are thought to be entropically stabilized, however such stabilization may strongly depend on the zeolite topology. Thus far, disentangling entropic and enthalpic effects has remained challenging for both theory and experiment. In this contribution, we study isobutene in H–SSZ–13 (CHA topology) to reveal the persistent nature of its various adsorbed states. This case is selected as isobutene is the smallest alkene with a tertiary carbon atom and the limited unit cell size of H–SSZ–13 allows for the highest level of theory calculations.

A large number of theoretical works have investigated isobutene chemisorption in zeolites, initially using gas-phase clusters[8–15] and then moving to more realistic, fully periodic models. These can be inquired with static approaches, only relying on a limited number of points on the potential energy surface (PES)[4,16–29], or with dynamic approaches, that better mimic operating conditions[5,30,31]. A more detailed overview of the theoretical investigations in periodic models performed thus far is taken up in Table S1 of the Supplementary Information. On the one hand, seminal work of Sauer and co-workers clearly shows that standard generalized gradient approximation density functional theory (GGA DFT), with or without empirical dispersion corrections, is unreliable when modeling species with a net charge separation such as carbenium ions in zeolites[20,22,29]. High-level (at least MP2) single point corrections based on cluster models increase the carbenium ion energy by tens of $kJ{\cdot}mol^{-1}$ with respect to the neutral alkene. On the other hand, typical industrial processes involving olefins—such as fluid catalytic cracking[32] (FCC)—are conducted at very high temperatures ($\sim$ 800 K). Static simulations with harmonic thermal corrections are not well-suited to model mobile and weakly bounded species in the zeolite voids[33,34]. Therefore, studies conducted with *operando* methodologies based on constant temperature molecular dynamics (MD)

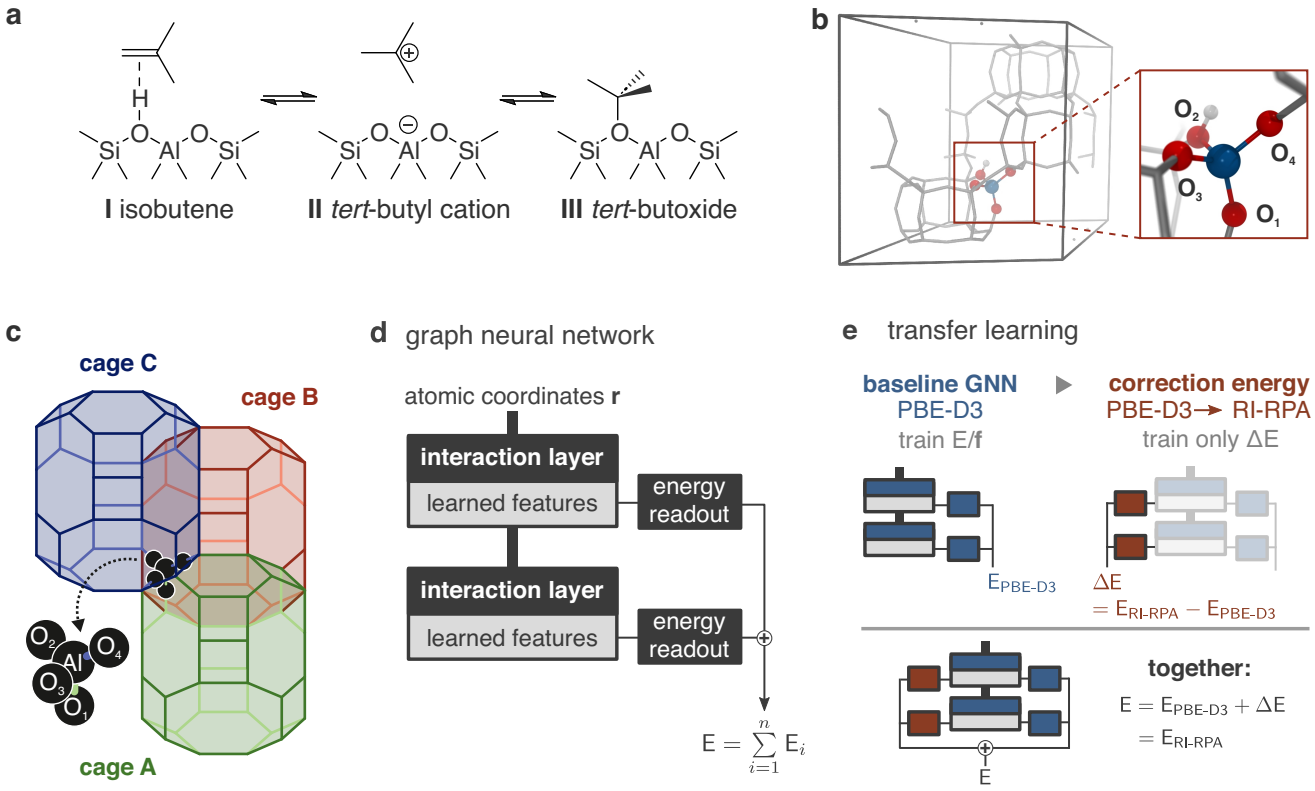


Figure 1. a. The three investigated intermediates formed upon isobutene adsorption in Brønsted acidic zeolites. b. The conventional H–SSZ–13 unit cell used in this work. The inset shows the nomenclature of the four oxygen atoms in the first coordination sphere of the Al defect adopted herein. Al is depicted in blue, O in red and H in white, while the rest of the framework is shown in grey sticks for the sake of clarity. c. Schematic depiction of the three nonequivalent H–SSZ–13 cages. The location of the Al tetrahedron is highlighted with grey spheres. d. Simplified representation of the MACE architecture[6]; the atomic coordinates are parsed in atomic environments which are passed through a series of interaction layers; this yields a set of features based on which the atomic energies are computed. e. Summary of the two-stage procedure proposed in this contribution: the interaction layers and the first set of energy readouts (blue) are trained based on PBE-D3-level reference data – energy and forces – obtained using active learning[7]. A second set of readouts (red) are trained on a small number of RI-RPA *energies*. See the Supplementary Sections 2.5 and 2.6 for more information.

suggest that thermal effects can remarkably stabilize the mobile carbenium ion, making it a prominent intermediate at reaction conditions[5,30]. This effect may however be highly dependent on the zeolite topology[5]. While MD-based approaches seem attractive to unravel the nature of adsorbed species at operating conditions, they require hundreds of thousands energy and forces evaluations, making it impossible to go beyond GGA DFT when evaluating the PES.

The extent to which carbenium ions are relevant intermediates in high temperature processes such as FCC has therefore remained unclear. Various experimental attempts have been performed to identify the predominant intermediate(s); because some studies report data in favor of transient tertiary carbenium ions[25,35–38] while others did not detect them[39–43] or did not provide definitive evidence[44], the experimental results remain inconclusive[3]. From a theoretical viewpoint, attempts have been made to reconcile enhanced sampling molecular dynamics with high-level calculations, using for instance corrections based on single-point cluster evaluations[31], machine learning assisted free energy perturbation theory[45,46], or similar methods[47]. These approaches have however been shown to produce wrong results when the high- and low-level method do not explore sufficiently overlapping regions in phase space[48], and were still based on *ab initio* molecular dynamics.

To fully solve these issues, we harness the power of state-of-the-art machine learning potentials (MLPs) and propose an efficient approach based on transfer learning[49–52]. Starting from a machine learning model trained on a production level of theory, we augment its accuracy towards a higher-level electronic structure method with outstanding data efficiency. We showcase the power of this methodology by investigating the conversion of isobutene in *tert*-butylcation and surface *tert*-butoxide species in the H–SSZ–13 zeolite with RI-RPA[53,54] (random phase approximation using a resolution-of-identity approach) quality enhanced sampling molecular dynamics (ESMD) simulations. Our results, arising from a total of 60 ns RI-RPA quality ESMD, show that within H–SSZ–13 the *tert*-butylcation is not entropically stabilized compared to the physisorbed neutral isobutene within the temperature range 323–823 K. By approaching chemical accuracy in both the enthalpy and entropy estimates of the reaction, physisorbed isobutene is predicted to be the only species present in significant amounts within the zeolite pores while the concentrations of *tert*-butylcation and *tert*-butoxide are negligible. The carbenium ion remains an elusive intermediate, with a computed lifetime that is however too short to be observed with currently available experimental techniques.

Results and Discussion

The proposed transfer learning approach is used to investigate the adsorption of isobutene in the H–SSZ–13 zeolite. The framework presents a relatively small unit cell where all tetrahedral atoms are equivalent by symmetry. We use the same equilibrated conventional unit cell of our previous investigation on the proton hopping reaction [55]. A single Si atom is substituted with Al (Si/Al = 35), breaking the unit cell symmetry and creating four nonequivalent oxygen atoms in the first coordination of the Al defect and three nonequivalent cages in which the isobutene can be located (Figure 1 b,c). The isobutene conversion in the *tert*-butylcation and the surface *tert*-butoxide is captured by means of two collective variables (CVs, Figure 2 a). CVs are functions of the atomic coordinates which should be able to describe the progress of the reaction under investigation. One CV describes the proton transfer from the zeolite to isobutene with a difference in coordination numbers (CNs). The first CN is between the primary carbon atoms (C_p) and all hydrogen atoms in the system, while the second between the four zeolite oxygen atoms in the first coordination sphere of the Al defect (O_z) and, again, all hydrogen atoms. The other CV is a CN between the tertiary carbon atom (C_t) and O_z (additional details are reported in Supplementary Section 2.3). Two assumptions will be made to keep the complexity of the problem manageable: (1) the formation of the primary surface isobutoxide species can be ignored. The reaction is known to proceed through a high energy barrier [22,24] and does not lead to the formation of a carbenium ion, the main focus of this study. (2) It will be assumed that the reactivity of isobutene is entirely independent of inter-cage diffusion and that it does not prefer to reside in any specific cage. An integrated reaction-diffusion model is beyond the scope of this study and does not impact its conclusions.

An accurate computational estimate of both enthalpic and entropic contributions to the reaction free energy requires millions of energy and force evaluations. We employ state of the art graph neural networks (GNNs) to learn a computationally efficient representation of the *ab initio* potential energy surface in order to be able to simulate the dynamics of the system on such a long time scale. The incorporation of equivariant feature representations into GNNs has reduced the required training set sizes to about a few thousand structures for a chemical reaction in zeolites [55], each of which is to be labeled with the correct *ab initio* energy and atomic forces [7]. The key challenge is that we cannot generate this amount of training data for the required (post-HF) level of theory. First, (RI-)MP2 or RPA energy evaluations require at least two orders of magnitude more memory and compute time than regular GGA DFT. Second, the calculation of the nuclear gradients introduces additional memory requirements and is moreover very hard to converge with respect to basis set and grid size. To overcome these issues, we developed a transfer learning approach which (1) reduces the required number of post-HF single-point evaluations by an order of magnitude, and (2) only requires the post-HF energy during training (and not the nuclear gradients, i.e. forces). The approach is schematically illustrated in Figure 1 (d, e), and is founded on the empirical observation that the interaction layers within the GNN implicitly learn an informative feature representation for the chemical environment of each atom in the system [51,52,56]. In particular, we observe that if we train a baseline GNN model to PBE-D3(BJ) energy and forces, the resulting interaction layers can be used to train a small additional energy correction ΔE which captures the difference between the learned PBE-D3(BJ) baseline output and the target post-HF energy. Importantly, the prediction of this ΔE requires only a small number of additional weights ($\sim 1\%$ extra), and can be learned based on just a few hundred *energy-only* post-HF single points—see Supplementary Section 2.6 for more information.

Based on our benchmarks, we determined RI-RPA/cc-QT(Al,Si)Z as our target level of theory (see Supplementary Section 2.2), whereas the baseline level of theory was PBE-D3(BJ). We employ a combination of active learning and enhanced sampling to train a baseline GNN model to PBE-D3(BJ) reference data using `psiflow` [7,57,58]. Quantum mechanical evaluations are performed with `CP2K` [59], using either a MOLOPT-TZVP or correlation-consistent cc-QT(Al,Si)Z basis set for respectively the PBE-D3 and RI-RPA levels of theory [60–62]. Umbrella sampling simulations are used to sample the phase space in the active learning loop as well as in the production runs to derive all free energy surfaces (FESs), which are computed with the weighted histogram analysis (WHAM) method as implemented in the `ThermoLIB` library [63]. All computational details are reported in Supplementary Section 2, whereas Supplementary Section 3 presents an extensive validation of the entire workflow. In particular, we show that (1) the baseline GNN perfectly reproduces the PBE-D3(BJ) reaction free energy profile as obtained from brute force AIMD umbrella sampling, and (2) that our transfer learning approach allows us to achieve an RMSE of only 0.3 meV/atom on the target RI-RPA potential energy surface. Nevertheless, the entire active learning workflow requires only 3682 PBE-D3(BJ) single-point calculations, with an additional ∼200 *energy-only* RI-RPA calculations to parameterize the ΔE prediction. Hereafter, all results of PBE-D3(BJ) quality are obtained using the vanilla baseline GNN. The results of RI-RPA quality are obtained with the same GNN augmented with an additional set of readout layers that predicts ΔE (Figure 1 e).

As previously mentioned, the introduction of the Al substitution breaks the CHA unit cell symmetry and creates 4 nonequivalent O atoms in the first coordination sphere and 3 nonequivalent cages (Figure 1 b,c). In principle, isobutene can be located in any of the 3 cages and the BAS/*tert*-butoxide can be bound to any of the 4 oxygen atoms. While some combinations can be excluded by simple steric considerations (for instance, when the BAS is on O_3 it is exclusively accessible from cage C), we tested all remaining paths at 623 K to select the most favorable one(s) (Supplementary Section 4.1). At the end, the two most promising ones were selected: in the first, isobutene is located in cage B and the BAS/*tert*-butoxide are formed on O_4. This path exhibits the lowest free energy barriers for both reactive steps. However, based on our previous investigation of the proton hopping reaction in H–SSZ–13 [55], O_4 is also the least populated site—which is in line with a higher activity. Hence, we also selected an additional path in which isobutene resides in cage C, the BAS is located on O_3 and the *tert*-butoxide is formed on O_2. This path presents moderate free energy barriers and, moreover, O_3 was also computed to be the most populated site.

The two paths are investigated with both PBE-D3(BJ) and RI-RPA quality MLP US simulations over the temperature range 323–823 K in steps of 100 K. While a com-

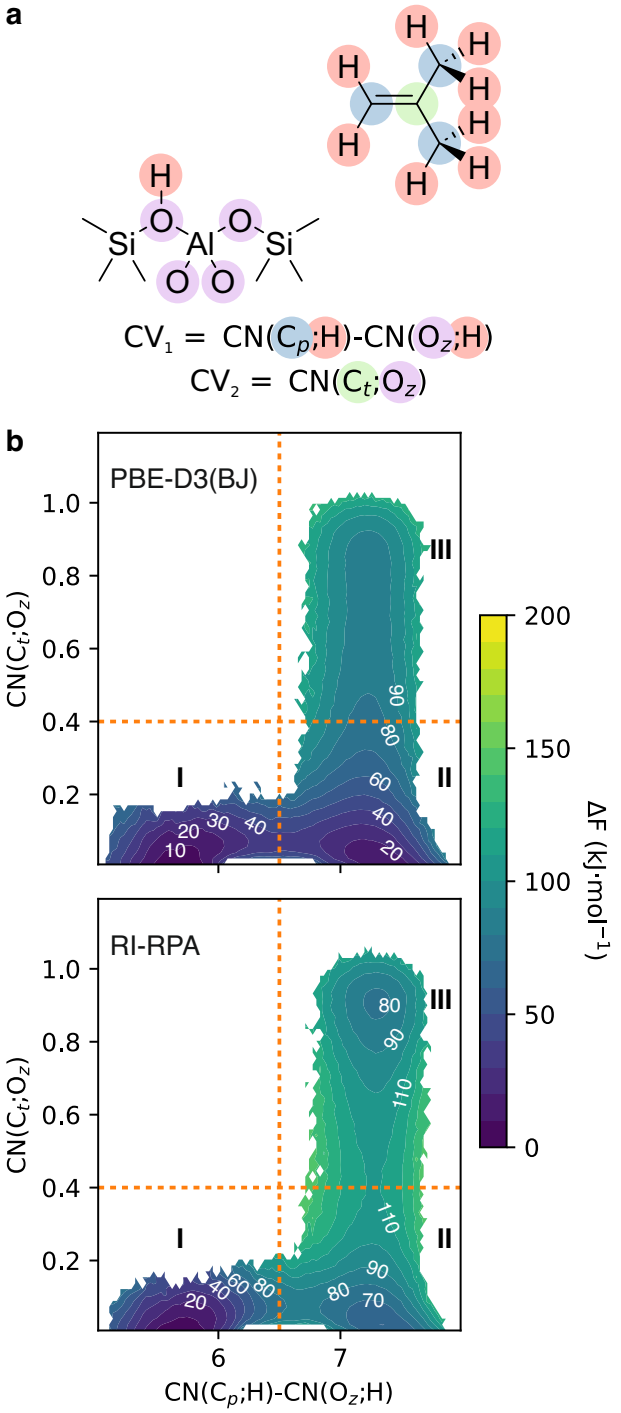


Figure 2. a. Graphical depiction of the atoms included in the CNs that constitute the two final CVs used in the US simulations. b. The computed FES changes remarkably when moving from the PBE-D3(BJ) to the RI-RPA PES. Showcased as examples are the 623 K FESs for isobutene conversion in cage B. The orange lines highlight the FES partition in the regions corresponding to isobutene (**I**), to the *tert*-butylcation (**II**) and to the *tert*-butoxide (**III**). A complete overview of the computed FESs is provided in Supplementary Section 4.2.

plete overview of all the computed FESs and free energy barriers is reported in Supplementary Section 4.2, we showcase the 623 K FESs of cage B in Figure 2 b to illustrate the main differences between PBE-D3(BJ) and RI-RPA, as these can be generalized to both cages and all temperatures.

Starting from the protonation reaction (**I**→**II**), a more accurate level of theory compensates—as expected—the deficiencies of GGA DFT. Indeed, the raw free energy barrier increases from $\sim$ 30 kJ·mol^{-1} to $\sim$ 80 kJ·mol^{-1} and the *tert*-butylcation is lifted in energy with respect to isobutene, going from $\sim$ +20 kJ·mol^{-1} to $\sim$ +70 kJ·mol^{-1}. This shift reflects the one observed for the 0 K electronic energy and agrees with the reference results of Plessow et al.[28] (Supplementary Section 3.1). Yet, the *tert*-butylcation remains a minimum on the FES at all temperatures, indicating that it is a transient reaction intermediate rather than a transition state. Quite some difference is also seen in the FES region concerning the *tert*-butoxide formation (**II**→**III**). While PBE-D3(BJ) does not predict the *tert*-butoxide to be a minimum, the RI-RPA FES has a clear transition state between the two intermediates. The overall stability of the alkoxide is similar for both levels of theory ($\sim$ +80 kJ·mol^{-1}), in line with the 0 K results (Supplementary Section 3.1).

From the computed 2-dimensional FESs a CV-independent free energy for an intermediate $\mathcal{A}$ can be retrieved by integration over the respective CV region:

$$F_{\mathcal{A}} = -k_{\mathrm{B}}T \log\left(\int_{\mathcal{A}} \exp\left(-\frac{F(q_1,q_2)}{k_{\mathrm{B}}T}\right) \mathrm{d}q_1 \mathrm{d}q_2\right) \qquad (1)$$

q_1 and q_2 are the two CVs, T is the reaction temperature and k_{B} Boltzmann's constant. The boundaries used to define the three regions associated with isobutene, *tert*-butylcation and *tert*-butoxide are highlighted with orange dotted lines in Figure 2 b. This allowed us to compute the free energy differences between the three intermediates for both cages and both levels of theory over the whole temperature range. Remark that for PBE-D3(BJ) the surface *tert*-butoxide (**III**) is not a minimum on the FES, *i.e.* there is no free energy barrier separating it from **II**. Hence, its computed free energy is arbitrary as the state boundaries are ill-defined. However, it can still provide an useful mean of comparison with the RI-RPA results where the state is metastable. The results are shown in Figure 3 a–d, where it can be seen how the free energy differences have in all cases an almost perfect linear behavior over the considered temperature range. This allowed us to rely on the simple relation

$$\Delta F_{\mathcal{A}\to\mathcal{B}} = \Delta U_{\mathcal{A}\to\mathcal{B}} - T\Delta S_{\mathcal{A}\to\mathcal{B}} \qquad (2)$$

and perform linear regression to estimate the internal energy ($\Delta U_{\mathcal{A}\to\mathcal{B}}$) and entropy ($\Delta S_{\mathcal{A}\to\mathcal{B}}$) variations of the reactions. Given the linear behavior of the data points, these variations can be approximated as temperature-independent over the considered temperature range.

Starting from the formation of the *tert*-butylcation (**I**→**II**), PBE-D3(BJ) predicts a small positive internal energy variation in both the considered cages (+16 kJ·mol^{-1}). The main cause for the lower free energy difference observed in cage B is the entropy variation, which goes from being slightly negative in cage C to moderately positive in cage B. Moving to RI-RPA has a dramatic effect on the results: first, the internal energy difference between **I** and **II** drastically increases to more than +55 kJ·mol^{-1} in both cages. Additionally, the entropy variation decreases by $\sim$ 10 J·K^{-1}·mol^{-1} in both cages, becoming moderately negative for cage C and close to zero for cage B. In any case, our results indicate that there exists no entropic stabilization of weakly bounded carbenium ions in H–SSZ–13 compared to the physisorbed complex, and their previous observation in MD simulations[5]

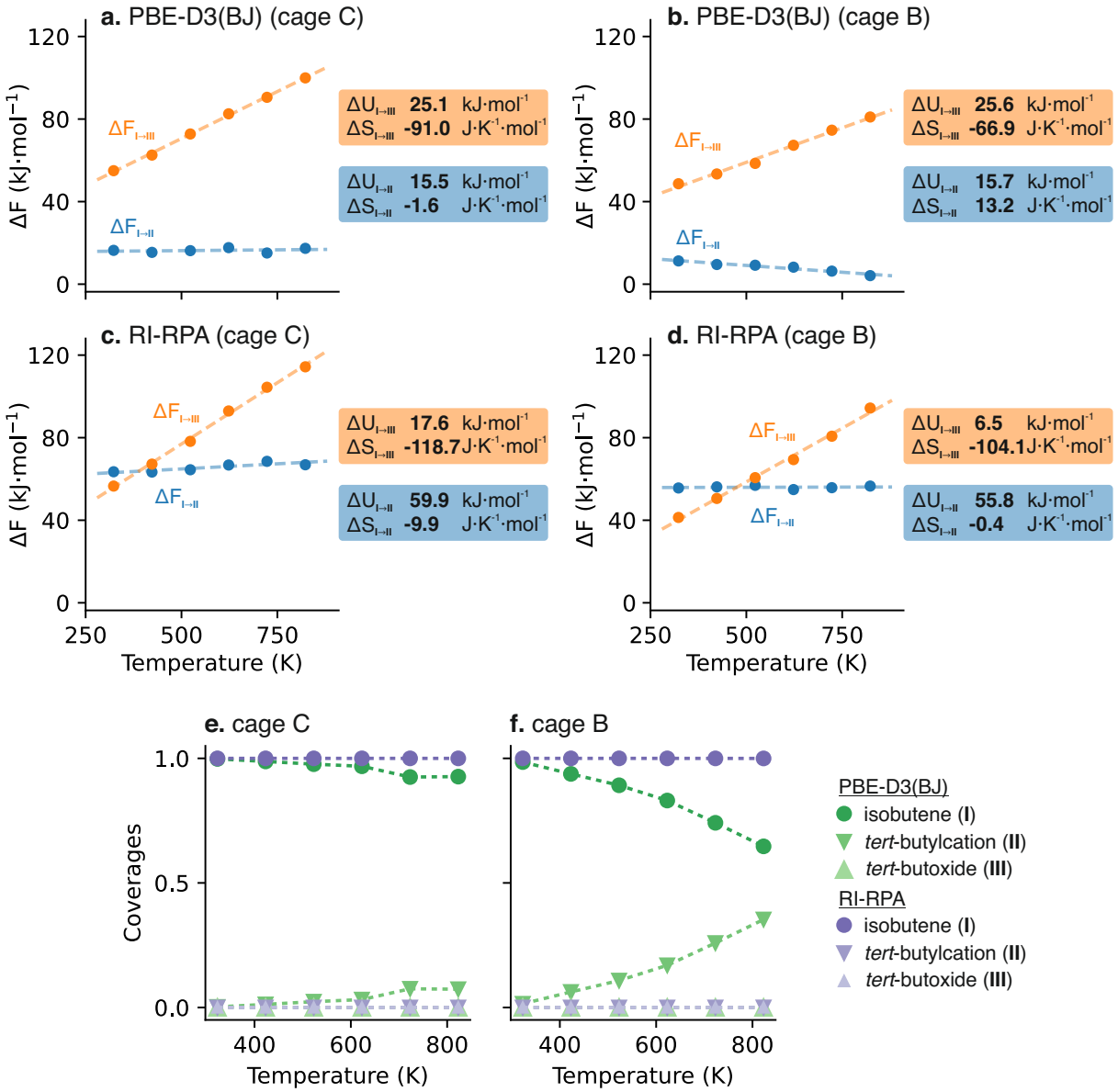


Figure 3. The *tert*-butylcation does not benefit from any entropic stabilization with respect to neutral isobutene in the H–SSZ–13 pores. a–d. Free energy difference between isobutene and *tert*-butylcation (**I**→**II**) and isobutene and *tert*-butoxide (**I**→**III**) computed with PBE-D3(BJ) and RI-RPA in cage B and cage C. The internal energy and entropy differences obtained from linear regression are also reported. e, f. Fractional coverages of the 3 intermediates derived from the PBE-D3(BJ) and RI-RPA free energy differences in cage C (e) and cage B (f).

is therefore exclusively related to the deficiencies of GGA DFT. This is made clear by observing the equilibrium population of the three intermediates (Figure 3 e,f). Especially for cage B, PBE-D3(BJ) predicts a significant concentration of *tert*-butylcation above 400 K, which also becomes easily observable in unbiased MD simulations given the underestimated free energy barrier. On the other hand, the RI-RPA results clearly foresee neutral isobutene as the only species present in the zeolite for any temperature, with the concentrations of *tert*-butylcation and *tert*-butoxide being nearly zero.

To support these conclusions, we used the MLPs to run 1 ns long unbiased MD simulations at 323, 623 and 823 K both in cage B and C (a complete overview of the results is reported in Supplementary Section 4.4). When using either isobutene or the *tert*-butylcation as initial structure, PBE-D3(BJ) allows frequent jumps between the two states, which become more frequent when increasing the temperature (Supplementary Figure S22). With RI-RPA, on the other hand, the *tert*-butylcation always deprotonates within 5 ps and neutral isobutene is the only observed species for all temperature. Only in one simulation at 823 K we observed a transient formation of *tert*-butylcation lasting for few ps, in line with the results of the enhanced sampling simulations. In order to better understand the differences in behavior between isobutene and the *tert*-butylcation, we performed two additional simulations in cage B and C at 623 K with half-quadratic potential walls to restrain the system in the *tert*-butylcation region (**II** in Figure 2 b). Interestingly, the *tert*-butylcation has a large mobility in cage B, because the Al defect is located at half the cage height. Hence, it explores a similar space volume as isobutene (Figure 4). Conversely,

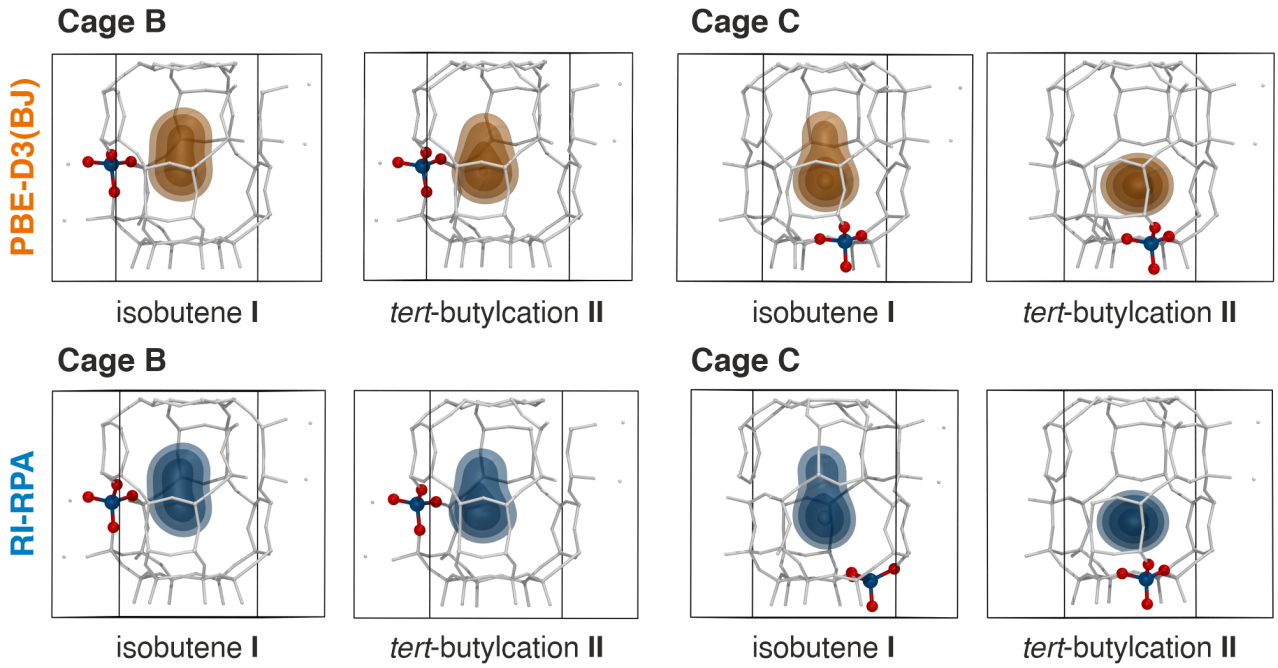


Figure 4. Density maps showing the location of the tertiary isobutene carbon atom during 1 ns long PBE-D3(BJ) and RI-RPA MD simulations of isobutene and *tert*-butylcation at 623 K. The surfaces are obtained with the `volmap` extension implemented in `VMD`[64] (resolution of 0.2 Å). The isovalues are arbitrarily chosen to be (from the exterior) 0.005, 0.01, 0.02 and 0.04. Zeolite atoms are represented with white sticks, only the Al tetrahedron is highlighted with colored spheres (Al in blue, O in red) for the sake of clarity. See also Supplementary Section 4.4.

cage C has the Al defect located at its base. While isobutene shifts between a BAS-bound π-complex and a free Van der Waals complex, the *tert*-butylcation remains in tight proximity to the negative Al. This reflects in a larger entropic penalty, which is in line with the lower $\Delta S_{\mathbf{I}\rightarrow\mathbf{II}}$ for cage C derived from enhanced sampling simulations. When comparing PBE-D3(BJ) and RI-RPA (Figure 4) no major differences can be seen, indicating that the entropy decrease of RI-RPA is not related to a reduced mobility of the intermediates in the framework and making it challenging to pinpoint a specific cause for the observation.

The formation of *tert*-butoxide (**I**→**III**) is predicted to be moderately endothermic by both levels of theory, with an internal energy variation with respect to neutral isobutene of +25 kJ·mol^{-1} in both cages according to PBE-D3(BJ) and +18 (cage C) or +7 kJ·mol^{-1} (cage B) according to RI-RPA. The difference could in this case be attributed to limitations of RI-RPA, which was shown to modestly overstabilize the surface-bonded species (Supplementary Section 3.1). As expected, the entropy variation of the reaction is in all cases largely negative, which makes the formation of *tert*-butoxide highly unfavorable at all temperatures (Figure 3 e,f).

The final question we tried to answer is whether the *tert*-butylcation, despite being clearly present in negligible concentrations in H–SSZ–13, might survive long enough to be detected experimentally. Tuma and Sauer[22] reported a half life $\tau_{1/2} = \ln 2/k_{\mathbf{II}\rightarrow\mathbf{I}}$ of 59 μs in ferrierite, which is sufficiently long for detection by fast spectroscopies such as UV-vis. To estimate a lifetime for the *tert*-butylcation in H–SSZ–13, we computed the backward kinetic constant $k_{\mathbf{II}\rightarrow\mathbf{I}}$ from our simulations within the transition state theory (TST) approximation (Supplementary Section 4.3). Our results indicate that the half life of the intermediate in H–SSZ–13 is extremely short, where for both cage B and C is in the order of hundreds of fs or few ps at most, in agreement with the results of the unbiased MD simulations (Supplementary Section 4.4). This, joined with its extremely low concentration, would make its experimental observation highly unlikely.

Conclusion

In this contribution, we propose a transfer learning approach based on state-of-the-art machine learning interatomic potentials to perform RI-RPA quality ESMD simulations. We select a relevant and challenging case study to showcase the capability of the methodology, namely the chemisorption of isobutene in the H–SSZ–13 zeolite. The reaction proceeds through a weakly bounded *tert*-butylcation intermediate, whose mobility is hardly captured by static simulations with harmonic corrections. Additionally, its charged nature makes standard GGA DFT unreliable in capturing its relative stability, with an overstabilization in the order of tens of kJ·mol^{-1}. Our approach allows for the first time to approach chemical accuracy both in the internal energy and entropy estimates with a modest computational cost.

By performing RI-RPA quality US simulations over a broad range of temperatures, we show that the entropy difference between the neutral isobutene and the *tert*-butylcation is around zero or even slightly negative independently on the level of theory. Yet, moving from PBE-D3(BJ) to RI-RPA drastically increases the endothermic nature of the

reaction, leading to equilibrium populations dominated by the neutral isobutene, while the concentrations of the *tert*-butylcation and *tert*-butoxide intermediates are negligible. Yet, the *tert*-butylcation remains a minimum on the FES for all considered temperatures, indicating that it is a transient reaction intermediate rather than a transition state species. This is in line with the available experimental evidences for H–ZSM–5, where the intermediate was captured with nucleophilic moieties [25,38]. Our estimate for the *tert*-butylcation half life is drastically lower than previous reports from static simulations in ferrierite, in the order of few picoseconds at most, making its direct experimental observation highly unlikely.

Of course, our conclusions strictly hold for H–SSZ–13 only and frameworks with different pore shapes and sizes could lead to different observations [4,5]. One of the main strength of our methodology, being based on graph neural networks, is that it provides a per-atom environment-dependent correction. Hence, as the parent model, it can be highly transferable provided that a sufficiently diverse training set is provided. Moreover, it also has the potential of being used with cluster-based corrections, where only the atoms directly involved in the correction (adsorbed molecule and active site) are modelled with the target level of theory. This could allow to model significantly larger systems (that are impossible to evaluate in full with RI-RPA) with unprecedented accuracy and further reduce the gap between theory and experiment.

Acknowledgements

The authors acknowledge funding from the Research Board of Ghent University (V.V.S.). S.V. wishes to thank the Research Foundation – Flanders (FWO) for a doctoral fellowship (grant n. 11H6821N). M.B. acknowledges financial support from the Fund for Scientific Research – Flanders and the Excellence of Science (EOS) Project BioFact (EOS ID 30902231). The resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government. Additionally, we acknowledge the EuroHPC Joint Undertaking for awarding this project access to the EuroHPC supercomputer LUMI, hosted by CSC (Finland) and the LUMI consortium, and the Luxembourg national supercomputer MeluXina through a EuroHPC Regular Access call. M.B. thanks Prof. Louis Vanduyfhuys for the assistance with the use of `ThermoLIB` and Dr. Frederick Stein and Prof. Jürg Hutter for their help with the RI-RPA calculations. S.V. wishes to thank Ben Clifford for his guidance during the development of `psiflow`.

Data Availability

An implementation of the modified MACE architecture will be made available in a public GitHub repository upon final publication; all additional input files, generated datasets, and models which were trained in this manuscript will be made available in a public Zenodo archive upon final publication.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: Carbocations • Machine Learning • Molecular Dynamics • Random Phase Approximation • Zeolites

References

[1] J. Cejka, A. Corma, S. Zones (Editors), *Zeolites and Catalysis: Synthesis, Reactions and Applications*, Wiley-VCH Verlag GmbH & Co. KGaA **2010**.

[2] T. Ennaert, J. Van Aelst, J. Dijkmans, R. De Clercq, W. Schutyser, M. Dusselier, D. Verboekend, B. F. Sels, *Chemical Society Reviews* **2016**, *45*, 584.

[3] R. Zhao, G. L. Haller, J. A. Lercher, *Microporous and Mesoporous Materials* **2023**, *358*, 112390.

[4] M. L. Sarazen, E. Iglesia, *Proceedings of the National Academy of Sciences* **2017**, *114*, E3900.

[5] P. Cnudde, M. Waroquier, V. Van Speybroeck, *Catalysis Science & Technology* **2023**, *13*, 4857.

[6] I. Batatia, D. P. Kovacs, G. Simm, C. Ortner, G. Csanyi, MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields, in S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, A. Oh (Editors), *Advances in Neural Information Processing Systems*, volume 35, Curran Associates, Inc. **2022** pages 11423–11436.

[7] S. Vandenhaute, M. Cools-Ceuppens, S. DeKeyser, T. Verstraelen, V. Van Speybroeck, *npj Computational Materials* **2023**, *9*, 1.

[8] P. Viruela-Martin, C. Zicovich-Wilson, A. Corma, *The Journal of Physical Chemistry* **1993**, *97*, 13713.

[9] A. Rigby, M. Frash, *Journal of Molecular Catalysis A: Chemical* **1997**, *126*, 61.

[10] V. Kazansky, *Catalysis Today* **1999**, *51*, 419.

[11] M. Boronat, P. Viruela, A. Corma, *Physical Chemistry Chemical Physics* **2001**, *3*, 3235.

[12] M. Boronat, A. Corma, *Applied Catalysis A: General* **2008**, *336*, 2.

[13] H. Fang, A. Zheng, S. Li, J. Xu, L. Chen, F. Deng, *The Journal of Physical Chemistry C* **2010**, *114*, 10254.

[14] H. Fang, A. Zheng, J. Xu, S. Li, Y. Chu, L. Chen, F. Deng, *The Journal of Physical Chemistry C* **2011**, *115*, 7429.

[15] Y. Chu, B. Han, A. Zheng, X. Yi, F. Deng, *The Journal of Physical Chemistry C* **2013**, *117*, 2194.

[16] P. Sinclair, A. Vries, R. van Santen, et al., *Journal of the Chemical Society, Faraday Transactions* **1998**, *94*, 3401.

[17] M. Boronat, C. M. Zicovich-Wilson, P. Viruela, A. Corma, *The Journal of Physical Chemistry B* **2001**, *105*, 11169.

[18] C. Tuma, J. Sauer, *Angewandte Chemie International Edition* **2005**, *44*, 4769.

[19] V. Nieminen, M. Sierka, D. Y. Murzin, J. Sauer, *Journal of Catalysis* **2005**, *231*, 393.

[20] C. Tuma, J. Sauer, *Physical Chemistry Chemical Physics* **2006**, *8*, 3955.

[21] B. A. De Moor, M.-F. Reyniers, M. Sierka, J. Sauer, G. B. Marin, *The Journal of Physical Chemistry C* **2008**, *112*, 11796.

[22] C. Tuma, T. Kerber, J. Sauer, *Angewandte Chemie International Edition* **2010**, *49*, 4678.

[23] N. Rosenbach Jr, A. P. dos Santos, M. Franco, C. J. Mota, *Chemical Physics Letters* **2010**, *485*, 124.

[24] C. M. Nguyen, B. A. De Moor, M.-F. Reyniers, G. B. Marin, *The Journal of Physical Chemistry C* **2012**, *116*, 18236.

[25] W. Dai, C. Wang, X. Yi, A. Zheng, L. Li, G. Wu, N. Guan, Z. Xie, M. Dyballa, M. Hunger, *Angewandte Chemie International Edition* **2015**, *54*, 8783.

[26] G. A. Ferguson, L. Cheng, L. Bu, S. Kim, D. J. Robichaud, M. R. Nimlos, L. A. Curtiss, G. T. Beckham, *The Journal of Physical Chemistry A* **2015**, *119*, 11397.

[27] T. J. Goncalves, P. N. Plessow, F. Studt, *ChemCatChem* **2019**, *11*, 4368.

[28] P. N. Plessow, F. Studt, *The Journal of Physical Chemistry Letters* **2020**, *11*, 4305.

[29] Q. Ren, M. Rybicki, J. Sauer, *The Journal of Physical Chemistry C* **2020**, *124*, 10067.

[30] P. Cnudde, K. De Wispelaere, J. Van der Mynsbrugge, M. Waroquier, V. Van Speybroeck, *Journal of Catalysis* **2017**, *345*, 53.

[31] K. De Wispelaere, P. N. Plessow, F. Studt, *ACS Physical Chemistry Au* **2022**, *2*, 399.

[32] E. T. Vogt, B. M. Weckhuysen, *Chemical Society Reviews* **2015**, *44*, 7342.

[33] G. Collinge, S. F. Yuk, M.-T. Nguyen, M.-S. Lee, V.-A. Glezakou, R. Rousseau, *ACS Catalysis* **2020**, *10*, 9236.

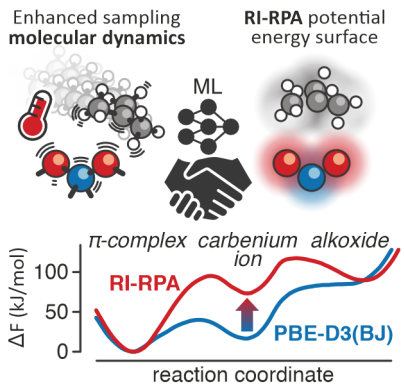
[34] V. Van Speybroeck, M. Bocus, P. Cnudde, L. Vanduyfhuys, *ACS Catalysis* **2023**, *13*, 11455.

[35] A. G. Stepanov, K. I. Zamaraev, J. M. Thomas, *Catalysis Letters* **1992**, *13*, 407.

[36] A. G. Stepanov, K. I. Zamaraev, *Catalysis Letters*

1993, *19*, 153.
[37] A. G. Stepanov, *Catalysis Today* **1995**, *24*, 341.
[38] A. G. Stepanov, M. V. Luzgin, V. N. Romannikov, V. N. Sidelnikov, K. I. Zamaraev, *Journal of Catalysis* **1996**, *164*, 411.
[39] J. F. Haw, B. R. Richardson, I. S. Oshiro, N. D. Lazo, J. A. Speed, *Journal of the American Chemical Society* **1989**, *111*, 2052.
[40] M. Aronson, R. Gorte, W. Farneth, D. White, *Journal of the American Chemical Society* **1989**, *111*, 840.
[41] N. Lazo, B. Richardson, P. Schettler, J. White, E. Munson, J. Haw, *The Journal of Physical Chemistry* **1991**, *95*, 9420.
[42] J. F. Haw, J. B. Nicholas, T. Xu, L. W. Beck, D. B. Ferguson, *Accounts of Chemical Research* **1996**, *29*, 259.
[43] H. Ishikawa, E. Yoda, J. N. Kondo, F. Wakabayashi, K. Domen, *The Journal of Physical Chemistry B* **1999**, *103*, 5681.
[44] K. Onda, K. Tanabe, H. Noguchi, K. Domen, A. Wada, *The Journal of Physical Chemistry B* **2003**, *107*, 11391.
[45] T. Bucko, M. Gesvandtnerova, D. Rocca, *Journal of Chemical Theory and Computation* **2020**, *16*, 6049.
[46] J. Rey, C. Chizallet, D. Rocca, T. Bučko, M. Badawi, *Angewandte Chemie International Edition* **2024**, *63*, e202312392.
[47] F. Berger, M. Rybicki, J. Sauer, *ACS Catalysis* **2023**, *13*, 2011.
[48] B. Herzog, M. Chagas da Silva, B. Casier, M. Badawi, F. Pascale, T. Bucko, S. Lebegue, D. Rocca, *Journal of Chemical Theory and Computation* **2022**, *18*, 1382.
[49] R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. Von Lilienfeld, *Journal of Chemical Theory and Computation* **2015**, *11*, 2087.
[50] V. Zaverkin, D. Holzmüller, L. Bonfirraro, J. Kästner, *Physical Chemistry Chemical Physics* **2023**, *25*, 5383.
[51] J. Falk, L. Bonati, P. Novelli, M. Parrinello, M. Pontil, Transfer learning for atomistic simulations using GNNs and kernel mean embeddings, https://doi.org/10.48550/arXiv.2306.01589 **2024**.
[52] A. Erlebach, M. Šípka, I. Saha, P. Nachtigall, C. J. Heard, L. Grajciar, *Nature Communications* **2024**, *15*, 4215.
[53] M. Del Ben, J. Hutter, J. VandeVondele, *Journal of chemical theory and computation* **2013**, *9*, 2654.
[54] M. Del Ben, O. Schütt, T. Wentz, P. Messmer, J. Hutter, J. VandeVondele, *Computer Physics Communications* **2015**, *187*, 120.
[55] M. Bocus, R. Goeminne, A. Lamaire, M. Cools-Ceuppens, T. Verstraelen, V. Van Speybroeck, *Nature Communications* **2023**, *14*, 1008.
[56] S. Vandenhaute, T. Braeckevelt, P. Dobbelaere, M. Bocus, V. V. Speybroeck, Rare Event Sampling using Smooth Basin Classification, https://doi.org/10.48550/arXiv.2306.01589 **2024**.
[57] S. Vandenhaute, Psiflow - interatomic potentials using online learning, https://github.com/molmod/psiflow **2023**.
[58] Y. Babuji, A. Woodard, Z. Li, D. S. Katz, B. Clifford, R. Kumar, L. Lacinski, R. Chard, J. M. Wozniak, I. Foster, M. Wilde, K. Chard, Parsl: Pervasive Parallel Programming in Python, in *Proceedings of the 28th International Symposium on High-Performance Parallel and Distributed Computing*, HPDC '19, Association for Computing Machinery, New York, NY, USA **2019** page 25–36.
[59] T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schütt, F. Schiffmann, et al., *The Journal of Chemical Physics* **2020**, *152*.
[60] J. P. Perdew, K. Burke, M. Ernzerhof, *Physical Review Letters* **1996**, *77*, 3865.
[61] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *The Journal of Chemical Physics* **2010**, *132*.
[62] S. Grimme, S. Ehrlich, L. Goerigk, *Journal of Computational Chemistry* **2011**, *32*, 1456.
[63] L. Vanduyfhuys, ThermoLIB, https://molmod.ugent.be/software/thermolib **2023**.
[64] W. Humphrey, A. Dalke, K. Schulten, *Journal of Molecular Graphics* **1996**, *14*, 33.

Entry for the Table of Contents



A transfer learning approach is proposed to perform enhanced sampling molecular dynamics simulations with arbitrary accuracy. We study the chemisorption of isobutene in the H–SSZ–13 zeolite with RI-RPA and show that carbenium ions are not present in significant concentrations at reaction temperatures.