

Methanol carbonylation over acid mordenite: Insights from *ab initio* molecular dynamics and machine learning thermodynamic perturbation theory

Monika Gešvandtnerová^a, Dario Rocca^c, Tomáš Bučko^{a,b}

^a*Department of Physical and Theoretical Chemistry, Faculty of Natural Sciences, Comenius University, Ilkovičova 6, SK-84215 Bratislava, SLOVAKIA*

^b*Institute of Inorganic Chemistry, Slovak Academy of Sciences, Dúbravská cesta 9, SK-84236 Bratislava, SLOVAKIA*

^c*Université de Lorraine and CNRS, LPCT UMR 7019, F-54000, Nancy, FRANCE*

Abstract

In this work we present a detailed *ab initio* study of the carbonylation reaction of methoxy groups in the zeolite mordenite, as it is the rate determining step in a series of elementary reactions leading to ethanol. For the first time we employ full molecular dynamics simulations to evaluate free energies of activation for the reactions in side pockets and main channels. Results show that the reaction in the side pocket is preferred and, when dispersion interactions are taken into account, this preference becomes even stronger. This conclusion is confirmed using multiple levels of density functional theory approximations with (PBE-D2, PBE-MBD, and vdW-DF2-B86R) or without (PBE, HSE06) dispersion corrections. These calculations, that in principle would require several demanding molecular dynamics simulations, were made possible at a minimal computational cost by using a newly developed approach that combines thermodynamic perturbation theory with machine learning.

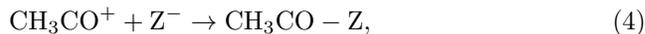
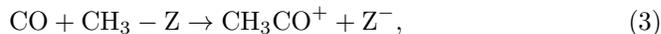
Keywords: acidic zeolites, methanol carbonylation, *ab initio* molecular dynamics, free energy perturbation theory, machine learning

Email addresses: gesvandtnerova2@uniba.sk (Monika Gešvandtnerová), dario.rocca@univ-lorraine.fr (Dario Rocca), tomas.bucko@uniba.sk (Tomáš Bučko)

1. Introduction

The increasing energetic demands of mankind and limited fossil resources along with their rising environmental costs create a significant pressure on the society to seek for renewable sources of energy [1]. Ethanol (EtOH) is an example of an attractive compound used as a fuel or a fuel additive [2, 3, 4] (thereby replacing up to 25 % of the fossil material [1]) that can be produced from renewable resources. Catalytic conversion of syngas, a mixture of CO, CO₂, H₂, and H₂O, that can be prepared by gasification of biomass has recently attracted a significant attention [5, 6, 7, 8] as an alternative to more conventional processes of ethanol production, such as fermentation [1] or hydration of ethylene [9]. Due to a low activity and selectivity of corresponding catalytic reactions, a direct conversion of syngas to EtOH turns out to be unfavorable [4, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20].

An alternative route of EtOH production from syngas proceeds via methanol (MeOH) synthesis catalysed by zeolites. On the basis of experimental evidence, mechanism of alcohol carbonylation was suggested to involve an alkoxy intermediate formed upon interaction with Brønsted acid sites [21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31]. The MeOH based process is catalyzed by acid zeolites (H-Z) and involves the following elementary steps [32]:



As known from previous theoretical reports [33, 32, 34, 35], the reaction of Eq. 3 represents the rate-determining step for the whole series of elementary reactions,

with the computed activation energy being about a factor two higher than the next highest barrier.

Among the zeolite catalysts tested for the use in methanol carbonylation, acid mordenite (H-MOR) has been found to be the most active and selective [36]. As discussed by Boronat et al. [33, 32], the unique feature of H-MOR responsible for its catalytic performance is the presence of Brønsted acid (BA) sites in small channels (SC, see Fig. 1) since the reactions of methoxy groups formed in these voids are faster than those in the main channels (MC) [37, 38].

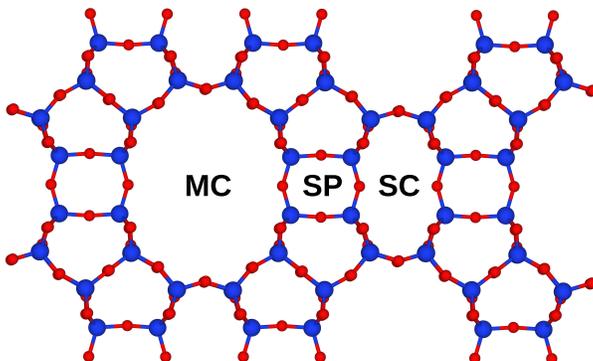


Figure 1: Structure of mordenite (MOR). Positions of important voids are shown: main channel (MC), side pocket (SP), and small channel (SC).

Boronat et al. [33] reported on cluster DFT calculations with semi-local functional B3PW91 [39, 40]. Among all tested active sites, only the one formed in SC was found to exhibit a special orientation of the methoxy group in relation to the side pocket (SP, see Fig. 1) allowing for a perfect fitting between the linear transition state for the rate determining step (Eq. 3) and the stabilizing framework O atoms, while impeding the attack of bulkier nucleofiles. The absence of such sites was offered as an explanation for a lower activity observed for the zeolite Y [36] (which is in some other aspects similar to MOR). The reported potential energy barriers for the reactions involving the sites in MC and SC ranged between 84.0 and 118.0 kJ mol⁻¹. Since the values for the SC sites were not significantly lower than those for the MC sites, it has been

hypothesized that the former sites are kinetically favored as they do not allow (due to spatial constraints) formation of bulkier intermediates of diverse by-reactions.

In the follow up work, Boronat et al. [32] used the B97D functional [41] involving the dispersion correction D2 proposed by Grimme [41]. No relevant mechanistic differences between reactions catalyzed by sites in MC and SC have been found. While the potential energy barriers for the rate determining step (Eq. 3) catalyzed by the site in SC was only 7.6 kJ mol^{-1} lower than that involving the site in MC when dispersion forces were neglected, their inclusion led to a dramatic change in computed values clearly favoring the reaction in SP (with the potential energy barrier being 26.5 kJ mol^{-1} lower compared to MC). As a product of the CO attack, acetyl intermediate was formed in two forms: the covalently bound and ionic. The latter has been proposed to be formed only in SP in reaction with the methyl group in SC.

Rasmussen et al. [34] studied the carbonylation reaction in MOR using the periodic DFT simulations at the dispersion corrected BEEF-vdW level [42]. Potential energy barrier for the rate determining step (Eq. 3) catalyzed by a site in SC was found to be slightly lower than that for MC, but the reported difference (99.4 kJ mol^{-1} vs. $105.2 \text{ kJ mol}^{-1}$) was much smaller than that suggested in dispersion corrected DFT simulations of Boronat et al. [32]. Free energies for $T=438 \text{ K}$ were determined using the static approach [34] and the difference in the reported values remained small ($\sim 10 \text{ kJ mol}^{-1}$), without changing the preference for the reaction in SP involving the methyl group in SC.

In a combined Monte Carlo and molecular dynamics (MD) simulations performed at the force field (COMPASS [43]) level, Liu et al. [44] investigated diffusion and adsorption dynamics reaction species relevant to the carbonylation reaction in MOR. It has been found that the CO diffusion from MC to SP is facile (linked with diffusion barriers of $\sim 4 \text{ kJ mol}^{-1}$ and $\sim 16 \text{ kJ mol}^{-1}$ for the MC \rightarrow SP and SP \rightarrow MC directions, respectively), and the CO molecules tend to aggregate in the SP. The diffusion dynamics of reaction species was suggested as one of the key factors contributing to the activity of the carbonylation reac-

tion in MOR, which was confirmed also by analyzing the diffusion and reaction kinetics of several other zeolites with different pore frameworks involving FER, GON, ATS, IRN, GME and OFF.

Despite the large amount of work that has been done to elucidate the kinetics of the methyl carbonylation reactions in MOR, some questions remain unanswered, while some other have actually arisen. First, the role of thermal and entropy effects is unclear as these have been essentially neglected [33, 32] or treated only approximately [34] in the previous theoretical studies. We note, however, that literature is rich in examples of reactions where these effects have a strong impact on kinetics [45, 46, 47] and/or mechanisms [48, 49]. Second, dispersion forces have been suggested to play an important role in theoretical predictions but since these interactions are treated only approximately within the density functional theory, it is not clear whether different approximations yield similar predictions for this specific reaction. The importance of the careful choice of the electronic structure method in simulations of adsorption and reactions in zeolites has been documented in a number of previous papers [50, 51, 52, 35, 53, 54, 55]. In this work, we address both problems. Specifically, we determine finite temperature free energies of activation for the rate determining step of methyl carbonylation by means of *ab initio* molecular dynamics at multiple levels of theory, involving the semi-local and hybrid density functional approximations, with or without correction for long-range dispersion interactions. Directly performing all the simulations required for this study is numerically cumbersome and hardly feasible (in particular for hybrid functional approximations). To this purpose we use our recently proposed method that couples machine learning and free energy perturbation theory (FEPT) [56, 57]. Starting from a molecular dynamics based on a numerically inexpensive functional, thermodynamic perturbation theory is used to obtain free energy profiles at different “target” levels of theory. The brute force application of perturbation might still require several thousands of single point calculations based on the target approximation. In this work most of these calculations are avoided by applying machine learning (ML): Only a small number of results is computed

explicitly and used to train a ML model, whose inexpensive predictions allow to achieve convergence. For the systems studied in this work, this so-called machine learning thermodynamic perturbation theory (MLPT) [56, 57] approach requires as few as 250 training data to achieve convergence.

This paper is organized as follows: in Section 2, the methodology and the computational details used in this study are described; the results for the conventional approach to reaction free energetics, based on the harmonic approximation to the transition state theory, is presented in Section 3.1; the finite temperature characterization of reactant and product states, based on *ab initio* MD, are discussed in Section 3.2; the MD results for the free energies of activation are analyzed in Section 3.3; the variation of predictions with the density functional approximation is investigated in Section 3.4, and conclusions are provided in Section 4.

2. Simulation details

2.1. Electronic structure calculations

Periodic DFT calculations were performed using the VASP code [58, 59, 60]. The Kohn-Sham equations have been solved variationally in a plane-wave basis set, using the projector-augmented-wave (PAW) method of Blöchl [61], as adapted by Kresse and Joubert [62]. The PBE exchange-correlation functional in the generalized gradient approximation proposed by Perdew et al. [63] was used. The D2 correction of Grimme [41] was applied to account for long-range dispersion interactions, which are particularly important in zeolitic systems [64, 65, 66]. Single point calculations employed in the perturbative free energy calculations (see Section 2.4) were performed using the uncorrected PBE [63], HSE06 [67, 68, 69, 70], PBE-MBD [71, 72, 73], and vdW-DF2-B86R functionals [74, 75, 76, 77, 78]. These functionals were selected so as to represent a reasonably broad sample of popular density functional approximations (DFAs) differing in the treatment of exchange and correlation energies, as well as long-range van der Waals interactions. Namely, the PBE represents an un-

corrected generalized gradient approximation (GGA) functional, HSE06 is an uncorrected hybrid functional, PBE-D2 includes a simple pairwise and atomic environment independent vdW correction, PBE-MBD includes sophisticated many-body and electronic density-dependent correction, and finally, vdW-DF2-B86R represents a non-local dispersion corrected functional. Owing to the large unit cell used in our simulations (*vide infra*), the Brillouin zone sampling was restricted to the Γ -point. Plane wave cutoff energy was set to 400 eV, and the convergence criterion for the SCF cycle was set to 10^{-6} eV/cell.

2.2. Geometric optimizations and molecular dynamics simulations

Geometric optimizations were stopped when all forces acting on atoms were smaller than 0.005 eV/Å. The improved dimer method [79, 80] was used to identify the first order saddle points on the PES representing the transition states (TS). Potential energy minima representing stable reactant and product states linked with TS via intrinsic reaction coordinate (IRC) [81, 82] were subsequently identified using the damped velocity Verlet algorithm [83]. Vibrational analysis was performed for all relaxed structures to test whether these states were correct stationary points of PES and, if needed, additional line-minimization calculations have been performed until the vibrational eigenspectra were correct (i.e., with zero and one imaginary frequency for the stable and transition states, respectively).

Born-Oppenheimer *ab initio* MD simulations were performed in the NVT ensemble, whereby the temperature was maintained by the Andersen thermostat [84] with the frequency of stochastic collisions set to 0.02 fs⁻¹ per atom. The choice of the simulation temperature (440 K) was inspired by experiment [36, 34] and previous theoretical work [34]. A step of 1 fs has been used in integration of classical equations of motion realized via leapfrog Verlet algorithm [85].

2.3. Free energy calculations

The Helmholtz free energy of activation can be expressed using the formula: [86, 87]

$$\Delta A_{R \rightarrow P}^\ddagger = \Delta A_{\xi_{ref,R} \rightarrow \xi^*} - k_B T \ln \left\{ \frac{h}{k_B T} \frac{\langle |\dot{\xi}^*| \rangle}{2} P(\xi_{ref,R}) \right\}, \quad (7)$$

where $\xi = \xi(\mathbf{r})$ is the reaction coordinate or its approximation which, in general, depends on all atomic positions \mathbf{r} , $\xi_{ref,R}$ is the value of ξ for an arbitrary reference point among the reactant configurations, $\dot{\xi}^*$ is the velocity associated with the reaction coordinate for configurations at the transition state (ξ^*), the angular bracket $\langle \dots \rangle$ represents an ensemble average over all reactant configurations, $P(\xi_{ref,R}) = \langle \delta(\xi_{ref,R} - \xi) \rangle$ is the probability density of $\xi_{ref,R}$ within the ensemble of reactant configurations, and $\Delta A_{\xi_{ref,R} \rightarrow \xi^*} = -k_B T \ln \left\{ \frac{P(\xi^*)}{P(\xi_{ref,R})} \right\}$ is the reversible work needed to shift the reaction coordinate from the value $\xi_{ref,R}$ to ξ^* . The individual terms of Eq. 7 are computed by means of molecular dynamics (MD) as follows. The probability density $P(\xi_{ref,R})$ is approximated by a histogram obtained using the data from a straightforward MD of the reactant state. The term $\Delta A_{\xi_{ref,R} \rightarrow \xi^*}$ is computed using the blue moon ensemble method. [86, 88] Finally, the generalized velocity term $\langle |\dot{\xi}^*| \rangle$ can be determined using the following equation [86]:

$$\langle |\dot{\xi}^*| \rangle = \sqrt{\frac{2k_B T}{\pi}} \frac{1}{\langle Z^{-1/2} \rangle_{\xi^*}}, \quad (8)$$

where Z is defined in Eq. B.3 and $\langle \dots \rangle_{\xi^*}$ indicates the constrained ensemble average with $\xi(\mathbf{r}) = \xi^*$. The term $\langle Z^{-1/2} \rangle_{\xi^*}$ is readily available from the constrained MD performed for the state ξ^* within the $\Delta A_{\xi_{ref,R} \rightarrow \xi^*}$ calculation.

2.4. The MLPT approach

Even the most sophisticated *ab initio* MD simulation protocol will still yield poor predictions if the electronic structure method used in the sampling does not accurately describe interactions within each configuration of the given ensemble. It is therefore desirable to have a tool at our disposal allowing us to explore

free energies at several different levels of theory without the need to repeat all the lengthy *ab initio* MD or Monte Carlo simulations multiple times. In our recent work [57], we proposed to use the free energy perturbation theory [89, 90] (FEPT) in combination with machine learning (ML) for this purpose. In this section, the method is briefly reviewed.

Let us assume that the free energy of activation ΔA^\ddagger has already been determined as described in Section 2.3 using the method (hereafter dubbed as the production method) with Hamiltonian $H(\mathbf{r}, \mathbf{p})$ defined by the potential energy $V(\mathbf{r})$ and the kinetic energy $T(\mathbf{p})$. The same quantity determined at a different (presumably higher) level of theory (the target method) with Hamiltonian $\tilde{H}(\mathbf{r}, \mathbf{p})$ can be formally expressed as:

$$\Delta \tilde{A}^\ddagger = \Delta \tilde{A}_{\xi_{Ref,R} \rightarrow \xi^*} - k_B T \ln \left(\frac{\hbar}{k_B T} \frac{\langle |\dot{\xi}^*(0)| \rangle}{2} \tilde{P}(\xi_{Ref,R}) \right), \quad (9)$$

where the tilde is used to distinguish the quantity computed using the target method from that obtained using the production method. In the following, we will suppose that production and target Hamiltonians differ only in position dependent potential energy term, namely $\tilde{H}(\mathbf{r}, \mathbf{p}) - H(\mathbf{r}, \mathbf{p}) = \tilde{V}(\mathbf{r}) - V(\mathbf{r}) = \Delta V(\mathbf{r})$. As shown in our previous work [57], the target free energy of activation $\Delta \tilde{A}^\ddagger$ can be obtained from the production free energy of activation ΔA^\ddagger in Eq. 7 using FEPT via the following formula:

$$\Delta \tilde{A}^\ddagger = \Delta A^\ddagger - k_B T \ln \left\{ \frac{\langle \exp[-\Delta V(\mathbf{r})/k_B T] \rangle_{\xi^*}}{\langle \exp[-\Delta V(\mathbf{r})/k_B T] \rangle} \right\}, \quad (10)$$

while for the free energy differences between two stable states ($\Delta \tilde{A}_{R \rightarrow P}$) is found [57]:

$$\Delta \tilde{A}_{R \rightarrow P} = \Delta A_{R \rightarrow P} - k_B T \ln \left\{ \frac{\langle \exp[-\Delta V(\mathbf{r})/k_B T] \rangle_P}{\langle \exp[-\Delta V(\mathbf{r})/k_B T] \rangle_R} \right\}, \quad (11)$$

where $\langle \dots \rangle_R$ and $\langle \dots \rangle_P$ indicate ensemble averages over initial and final stable states, respectively.

In practice, it is often useful to replace the straightforward FEPT expressions such as Eq. 10 by an approximation allowing to compute the free energy

differences from an appropriate analytical model for the $\Delta V(\mathbf{r})$ distribution [90]. If the latter is approximately a Gaussian-shaped function, the second order cumulant expansion [90] (CE):

$$\langle \exp \left[-\frac{\Delta V(\mathbf{r})}{k_B T} \right] \rangle \approx \exp \left[-\frac{\langle \Delta V(\mathbf{r}) \rangle}{k_B T} + \frac{\langle \Delta V^2(\mathbf{r}) \rangle - \langle \Delta V(\mathbf{r}) \rangle^2}{2(k_B T)^2} \right] \quad (12)$$

represents a good approximation whose advantage compared to the straightforward FEPT is that it allows to reduce the numerical problems occurring in simulations of large systems where naturally occurring large ΔV fluctuations might lead to huge statistical weights of few configurations which would subsequently dominate the statistical ensembles. We employ the CE approach throughout this work.

As shown in our previous work [56, 57], a machine learning regression model can be used to significantly decrease the number of calculations of ΔV needed for applying FEPT approach. Based on the well established Δ -ML method [91], which assumes a reasonable statistical correlation between production and target method energies, we exploit the rather smooth behavior of $\Delta V(\mathbf{r})$ to make the learning process significantly more efficient with respect to the direct learning of \tilde{V} . Following Ref. [56], the kernel ridge regression (KRR) [92] machine learning algorithm is used in this work along with the global REMatch kernel [93] constructed from smooth overlap of atomic positions (SOAP) kernels [94, 93] of local atomic environments. The numerical implementation of the ML models considered in this work is based on the Dscribe and scikit-learn libraries [95, 96, 97]; the in-house programs implementing the MLPT method are available from the authors upon request.

2.5. Structural model

A monoclinic supercell of mordenite doubled in the c -direction (parallel with the main channel) was constructed on the basis of experimental data from the database of zeolite structures of the International Zeolite Association (IZA) [98]. This supercell, shown in Fig. 2, contained 48 tetrahedral (SiO_2 or AlO_2) units (altogether 154 atoms) and was defined by the following lattice parameters:

$a = b = 13.738 \text{ \AA}$, $c = 15.084 \text{ \AA}$, and $\gamma = 83.3^\circ$. In our model, the Si/Al ratio was set to 23, whereby the two framework Al atoms occupied the tetrahedral positions T3 and T4 (see Fig. 2). As shown in Fig. 2, the reactant structures contained two CH_3 groups bound to O atoms O10 and O9. This structural model was chosen so as to allow for a direct and consistent comparison of energetics of the carbonylation reaction (see Eq. 3) in the main channel (MC) with that in the side pocket (SP). In particular, the methyl group initially sitting next to the T3 site in SC pointing towards SP and oriented perpendicular to it has previously been identified as the one allowing for a perfect fitting between the linear TS for the CO attack of methyl group and the stabilizing framework O atoms [33]. In contrast, the methyl group next to the T4 site is inside MC and, again, it is oriented perpendicular to the channel. Furthermore, both Al sites are located on an eight membered ring (8MR) of T sites with a very similar local environment. Hence the location of reacting molecule will remain the only important factor making difference in the catalytic performance of the sites in MC and SC. Note that the reaction involving the methyl group located in SC does not occur strictly in SC or SP but, instead, all the relevant configurations are included in both voids simultaneously (*vide infra*). For an easier reference, we shall refer to this reaction simply as the "reaction in the side pocket" and label the corresponding reaction states with "SP". Similarly, the reaction involving the methyl group initially located in MC will be referred to, regardless of any further details, as the "reaction in the main channel" and the corresponding reaction states will be labeled with "MC". We remark that the same locations of the Al-O- CH_3 species were considered also in the previous theoretical work [33, 32, 44]. Furthermore, the location of Al and acid protons in MOR has been very recently investigated in a combined experimental ^1H DQ NMR and theoretical DFT study [99] and the Brønsted sites potentially giving rise to the methoxy sites considered in our work were found to belong to the most likely ones. Drawings of structures presented in this work have been created using the program VESTA [100].

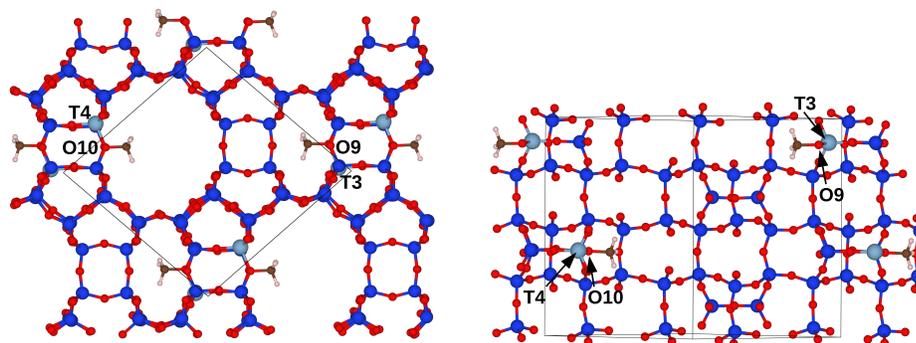


Figure 2: Top (left) and side (right) views on the structural model of MOR employed in this work. The Al atoms occupy the tetrahedral sites T3 and T4, and the methyl groups present in the reactant state are attached to the oxygen atoms O10 and O9 next to Al. The solid lines indicate the supercell used in the calculations. The labeling of the framework tetrahedral (T) and oxygen (O) sites follows Ref. [101]. (Light blue: Al, dark blue: Si, red: O, white: H, brown: C.)

3. Results

3.1. Static calculations

In the previous theoretical work [33, 32, 34], the methanol carbonylation has been studied by means of the static approach based on atomic relaxations, in Ref. [34] also complemented by the harmonic approximation to transition state theory. Since our computational setting differs from that used in the previous studies in several important aspects (e.g., the structural model or the treatment of long-range dispersion interactions), we find useful to start our analysis by calculations involving this simple model.

The relaxed reactant structures are shown in Fig. 3. In these configurations, adsorption complexes are formed in which CO interacts weakly with the methyl group. The two complexes (hereafter labeled as R(MC) and R(SP)) differ significantly in the C...C distances, which are 3.19 Å and 4.46 Å for MC and SP, respectively. This difference can be understood: in SP, CO is pulled deeper into the small cavity, where it is stabilized by stronger dispersion interactions due to a higher local density of polarizable framework atoms, causing thus the increase

in the C...C separation. In MC, on the other hand, CO is located in a much larger empty space and hence the C...C distance is given only by the specific CO...CH₃ interaction.

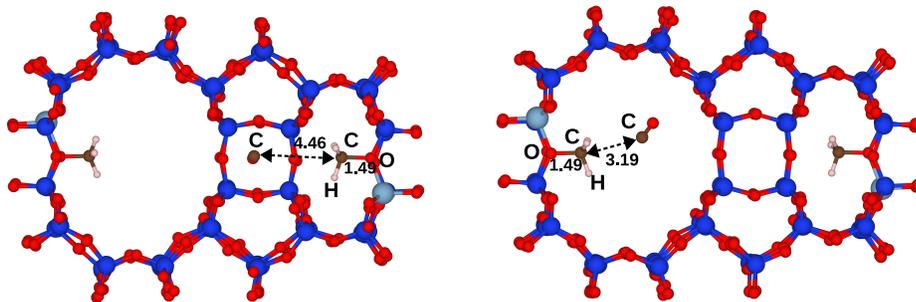


Figure 3: Relaxed reactant for the carbonylation reaction (Eq. 3) in SP (left) and MC (right). Selected interatomic distances (in Å) are indicated. (Light blue: Al, dark blue: Si, red: O, white: H, brown: C.)

We shall show in Section 3.2 that the relaxed configuration R(MC) is not stable at our target temperature (440 K) as the CO molecule tends to diffuse spontaneously from MC into SP. For the sake of completeness, we therefore consider here also the reactant state (hereafter labeled as R') created upon relaxation of the structure obtained from molecular dynamics simulation of R(MC). As shown in Fig. 4, the configuration R' is similar to R(SP) as CO is located in SP in both cases but the two configurations differ in their position with respect to the nearest CH₃ group. Further on, we shall use the symbol MC' to refer to the reaction occurring in MC but starting from R'.

The relaxed transition states are shown in Fig. 5. In these structures, the C-O bonds are just being broken and C-C bonds are being formed. Unlike in the case of the reactant structures, the interatomic distances in TS are hardly affected by the interactions with the zeolitic environment. Nevertheless, the environment affects, to some extent, the orientation of the CO fragment with respect to the CH₃ group: the corresponding O-C-C angles are 175° and 157° for the TS structures formed in SP and MC, respectively. As pointed out in previous

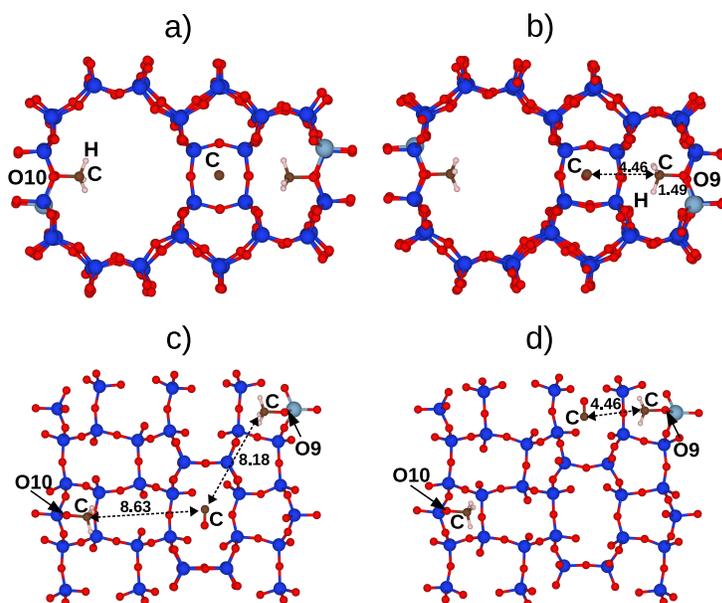


Figure 4: Comparison of the reactant states R' (a,c) and R(SP) (b,d) for the carbonylation of methoxy group in MOR (Eq. 3). Top (a,b) and side (c,d) views are shown for both structures. Selected interatomic distances (in Å) are indicated. (Light blue: Al, dark blue: Si, red: O, white: H, brown: C.)

theoretical work [34, 32], the nearly linear structure of TS fits perfectly in the void of SP, while the large open space of MC allows for a more significant deviation from linearity. The values of the interatomic distances determined for the TS are in a good agreement with those reported in previous studies [32, 34], with the maximum difference being ~ 0.05 Å.

As reaction products, acetyl cations have been identified by the IRC analysis (see Section 2), the corresponding optimized structures are shown in Fig. 6. These species have been shown in previous theoretical work [32, 34] to swiftly undergo a further transformation (Eq. 4), whereby the corresponding free energy barrier of activation was estimated to be only ~ 2 kJ mol $^{-1}$. This result indicates that the stability of the free cationic species is extremely low and our MD results presented in Section 3.2 are consistent with this expectation.

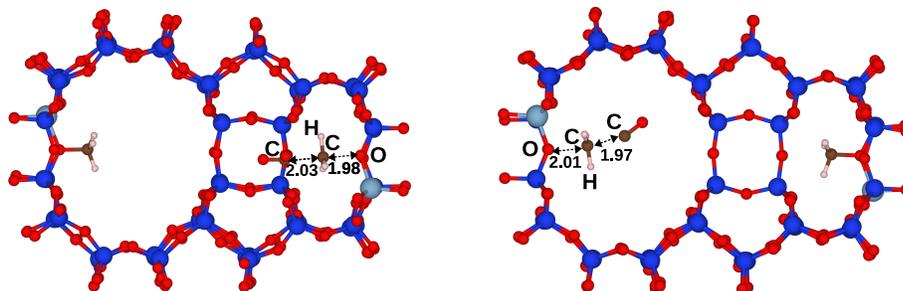


Figure 5: Relaxed transition states for the carbonylation reaction (Eq. 3) formed in SP (left) and MC (right). Selected interatomic distances (in Å) are indicated. (Light blue: Al, dark blue: Si, red: O, white: H, brown: C.)

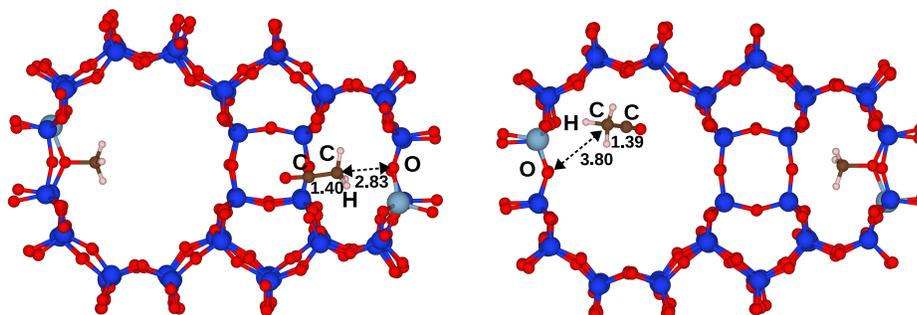


Figure 6: Relaxed product states for the carbonylation reaction (Eq. 3) formed in SP (left) and MC (right). Selected interatomic distances (in Å) are indicated. (Light blue: Al, dark blue: Si, red: O, white: H, brown: C.)

The free energy profiles for the reaction occurring in SP and MC are shown in Fig. 7 and the relevant numerical data are presented in Tab. 1. As expected, the free energy difference between the states R(SP) and R' is only 0.5 kJ mol^{-1} and these two states are very similar also in terms of internal energy and entropy.

The slightly lower free energy of R(SP) can be explained by the presence of a weak interaction of CO with CH₃ group, which is absent in the configuration R' (see Fig. 4). The free energy barrier for the reaction involving R(SP) and

Table 1: Relative free energies (ΔA), internal energies (ΔU), and entropies (ΔS) for the states R', R(MC), TS(SP), TS(MC), P(SP), and P(MC) computed using the static approach at the PBE-D2 level. Contribution of dispersion interactions to internal energies is shown in parentheses. All values are referenced to the lowest free energy reactant state R(SP) (cf. Fig. 3).

Configuration	ΔA (kJ mol ⁻¹)	ΔU (kJ mol ⁻¹)	ΔS (J mol ⁻¹ K ⁻¹)
R'	0.5	-2.3 (-0.6)	-6
R(MC)	21.5	9.5 (6.2)	-27
TS(SP)	99.8	80.2 (-8.6)	-44
TS(MC)	115.5	103.0 (7.1)	-28
P(SP)	-39.3	-52.1 (-11.6)	-29
P(MC)	-16.9	-46.1 (-4.0)	-66

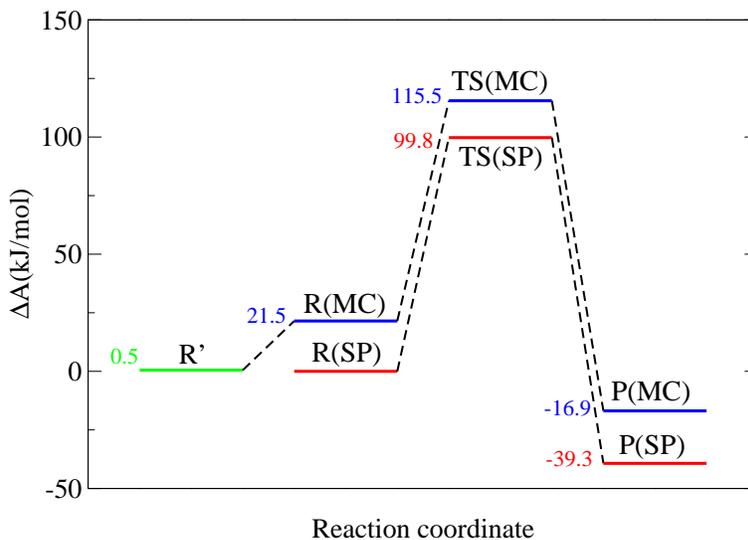


Figure 7: Free energy profile for the for the methanol carbonylation reaction in SP and MC computed using the static approach at the PBE-D2 level. Zero on the energy axis corresponds to the free energy of the state R(SP).

TS(SP) is 99.8 kJ mol⁻¹ and the corresponding internal energy contribution is 80.2 kJ mol⁻¹. The negative sign of the predicted entropy of activation

($-44 \text{ J mol}^{-1}\text{K}^{-1}$) can be explained by the loss of a part of soft degrees of freedom (such as hindered rotations and translations) upon the TS formation. This reaction is both exergonic and exothermic, as evident from the values of free and internal energies of reaction ($-39.3 \text{ kJ mol}^{-1}$ and $-52.1 \text{ kJ mol}^{-1}$, respectively). Compared with TS, the entropy of product is $15 \text{ J mol}^{-1} \text{ K}^{-1}$ higher.

The state R(MC) is 21.5 kJ mol^{-1} higher in free energy than R(SP), which is, at least in part, a consequence of a weaker stabilization by dispersion interactions in the large MC (see Tab. 1). A similar argument can be used to explain the higher free energy of TS(MC) as compared to TS(SP), with the difference being as large as 15.7 kJ mol^{-1} . Furthermore, one can recognize from the results presented in Tab. 1 that the state TS(MC) is less constrained than TS(SP) (as evident from higher entropy), which can be explained by larger space available in MC for the motion of the O-C-C-O fragment. The free energy barrier involving the states R(MC) and TS(MC) is 94.0 kJ mol^{-1} , whereby the contribution of internal energy to this value is 93.5 kJ mol^{-1} . The corresponding entropy of activation is therefore only $-1 \text{ J mol}^{-1} \text{ K}^{-1}$. However, the state R(MC) is not the adsorption ground state. Hence the free energy difference between the states TS(MC) and R', which is $115.0 \text{ kJ mol}^{-1}$ (with the internal energy contribution of $105.3 \text{ kJ mol}^{-1}$), is a more relevant estimate of effective free energy of activation for the reaction in MC. Since the free energy barrier for the reaction in SP is 15.2 kJ mol^{-1} lower, this result supports the previous experimental reports [37, 38], claiming that the carbonylation reaction is catalyzed preferentially by the sites in SP.

The product formed in MC is 22.4 kJ mol^{-1} higher in free energy compared to that formed in SP and, once again, the dispersion interactions contribute significantly (7.5 kJ mol^{-1}) to this difference. As in the case of SP, the reaction in MC is both exergonic and exothermic, with the free and internal energies of reaction being $-38.4 \text{ kJ mol}^{-1}$ and $-55.6 \text{ kJ mol}^{-1}$, respectively.

Our results can be compared with previous theoretical studies [32, 34] in which a similar computational setup as discussed here was used. Boronat et

al. [32] obtained the zero temperature apparent reaction barriers (defined with respect to non-interacting reference systems) of ~ 54 kJ mol $^{-1}$ for the reaction in SP and ~ 80 kJ mol $^{-1}$ for MC, when the D2 correction [41] (within the B97D functional [41]) was used. Hence the difference in internal energies of the two transition states was predicted to be ~ 26 kJ mol $^{-1}$ in favor of TS(SP), which is to be compared with our zero temperature result of 24.8 kJ mol $^{-1}$ (see Tab. 2). Furthermore, it was reported in Ref. [32] that when the dispersion interactions were not included, the reaction barriers for SP and MC were rather similar, with the difference being only 8 kJ mol $^{-1}$. Once again, this result is in a good agreement with our difference in internal energies of activation when the dispersion corrections are removed, which is 4 kJ mol $^{-1}$ (see Tab. 1). Rasmussen et al. [34] reported the apparent finite temperature (438 K) free energy barriers obtained using the BEEF-vdW functional [42] (whose performance in the zeolite chemistry has been, however, criticized [102]). According to these results, TS(SP) is 10 kJ mol $^{-1}$ lower in free energy compared to TS(MC), which is to be compared with our result of 16 kJ mol $^{-1}$ (a part of this difference is probably due to a different location of the methyl group in MC considered in Ref. [34]). Importantly, the difference in finite temperature barriers reported by Rasmussen et al. [34] is much smaller than the zero temperature difference predicted by Boronat et al. [32]. This result indicates that entropy and thermal effects play important role in kinetics of the methanol carbonylation reaction and their correct treatment will be essential for reliable predictions. We address this problem in Section 3.2. It is also evident from our results and from the published data, that dispersion interactions are essential for the stabilization of species formed in SP. Given the approximate treatment of these interactions within DFT [103, 104], it is of interest to examine the variation of predictions on energetics with the choice of method. In this work, we tested the following popular DFAs: uncorrected PBE[63], PBE-D2[63, 41] with a simple pairwise and electronic structure independent correction, PBE-MBD[71, 72, 73] with many-body electronic density-dependent correction, vdW-DF2-B86R[74, 75, 76, 77, 78] representing a non-local dispersion

corrected functional, and hybrid functional HSE06[67, 68, 69, 70], where the quality of description of bonding is presumably improved by adding exact exchange energy. As shown in Tab. 2, these approximations yield electronic energy contributions to free energy barriers, determined using geometries optimized at the PBE-D2 level, that differ by as much as 30 kJ mol⁻¹. A detailed analysis of impact of the DFA choice (involving accurate treatment of thermal effects) is presented in Section 3.4. Additional static calculations results are presented in Section SI of the Supporting Information.

Table 2: Electronic contributions to free energies of activation (in kJ mol⁻¹) of the carbonylation of methoxy groups in SP and MC of MOR determined using different density functional approximations. Reactant and transition states were fixed at the geometries determined at the PBE-D2 level. MC' indicates reaction in MC starting from the state R' (cf. Fig. 4).

Method	SP	MC'	MC
PBE-D2	86.6	111.4	99.4
PBE	95.2	103.7	98.5
HSE06	115.9	125.9	118.3
PBE-MBD	89.6	112.1	98.1
vdW-DF2-B86R	92.6	123.4	103.9

3.2. Finite Temperature Simulations of Reactant and Product States

In this section, we report on the finite temperature (T=440 K) behavior of reactant and product states in SP and MC. To this end, an independent MD simulation of a length of 200 ps has been performed for each state discussed here. Unless stated otherwise, the calculations have been initialized from the relaxed structures and the equilibration period of 5 ps has been considered. For the reactant states, we monitored the parameter ξ :

$$\xi = r_{Z-C} - r_{C-C}, \quad (13)$$

where r_{Z-C} is the interatomic separation between carbon in the methyl group and the framework O attached to the methyl group in the reactant configuration,

and r_{C-C} is the distance between the carbon atoms in CO and the CH₃ group participating in reaction. This geometric parameter is used as an approximation to the reaction coordinate in our calculations of the free energy of activation presented in Section 2.3, and it takes the values $\xi \approx 0$ for TS, $\xi \lesssim 0$ for reactant, and $\xi \gtrsim 0$ for product.

The CO molecule adsorbed in SP (see Fig. 3) rotated and translated within the small cavity in which it was initially located. Nevertheless, the molecule did not leave SP during the whole simulation time considered here. A direct consequence of the small variation in the position of CO is the shape of the probability distribution function $P(\xi)$, which is approximately Gaussian with the center at $\xi \approx -3 \text{ \AA}$ (see Fig. 8). As the reference reactant state for the free energy calculations (see Eq. 7), the point $\xi_{Ref,R(SP)} = -1.89 \text{ \AA}$ was chosen for which the probability distribution function takes the value $P(\xi_{Ref,R(SP)}) = 0.14 \text{ \AA}^{-1}$.

As discussed in Section 3.1, dispersion interactions are an important driving force causing diffusion of CO into SP. Due to very weak specific interactions with the methoxy group, the CO molecule adsorbed in MC (see Fig 3(b)) was free to move within the whole volume of the main channel. After 29 ps, however, the molecule spontaneously diffused into the side pocket where it stayed for the rest of the simulation (note that this initial period was considered as equilibration and the corresponding data were not used in ensemble averages discussed here). Indeed, the internal energy of CO in MC computed as an average of the total energy is $\sim 9 \text{ kJ mol}^{-1}$ higher than that for R(SP) which is almost identical to the difference between these two states in average dispersion energy ($\sim 10 \text{ kJ mol}^{-1}$). Note that these MD results are similar to our estimates obtained within the static approach (internal energy difference of $\sim 10 \text{ kJ mol}^{-1}$, dispersion energy difference of $\sim 6 \text{ kJ mol}^{-1}$, see Tab. 1). Upon diffusion of CO from MC to SP, the adsorption state R' was formed. This state is very similar to R(SP), with the only difference being the absence of Al and of the weakly interacting CH₃ group (see Fig. 4) in the side pocket containing the reactant molecule. Consequently, both the internal energies and the average dispersion

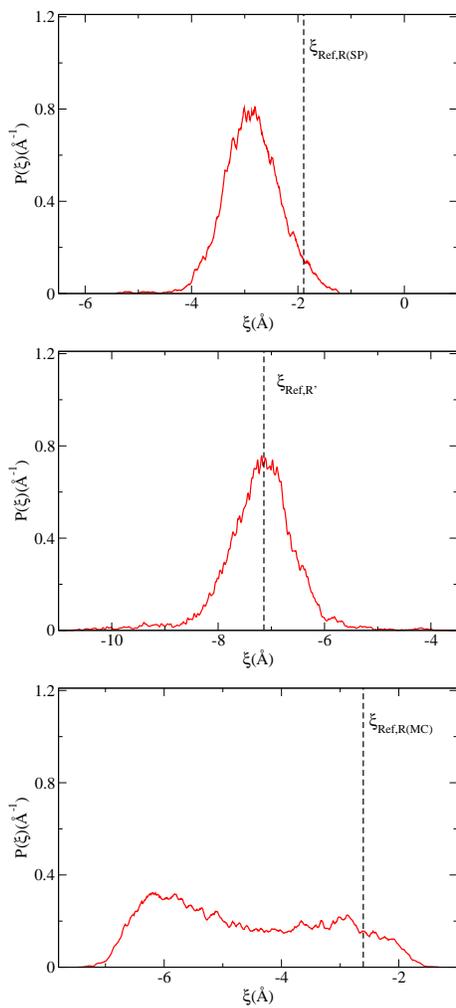


Figure 8: Probability distribution functions of approximation to the reaction coordinate ξ determined using the MD simulations with $T=440$ K performed at the PBE-D2 level for the reactant configurations R(SP) (above) and R' (below), and at the PBE level for the reactant configurations R(MC) (below). Dashed lines indicate the reference reactant states used in the free energy calculations.

interactions determined for R(SP) and R' differ only negligibly (by 2 kJ mol^{-1} and 0.4 kJ mol^{-1} , respectively). From the analysis of the radial distribution function (RDF) computed for the X-Y pairs of atoms, where X represents an

atom (carbon or oxide) from the CO molecule, and Y is a tetrahedral center (Al or Si atom) of the zeolite framework, one can see that the motions of CO in R(SP) and R' are very similar too (see Fig. 9) and hence both states are similar also in terms of entropy.

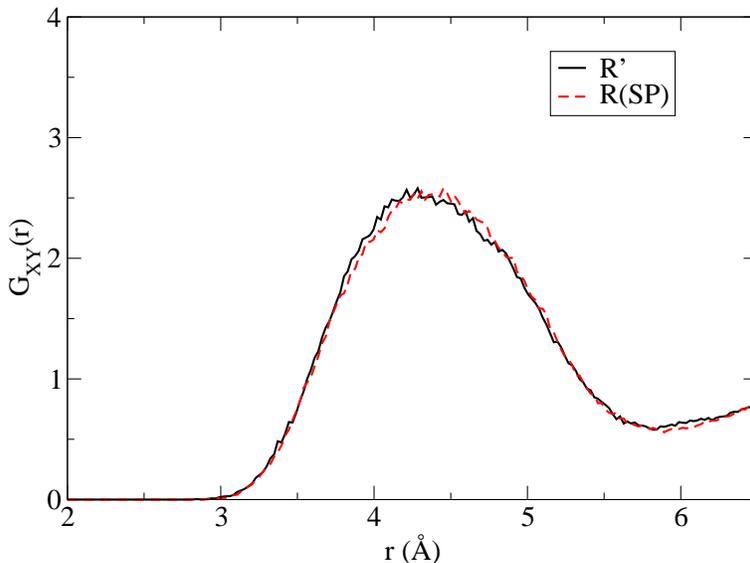


Figure 9: Radial distribution functions (G_{XY}) for the pairs of atoms X (C and O atoms from the CO molecule) and Y (Al and Si atoms from the zeolite framework) in the reactant configurations R' and R(SP) at T=440 K determined using MD performed at the PBE-D2 level.

Given the small differences in internal energies and entropies (observed also in our static approach calculations, see Tab. 1), as well as the similarity in the structure, it is reasonable to consider the free energies of R' and R(SP) as being approximately equal. We shall make use of this assumption in Sections 3.3 and 3.4, where free energies of activation are discussed. The probability distribution function $P(\xi)$ determined for the state R' (where ξ was defined with respect to the methyl group in MC) is shown in Fig. 8. As the reference reactant state for our free energy calculations, the point $\xi_{Ref,R} = -7.14 \text{ \AA}$ was chosen for which the probability distribution function takes the value $P(\xi_{Ref,R'}) = 0.72 \text{ \AA}^{-1}$.

Next, the MD simulations of the product formed in SP and MC are considered. In both cases, the acetyl cation CH_3CO^+ eventually transformed, after short episodes involving temporary formation of neutral ketene CH_2CO (see Section SII in the Supplemental Information), into the surface acetate species (see Eq. 4), whose existence within the MTO process has recently been proven experimentally by Chowdhury et al. [29]. The formation of the surface acetate occurred almost immediately (~ 3 ps after the equilibration) in the case of MC. The more confining environment in SP seemed to have partially stabilized the free cationic form, which transformed into another form only after >90 ps after equilibration. A stabilizing effect of SP on the free cationic form of product has been suggested also in the previous theoretical study [32]. The low stability of the free acetyl cation is not surprising, as the potential energy barrier for its transformation into the acetyloxy species is as low as 2 kJ mol^{-1} [32, 34].

Finally, we performed MD simulations of reactants using the PBE functional without the dispersion correction. The observed behavior was exactly opposite to that found in MD at the PBE-D2 level: the CO molecule initially located in SP diffused to MC (after ~ 160 ps), while the molecule initially located in MC did not leave the main channel. This observation is in line with our conclusion on decisive role of dispersion interactions in stabilization of the molecule in SP. Indeed, our results suggest that the presence or absence of dispersion interactions leads to the change of the adsorption ground state. For the purpose of analysis presented in Section 3.4, the reference reactant state for CO in MC was chosen as $\xi_{Ref,R(MC)} = -2.61 \text{ \AA}$ and the corresponding value of $P(\xi_{Ref,R(MC)}) = 0.16 \text{ \AA}^{-1}$ was determined at the PBE level (see Fig.8).

3.3. Free Energy of Activation

In this section we discuss the free energy barriers (ΔA^\ddagger) for the methyl group carbonylation occurring in the side pocket and in the main channel of mordenite at $T=440 \text{ K}$ that we computed using the approach based on the blue moon (BM) ensemble technique [86, 88] described in Section 2.3.

For the reaction in SP, 15 evenly spaced integration points ranging between

the values $-1.89 \text{ \AA} \leq \xi \leq 1.33 \text{ \AA}$ were used for the thermodynamic integration of free energy gradients (see Eq. B.1), whereby the initial integration point was chosen to correspond to the reference reactant configuration defined in Section 3.2. The free energy transition state $\xi^* = -0.02 \text{ \AA}$ was determined as the point with vanishing free energy gradient on the top of the free energy profile $A(\xi)$ shown in Fig. 10. The computed value of $\Delta A_{\xi_{Ref,R} \rightarrow \xi^*}$ is 93.3 kJ mol^{-1} which, combined with the values of $P(\xi_{Ref,R})$ and $\langle |\dot{\xi}^*| \rangle$ summarized in Tab. 3, yield the result $\Delta A^\ddagger = 102.4 \text{ kJ mol}^{-1}$. This value is very close to our result obtained in the static approach, which is 99.8 kJ mol^{-1} (see Section 3.1). Furthermore, we determined the internal energy of activation via the formula:

$$\Delta U^\ddagger = \frac{\langle Z^{-1/2} V \rangle_{\xi^*}}{\langle Z^{-1/2} \rangle_{\xi^*}} - \langle V \rangle_R \quad (14)$$

where the term $\langle V \rangle_R$ stands for the average potential energy of reactant computed in unconstrained MD simulation, and the term $\frac{\langle Z^{-1/2} V \rangle_{\xi^*}}{\langle Z^{-1/2} \rangle_{\xi^*}}$ is the potential energy of transition state from the constrained MD simulation. The resulting value is $\Delta U^\ddagger = 90.2 \text{ kJ mol}^{-1}$, which is somewhat higher than our static approach estimate of 80.2 kJ mol^{-1} .

Table 3: Values of individual terms used in the calculation of free energy of activation ΔA^\ddagger obtained via Eq. 7 for the methyl group cabonylation in SP and MC of MOR at $T=440 \text{ K}$. Note that MC' indicates the reaction in MC starting from the state R' (cf. Fig. 4).

	SP	MC'
$\Delta A_{\xi_{Ref,R} \rightarrow \xi^*} \text{ (kJ mol}^{-1}\text{)}$	93.3	126.3
$\langle \dot{\xi}^* \rangle \text{ (\AA s}^{-1}\text{)}$	$1.05 \cdot 10^{13}$	$1.05 \cdot 10^{13}$
$P(\xi_{Ref,R}) \text{ (\AA}^{-1}\text{)}$	0.14	0.72
$\Delta A^\ddagger \text{ (kJ mol}^{-1}\text{)}$	102.4	129.6

In the calculations involving reaction in MC, 22 evenly spaced integration points defined on the interval $-7.14 \text{ \AA} \leq \xi \leq 0.52 \text{ \AA}$ were used for the integration of free energy gradients (see Eq. B.1), whereby the initial integration point corresponded to the reference reactant configuration defined in Section 3.2. The

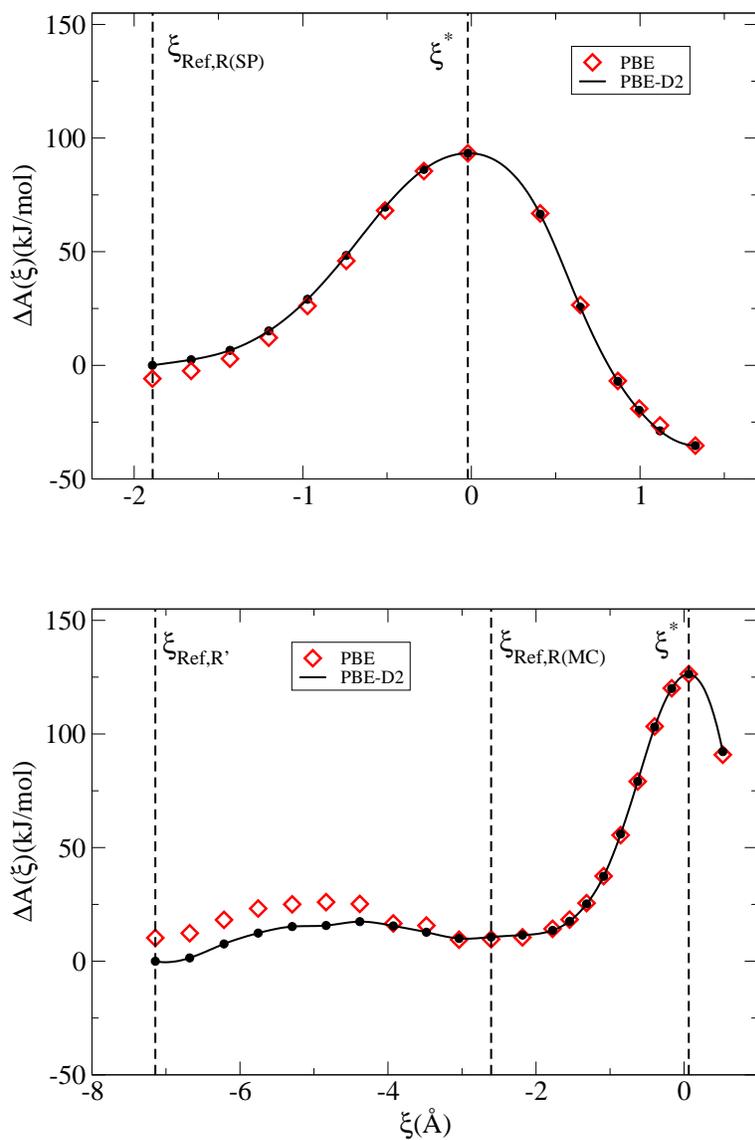


Figure 10: Free energy profiles $\Delta A(\xi)$ for the methyl group cabonylation occurring in SP (above) and MC (below) computed using MD at the PBE-D2 level and perturbatively at the PBE level. Dashed lines indicate the reference reactant states and the free energy transition states used in free energy calculations via Eq. 7.

free energy transition state $\xi^*=0.06 \text{ \AA}$ was determined as the point with vanishing free energy gradient on the top of the free energy profile $A(\xi)$ shown in Fig. 10. The values of $\Delta A_{\xi_{Ref,R} \rightarrow \xi^*}$, $P(\xi_{Ref,R})$, and $\langle |\dot{\xi}^*| \rangle$, used in the calculation of the free energy of activation are summarized in Tab. 3. The resulting value of ΔA^\ddagger is $129.6 \text{ kJ mol}^{-1}$, which is $\sim 15 \text{ kJ mol}^{-1}$ higher than the result obtained using the static approach. The larger disagreement compared to the reaction in SP is not surprising, as the static approach is known to fail in cases where soft degrees of freedom such as the hindered rotations and translations contribute significantly to free energetics [105, 106, 107, 46]. The internal energy of activation computed using Eq. 7 is $113.5 \text{ kJ mol}^{-1}$, which is to be compared with the static approach result of $105.3 \text{ kJ mol}^{-1}$. In principle, one can use the thermodynamic relation $A = U - TS$ to determine also the entropy of activation (ΔS^\ddagger). Nevertheless, the quality of such a result is poor as it is obtained by subtracting two large numbers. Hence the statistical uncertainty in ΔS^\ddagger is comparable to its actual value (see Tab. 4). Since the difference in ΔS^\ddagger computed for the reactions in SP and MC is not statistically significant (because the difference in computed values is smaller than the total statistical uncertainty), we can only draw a qualitative conclusion that the entropy of activation is negative in both cases, corresponding to a loose reactant and tight transition state picture. We note that a qualitatively same picture has been obtained also using the static approach, see Tab. 1.

Table 4: Values of the free energy ΔA^\ddagger , internal energy ΔU^\ddagger and entropy ΔS^\ddagger of activation computed using MD for the methyl group carbonylation in SP and in MC of MOR at $T=440 \text{ K}$. The numbers after the \pm symbol represent the standard error multiplied by the factor 1.96, corresponding to the confidence interval of 95%. Note that MC' indicates reaction in MC starting from the reactant state R' (cf. Fig. 4).

	SP	MC'
$\Delta A^\ddagger \text{ (kJ mol}^{-1}\text{)}$	102.4 ± 0.8	129.6 ± 2.9
$\Delta U^\ddagger \text{ (kJ mol}^{-1}\text{)}$	90.2 ± 2.1	113.5 ± 2.7
$\Delta S^\ddagger \text{ (J mol}^{-1} \text{ K}^{-1}\text{)}$	-28 ± 5	-38 ± 9

Free energy profile $A(\xi)$ determined for the reaction in MC exhibits two minima (Fig. 10) corresponding to the states R' ($\xi \approx -7.14 \text{ \AA}$) and R(MC) ($\xi \approx -2.6 \text{ \AA}$). On the basis of this profile, we estimate that the state R' is $\sim 11 \text{ kJ mol}^{-1}$ lower in free energy than R(MC) and the free energy barrier separating the latter from the former is only $\sim 7 \text{ kJ mol}^{-1}$. This result explains the behavior observed in our straightforward MD simulations presented in Section 3.2, where the CO molecule initially located in the main channel spontaneously diffused into the neighboring side pocket. The CO diffusion from MC to SP has recently been studied in a recent theoretical work of Liu et al. [44]. Despite using a very different computational approach (force field based Monte Carlo), the reported diffusion barriers ($\sim 4 \text{ kJ mol}^{-1}$ and $\sim 16 \text{ kJ mol}^{-1}$ for the MC \rightarrow SP and SP \rightarrow MC directions, respectively) are similar to our free energy barriers determined at the PBE-D2 level (7 kJ mol^{-1} for MC \rightarrow SP and 18 kJ mol^{-1} for SP \rightarrow MC).

3.4. Effect of Density Functional Approximation on the Free Energy of Activation

Insofar, the results obtained using the PBE-D2 method were discussed. Since the long range dispersion interactions are found to play a major role in the stabilization of reactant in SP (see Section 3.2), and since this contribution is treated only approximately within the DFT approach adapted here, it is useful to examine predictions made by other, more sophisticated, dispersion correction methods. The exploratory static results presented in Tab. 2 indeed suggest a significant variation in free energy barriers. It would be, however, prohibitively time consuming to repeat all the free energy calculations presented in Section 3.3 with different DFAs. Instead, we adapt here the recently proposed machine learning perturbation theory (MLPT) approach [56, 57] (see Section 2.4). Since the MD data for the PBE-D2 method are available, we shall consider this DFA as the production method. As the target methods, uncorrected PBE [63], PBE-MBD [71, 72, 73], vdW-DF2-B86R [74, 75, 76, 77, 78] and HSE06 [67, 68, 69, 70] are considered.

Since the PBE energies are readily available in the PBE-D2 calculations used

in Section 3.3, we shall start our discussion by this method treated at the brute force (i.e., without the machine learning acceleration) FEPT (direct application of Eq. 10). In this treatment, every tenth point of all relevant trajectories (i.e., 17100, 20000, 5000, and 5000 configurations for R', R(SP), TS(MC), and TS(SP), respectively) were employed to determine $\Delta\tilde{A}^\ddagger$ via Eq. 10 combined with the CE formula of Eq. 12. Note that following the notation of Section 2.4, a tilde is introduced to distinguish the target method quantities from those obtained using the production method. The computed values of activation free energies are $110.0 \text{ kJ mol}^{-1}$ (SP) and $118.7 \text{ kJ mol}^{-1}$ (MC'), which are to be compared with the PBE-D2 results of $102.4 \text{ kJ mol}^{-1}$ (SP) and $129.6 \text{ kJ mol}^{-1}$ (MC'). Hence, the absence of the long-range dispersion interactions leads to an increase in the free energy barrier of the reaction in SP and decrease for that in MC, which is a direct consequence of (i) a destabilization of the reactant (located in SP in both cases) and (ii) a stronger destabilization of TS(SP) compared to TS(MC) (cf. Tab. 1). Assuming that the free energies of the reactant states R(SP) and R' can be considered identical (as argued in Section 3.2), we conclude that the PBE method leads to qualitatively same prediction as PBE-D2, namely that the reaction involving TS formed in SP proceeds via lower free energy of activation and therefore it is kinetically preferred.

As already qualitatively discussed in Section 3.2, and as it is evident also from the free energy profile $\Delta\tilde{A}(\xi)$ shown in Fig. 10, which we determined at the PBE level using a generalization of Eq. 10 (see Ref. [57]):

$$\begin{aligned} \Delta\tilde{A}_{\xi_{ref},R \rightarrow \xi'} - \Delta A_{\xi_{ref},R \rightarrow \xi'} &= -k_B T \ln \frac{\langle Z^{-1/2} \exp[-\Delta V(\mathbf{r})/k_B T] \rangle_{\xi'}}{\langle Z^{-1/2} \rangle_{\xi'}} \\ &+ k_B T \ln \frac{\langle Z^{-1/2} \exp[-\Delta V(\mathbf{r})/k_B T] \rangle_{\xi_{ref},R}}{\langle Z^{-1/2} \rangle_{\xi_{ref},R}} \end{aligned} \quad (15)$$

the presence/absence of dispersion interactions affects also the adsorption ground state. Although this effect can not affect the conclusion made above (since the relative kinetics of competing reactions is given by the relative free energies of their respective transition states), it is conceptually interesting to determine $\Delta\tilde{A}^\ddagger$ with respect to the correct initial state which is, in the case of

the PBE method, the CO molecule located in MC. To this end, we determined the free energy difference between the states R' and R(MC) ($\Delta\tilde{A}_{R'\rightarrow R(MC)}$) using Eq.11, whereby $\Delta\tilde{A}_{\xi_{Ref,R'}\rightarrow\xi_{Ref,R(MC)}}=-0.6\text{ kJ mol}^{-1}$ was determined from $\Delta\tilde{A}(\xi)$, the probability density $P(\xi_{Ref,R(MC)})=0.16\text{ \AA}^{-1}$ was obtained using the MD simulations performed at the PBE level (see Section 3.2), and the density $P(\xi_{Ref,R'})=0.71\text{ \AA}^{-1}$ was determined by reweighting the data obtained from the straightforward MD performed at the PBE-D2 level as follows:

$$\tilde{P}(\xi_{ref,R'}) = \frac{\int d\mathbf{p} d\mathbf{r} \delta(\xi(\mathbf{r}) - \xi_{ref,R}) \exp[-\Delta V(\mathbf{r})/k_B T] \exp[-H(\mathbf{p}, \mathbf{r})/k_B T]}{\int_{\mathbf{r} \in R'} d\mathbf{p} d\mathbf{r} \exp[-\Delta V(\mathbf{r})/k_B T] \exp[-H(\mathbf{p}, \mathbf{r})/k_B T]}, \quad (16)$$

where $\mathbf{r} \in R'$ in the integral signifies all configurations belonging to the reactant state R'. Since the computed value of $\tilde{A}_{R'\rightarrow R(MC)}$ is -6.1 kJ mol^{-1} , the PBE activation energies $\Delta\tilde{A}(\xi)$ computed with respect to the correct adsorption groundstate are 116.1 kJ mol^{-1} (SP) and 124.8 kJ mol^{-1} (MC). A visual comparison of free energetics determined at the PBE and PBE-D2 levels is offered in Fig. 11.

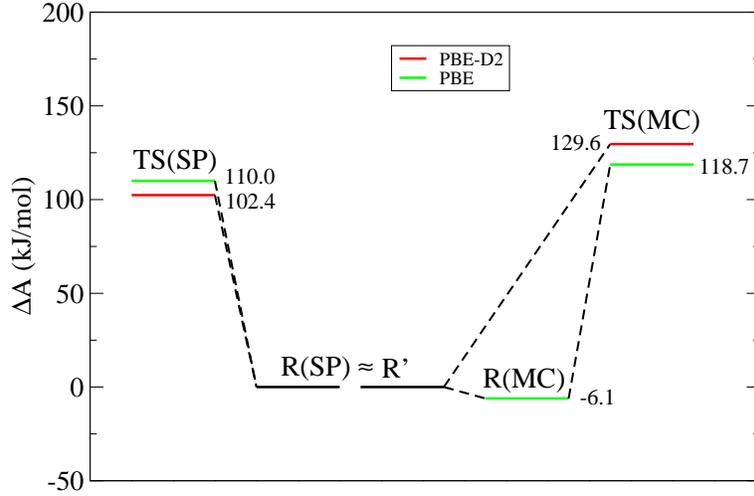


Figure 11: Free energies of activation of the methyl group carbonylation in SP and in MC of MOR at T=440 K computed via MD using the PBE-D2 method and perturbatively using the PBE method. The state R(SP) was chosen as a reference.

For the test purposes, we next determine the PBE free energy barriers also using the MLPT method (see Section 2.4), where the terms $\Delta V(\mathbf{r})$ needed in Eq. 10 are computed for the same points as considered in the brute force FEPT discussed above, but instead of using the exact values, the machine learning model trained on 250 configurations for each relevant state (unconstrained R and constrained TS) was applied. Only the reactions starting from the reactant in SP (i.e., R(SP) and R') were considered. In order to evaluate the quality of predictions of the ML model, we determined the mean error:

$$ME = \frac{\sum_{i=1}^N (\Delta V_{ref,i} - \Delta V_i)}{N}, \quad (17)$$

and root mean square error (RMSE) of predicted potential energy differences:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\Delta V_{ref,i} - \Delta V_i)^2}{N}}, \quad (18)$$

where $\Delta V_{ref,i}$ is the potential energy difference between the target and the production method explicitly computed for the configuration i , ΔV_i is that predicted using the ML, and the sum runs over all N members of the test set defined as every tenth configuration generated by MD at the production level. As evident from the numerical data compiled in Tab. 5, the ML model yields reliable predictions, with both the ME and RMSE being well below 1 kJ mol⁻¹. Consequently, the computed values of $\Delta \tilde{A}^\ddagger$ (110.2 kJ mol⁻¹ (SP) and 118.6 kJ mol⁻¹ (MC')) are found to be in excellent agreement with the results determined by the brute force FEPT (110.0 kJ mol⁻¹ (SP) and 118.6 kJ mol⁻¹ (MC')).

Next, we used the MLPT method trained on 250 configurations for each relevant state to determine the activation energies at the PBE-MBD, vdW-DF2-B86R, and HSE06 levels of theory. As in the case of the PBE method, we used the large test sets involving all configurations for which the MLPT predictions are made (16850, 19750, 4750, and 4750 configurations for R', R(SP), TS(MC), and TS(SP), respectively). The use of the set, for which the target method calculations have been performed, was twofold: (i) it allowed us to test the performance of our ML model in the $\Delta V(\mathbf{r})$ predictions, and (ii) it also provided us with an 'exact' FEPT reference allowing us to examine the effect

Table 5: Root mean square error and mean error (in parentheses), both in kJ mol^{-1} , calculated for energies ΔV_i predicted by the ML algorithm for different target methods. As a reference, exact values computed for every tenth configuration generated by MD at the production method level were used (except for HSE06, where exact values were computed for every 200th configuration).

	R(SP)	TS(SP)	R'	TS(MC)
PBE	0.9 (0.0)	0.9 (-0.1)	0.9 (0.0)	0.8 (0.0)
HSE06	4.1 (0.1)	4.2 (-0.6)	3.9 (-0.1)	3.8 (0.2)
PBE-MBD	1.2 (0.0)	1.3 (0.1)	1.2 (0.0)	1.4 (0.1)
vdW-DF2-B86R	1.2 (-0.1)	1.1 (0.0)	1.2 (0.0)	1.2 (0.0)

of ML on the free energy calculations. Owing to its very high computational cost, much smaller test sets (583, 750, 250 and 250 configurations for R', R(SP), TS(SP) and TS(MC), respectively) were defined at the HSE06 level. As shown in Tab. 5, the quality of $\Delta V(\mathbf{r})$ predicted for the PBE-MBD and vdW-DF2-B86R methods slightly deteriorated compared to the predictions done for the PBE target method, nevertheless, the predictions are still very accurate (RMSE of 1.1 - 1.4 kJ mol^{-1}). The very good performance of ML is not unexpected as both the production and the target methods are based on similar semilocal functionals (PBE or B86). A somewhat worse performance is found for the HSE06 functional, where RMSE values of 3.8 - 4.2 kJ mol^{-1} are found. Since the HSE06 differs more significantly from the production method than any other target method considered here, one can indeed expect that a bigger training set would be needed in order to achieve a similar quality of predictions as for the other DFAs considered here. In this respect it is important to note that the the systematic errors in ML model for R and TS are similar and hence their contribution to the error in ΔA^\ddagger will largely cancel, as showed in our previous work [57]. This fact is obvious also from the comparison of free energies of activation computed using MLPT with the exact FEPT reference presented in Fig. 12.

On the basis of these results, one can conclude that the MLPT method reaches virtually the same quality of predictions as the brute force FEPT, although the number of target method calculations is significantly reduced (250 vs. 5000-20000). Compared to the PBE-D2 results, the target method results follow closely the trend obvious from our exploratory single point calculations discussed in Section 3.1. In the case of the reaction in SP, for instance, the PBE-MBD, vdW-DF2-B86R, PBE, and HSE06 barriers are higher, respectively, by $\sim 3 \text{ kJ mol}^{-1}$, 5 kJ mol^{-1} , 8 kJ mol^{-1} , and 34 kJ mol^{-1} compared to the PBE-D2 result, and these differences are similar to the differences in electronic energy contributions to ΔA^\ddagger determined for fixed reactant and transition state geometries (see Tab. 2). The comparison of the MD results with those of the static approach (see Section SI in the Supporting Information) indicates that the two approaches yield very similar results for the reaction SP, while they differ significantly for the reaction in MC where the static approach predicts significantly lower free energy barriers. This result is not unexpected and it derives from the fact the static approach fails especially in cases where soft degrees of freedom such as the hindered rotations and translations contribute significantly to free energetics. In this respect, clearly, the reaction in MC is more challenging than in SP where both the reactant and transition states experience much more confining environment. Importantly, despite the large variation in numerical values of ΔA^\ddagger , all methods considered here lead to the qualitatively same prediction, namely that the reaction in SP is favored over that in MC. The same conclusion has been made also on the basis of experiment [37, 38].

4. Conclusions

In this work, detailed DFT simulations of the carbonylation of methoxy groups located in zeolite mordenite have been conducted. For the first time, the free energies have been determined via explicit sampling of all degrees of freedom at multiple levels of theory, involving the semi-local and hybrid density functional approximations PBE [63] and HSE06 [67, 68, 69, 70], as well as the

dispersion corrected methods PBE-D2 [63, 41], PBE-MBD [71, 72, 73], and vdW-DF2-B86R [74, 75, 76, 77, 78]. This was made possible by coupling the molecular dynamics simulations based on the blue moon ensemble technique [86, 88] with the recently introduced free energy perturbation theory accelerated by machine learning (MLPT) [56, 57]. Our approach allowed us to address the intensely discussed [33, 32, 34] problem of activity of active centers located in two different cavities available in MOR, namely the main channels and small channels. The overall picture drawn on the basis of our calculations is shown in Fig. 12.

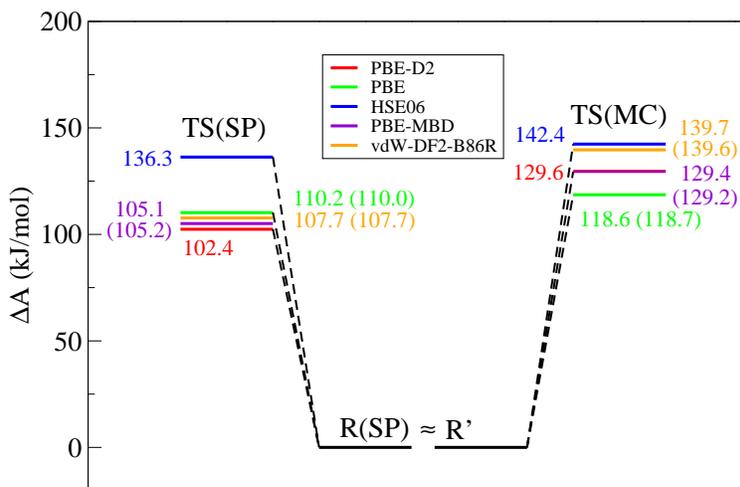


Figure 12: Overview of the free energies of activation of the methyl group carbonylation in SP and in MC of MOR at T=440 K computed using MD at the PBE-D2 level, and perturbatively at the PBE, HSE06, PBE-MBD, and vdW-DF2-B86R levels of theory. The brute force FEPT results (where the FEPT was applied to every tenth configuration generated by the production method (PBE-D2)) are given in parentheses.

The presence or absence of the long range dispersion interactions in the density functional approximation used in calculations is found to affect the adsorption ground state of reactant, which corresponds to the CO molecule adsorbed in MC or SP. This effect, however, does not have any impact on the relative kinetics of reactions in MC and SP, which is determined by the free energy

differences between the related transition states. Furthermore, including the long-range dispersion interactions in the DFA clearly tends to increase the differences between ΔA^\ddagger predicted for MC and SP, which are 6.1-8.4 kJ mol⁻¹ and 16.1-32.0 kJ mol⁻¹ when uncorrected (PBE and HSE06) and corrected (PBE-D2, PBE-MBD, and vdW-DF2-B86R) functionals are used, respectively. This result is consistent with the previous report based on the static approach [32], where inclusion of dispersion correction led to increase of difference in ΔA^\ddagger from 7.6 to 26.5 kJ mol⁻¹. Despite the large variation in computed free energies of activation, however, the qualitative prediction made by different DFAs remains the same, namely, the reaction in SP is kinetically preferred over that in MC. The strong preference of the reaction in SP has also been suggested experimentally [37, 38], although a rather indirect evidence has been offered. We note on passing that a definite proof of a higher activity of sites in SP will require systematic free energy calculations for all 14 possible CH₃ locations hypothetically allowed by the mordenite structure, which is beyond the scope of the present paper.

Acknowledgments

T.B. acknowledges support from Slovak Research and Development Agency under the Contracts No. APVV-15-0105 and No. VEGA-1/0777/19. D.R. acknowledges financial support through the COMETE project (COncEption in silico de Matériaux pour l'Environnement et l'Energie) co-funded by the European Union under the program "FEDER-FSE Lorraine et Massif des Vosges 2014-2020". Part of the results of this research have been achieved using the Computing Center of the Slovak Academy of Sciences acquired in projects ITMS 26230120002 and 26210120002 supported by the Research and Development Operational Program funded by the ERDF.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at (to be filled in by publisher).

Appendix B. Blue moon ensemble approach

In the blue moon ensemble approach, the free energy differences are determined by integrating free energy gradients along the reaction path connecting the states $\xi_{ref,R}$ and ξ^* :

$$\Delta A_{\xi_{ref,R} \rightarrow \xi^*} = \int_{\xi_{ref,R}}^{\xi^*} \left(\frac{\partial A}{\partial \xi} \right)_{\xi'} d\xi'. \quad (\text{B.1})$$

The free energy gradients are obtained using constrained MD simulations via the following formula [88]:

$$\left(\frac{\partial A}{\partial \xi} \right)_{\xi'} = \frac{1}{\langle Z^{-1/2} \rangle_{\xi'}} \langle Z^{-1/2} [-\lambda + \frac{k_B T}{Z^2} \sum_{i=1}^N \sum_{j=1}^N \frac{1}{m_i m_j} \sum_{\mu=x,y,z} \sum_{\nu=x,y,z} \frac{\partial \xi}{\partial r_{\mu,i}} \frac{\partial^2 \xi}{\partial r_{\mu,i} \partial r_{\nu,j}} \frac{\partial \xi}{\partial r_{\nu,j}}] \rangle_{\xi'}, \quad (\text{B.2})$$

where $\langle \dots \rangle_{\xi'}$ indicates the constrained ensemble average with $\xi(\mathbf{r}) = \xi'$, m_i and $r_{\mu,i}$ are the mass and the μ component of a Cartesian position vector of an atom i , respectively, λ is a Lagrange multiplier associated with the parameter ξ used in the SHAKE algorithm, [108] and Z is the inverse of the mass metric tensor defined as follows:

$$Z = \sum_{i=1}^N \frac{1}{m_i} \sum_{\mu=x,y,z} \left(\frac{\partial \xi}{\partial r_{\mu,i}} \right)^2. \quad (\text{B.3})$$

The numerical integration of Eq. B.1 is performed using the Simpson integration scheme adapted to irregularly spaced grid points [109, 110, 111].

References

- [1] J. Goldemberg, *Science* 315 (2007) 808–810.
- [2] J. Keller, *Hydrocarb. Process.* 58 (1979) 127–138.
- [3] I. Wender, *Fuel Process. Technol.* 48 (1995) 189–297.
- [4] R. G. Herman, *Catal. Today* 55 (2000) 233–245.
- [5] J. J. Spivey, A. Egbebi, *Chem. Soc. Rev.* 36 (2007) 1514–1528.
- [6] K. Cheng, J. Kang, D. L. King, V. Subramanian, C. Zhou, Q. Zhang, Y. Wang, Chapter three - advances in catalysis for syngas conversion to hydrocarbons, 2017.
- [7] W. Zhou, K. Cheng, J. Kang, C. Zhou, V. Subramanian, Q. Zhang, Y. Wang, *Chem. Soc. Rev.* 48 (2019) 3193–3228.
- [8] J. Li, Y. He, L. Tan, P. Zhang, X. Peng, A. Oruganti, G. Yang, H. Abe, Y. Wang, N. Tsubaki, *Nat. Catal.* 1 (2018) 787–793.
- [9] C. M. Fougret, W. F. Hölderich, *Appl. Catal. A-Gen.* 207 (2001) 295–301.
- [10] M. Bhasin, W. Bartley, P. Ellgen, T. Wilson, *J. Catal.* 54 (1978) 120–128.
- [11] M. Ichikawa, *Bull. Chem. Soc. Jpn.* 51 (1978) 2273–2277.
- [12] P. Courty, D. Durand, E. Freund, A. Sugier, *J. Mol. Catal.* 17 (1982) 241–254.
- [13] K. Smith, R. Anderson, *Can. J. Chem. Eng.* 61 (1983) 40–45.
- [14] J. Nunan, C. Bogdan, K. Klier, K. Smith, C. Young, R. Herman, *J. Catal.* 116 (1989) 195–221.
- [15] E. Tronconi, L. Lietti, P. Forzatti, I. Pasquon, *Appl. Catal.* 47 (1989) 317–333.

- [16] P. Courty, P. Chaumette, C. Raimbault, P. Travers, *Revue De L Institut Francais Du Petrole* 45 (1990) 561–578.
- [17] J. Dalmon, P. Chaumette, C. Mirodatos, *Catal. Today* 15 (1992) 101–127.
- [18] J. M. Christensen, P. M. Mortensen, R. Trane, P. A. Jensen, A. D. Jensen, *Appl. Catal. A-Gen.* 366 (2009) 29–43.
- [19] J. M. Christensen, P. A. Jensen, N. C. Schiodt, A. D. Jensen, *Chemcatchem* 2 (2010) 523–526.
- [20] J. M. Christensen, P. A. Jensen, A. D. Jensen, *Industrial & Engineering Chemistry Research* 50 (2011) 7949–7963.
- [21] Y. Ono, T. Mori, *J. Chem. Soc., Faraday Trans. 1* 77 (1981) 2209–2221.
- [22] A. G. Stepanov, M. V. Luzgin, V. N. Romannikov, V. N. Sidelnikov, K. I. Zamaraev, *J. Catal.* 164 (1996) 441–421.
- [23] M. V. Luzgin, V. N. Romannikov, A. G. Stepanov, K. I. Zamaraev, *J. Am. Chem. Soc.* 118 (1996) 10890–10891.
- [24] A. G. Stepanov, M. V. Luzgin, V. N. Romannikov, V. N. Sidelnikov, E. A. Paukshtis, *Journal of Catalysis* 178 (1998) 466–477.
- [25] H. Mori, A. Wada, Q. Xu, Y. Souma, *Chem. Lett.* 29 (2000) 136–137.
- [26] Q. Xu, S. Inoue, N. Tsumori, H. Mori, M. Kameda, M. Tanaka, M. Fujiwara, Y. Souma, *J. Mol. Catal. A Chem.* 170 (2001) 147–153.
- [27] N. Tsumori, Q. Xu, Y. Souma, H. Mori, *J. Mol. Catal. A Chem.* 179 (2002) 271–277.
- [28] T. Li, N. Tsumori, Y. Souma, Q. Xu, *Chem. Commun.* (2003) 2070–2071.
- [29] A. D. Chowdhury, K. Houben, G. T. Whiting, M. Mokhtar, A. M. Asiri, S. A. Al-Thabaiti, S. N. Basahel, M. Baldus, B. M. Weckhuysen, *Angew. Chem. Int. Ed.* 55 (2016) 15840–15845.

- [30] A. D. Chowdhury, A. L. Paioni, K. Houben, G. T. Whiting, M. Baldus, B. M. Weckhuysen, *Angew. Chem. Int. Ed.* 57 (2018) 8095–8099.
- [31] I. Yarulina, A. D. Chowdhury, F. Meirer, B. M. Weckhuysen, J. Gascon, *Nat. Catal.* 1 (2018) 398–411.
- [32] M. Boronat, C. Martínez, A. Corma, *Phys. Chem. Chem. Phys.* 13 (2011) 2603–2612.
- [33] M. Boronat, C. Martinez-Sanchez, D. Law, A. Corma, *J. Am. Chem. Soc.* 130 (2008) 16316–16323.
- [34] D. B. Rasmussen, J. M. Christensen, B. Temel, F. Studt, P. G. Moses, J. Rossmeisl, A. Riisager, A. D. Jensen, *Catal. Sci. Technol.* 7 (2017) 1141–1152.
- [35] P. N. Plessow, A. Smith, S. Tischer, F. Studt, *J. Am. Chem. Soc.* 141 (2019) 5908–5915.
- [36] K. Fujimoto, T. Shikada, K. Omata, H.-o. Tominaga, *Chem. Lett.* 13 (1984) 2047–2050.
- [37] A. Bhan, A. D. Allian, G. J. Sunley, D. J. Law, E. Iglesia, *J. Am. Chem. Soc.* 129 (2007) 4919–4924.
- [38] A. Bhan, E. Iglesia, *Accounts Chem. Res.* 41 (2008) 559–567.
- [39] A. D. Becke, *J. Chem. Phys.* 98 (1993) 5648–5652.
- [40] J. P. Perdew, Y. Wang, *Phys. Rev. B* 45 (1992) 13244–13249.
- [41] S. Grimme, *J. Comput. Chem.* 27 (2006) 1787–1799.
- [42] J. Wellendorff, K. T. Lundgaard, A. Mogelhoff, V. Petzold, D. D. Landis, J. K. Nørskov, T. Bligaard, K. W. Jacobsen, *Phys. Rev. B* 85 (2012) 235149.
- [43] H. Sun, *J. Phys. Chem. B* 102 (1998) 7338–7364.

- [44] Z. Liu, X. Yi, G. Wang, X. Tang, G. Li, L. Huang, A. Zheng, *Journal of Catalysis* 369 (2019) 335–344.
- [45] P. Cnudde, K. De Wispelaere, J. Van der Mynsbrugge, *J. Catal.* 345 (2017) 53–69.
- [46] J. Rey, A. Gomez, P. Raybaud, C. Chizallet, T. Bučko, *J. Catal.* 373 (2019) 361–373.
- [47] J. Rey, P. Raybaud, C. Chizallet, T. Bučko, *ACS Catal.* 9 (2019) 9813–9828.
- [48] T. Bučko, L. Benco, O. Dubay, C. Dellago, J. Hafner, *J. Chem. Phys.* 131 (2009) 214508.
- [49] J. Rey, C. Bignaud, P. Raybaud, T. Bučko, C. Chizallet, *Angew. Chem. Int. Ed.* 59 (2020) 18938–18942.
- [50] C. Tuma, T. Kerber, J. Sauer, *Angew. Chem. Int. Ed.* 49 (2010) 4678–4680.
- [51] G. Piccini, M. Alessio, J. Sauer, *Angew. Chem. Int. Ed.* 55 (2016) 5235–5237.
- [52] P. N. Plessow, F. Studt, *ACS Catal.* 7 (2017) 7987–7994.
- [53] D. Rocca, A. Dixit, M. Badawi, S. Lebègue, T. Gould, T. Bučko, *Phys. Rev. Materials* 3 (2019) 040801.
- [54] Q. Ren, M. Rybicki, J. Sauer, *J. Phys. Chem. C* 124 (2020) 10067–10078.
- [55] F. Berger, M. Rybicki, J. Sauer, *Journal of Catalysis* 395 (2021) 117 – 128.
- [56] B. Chehaibou, M. Badawi, T. Bučko, T. Bazhиров, D. Rocca, *J. Chem. Theory Comput.* 15 (2019) 6333–6342.
- [57] T. Bučko, M. Gešvandtnerová, D. Rocca, *J. Chem. Theory Comput.* 16 (2020) 6049–6060.

- [58] G. Kresse, J. Hafner, *Phys. Rev. B* 47 (1993) 558–561.
- [59] G. Kresse, J. Hafner, *Phys. Rev. B* 49 (1994) 14251–14269.
- [60] G. Kresse, J. Furthmuller, *Phys. Rev. B* 54 (1996) 11169–11186.
- [61] P. E. Blochl, *Phys. Rev. B* 50 (1994) 17953–17979.
- [62] G. Kresse, D. Joubert, *Phys. Rev. B* 59 (1999) 1758–1775.
- [63] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* 77 (1996) 3865–3868.
- [64] F. Göttl, A. Grüneis, T. Bučko, J. Hafner, *J. Chem. Phys.* 137 (2012) 114111.
- [65] M.-C. Silaghi, C. Chizallet, E. Petracovschi, T. Kerber, J. Sauer, P. Raybaud, *ACS Catal.* 5 (2015) 11–15.
- [66] T. Bučko, J. Hafner, S. Lebégue, J. G. Ángyán, *J. Phys. Chem. A* 114 (2010) 11814–11824.
- [67] J. Heyd, G. E. Scuseria, M. Ernzerhof, *J. Chem. Phys.* 118 (2003) 8207–8215.
- [68] J. Heyd, G. E. Scuseria, *J. Chem. Phys.* 121 (2004) 1187–1192.
- [69] J. Heyd, G. E. Scuseria, M. Ernzerhof, *The Journal of Chemical Physics* 124 (2006) 219906.
- [70] A. V. Kruckau, O. A. Vydrov, A. F. Izmaylov, G. E. Scuseria, *J. Chem. Phys.* 125 (2006) 224106.
- [71] A. Ambrosetti, A. M. Reilly, R. A. DiStasio, Jr., A. Tkatchenko, *J. Chem. Phys.* 140 (2014).
- [72] A. Tkatchenko, R. A. DiStasio, R. Car, M. Scheffler, *Phys. Rev. Lett.* 108 (2012) 236402.

- [73] T. Bučko, S. Lebègue, T. Gould, J. G. Ángyán, *J. Phys.: Condens. Matter* 28 (2016) 045201.
- [74] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, B. I. Lundqvist, *Phys. Rev. Lett.* 92 (2004) 246401.
- [75] G. Román-Pérez, J. M. Soler, *Phys. Rev. Lett.* 103 (2009) 096102.
- [76] J. Klimeš, D. R. Bowler, A. Michaelides, *Phys. Rev. B* 83 (2011) 195131.
- [77] K. Lee, E. D. Murray, L. Kong, B. I. Lundqvist, D. C. Langreth, *Phys. Rev. B* 82 (2010) 081101.
- [78] I. Hamada, *Phys. Rev. B* 89 (2014) 121103.
- [79] G. Henkelman, H. Jónsson, *J. Chem. Phys.* 111 (1999) 7010–7022.
- [80] A. Heyden, A. T. Bell, F. J. Keil, *J. Chem. Phys.* 123 (2005) 224101.
- [81] K. Fukui, *J. Phys. Chem.* 74 (1970) 4161–4163.
- [82] K. Fukui, *Acc. Chem. Res.* 14 (1981) 363–368.
- [83] H. P. Hratchian, H. B. Schlegel, *J. Phys. Chem. A* 106 (2002) 165–169.
- [84] D. Frenkel, B. Smit, in: D. Frenkel, B. Smit (Eds.), *Understanding Molecular Simulation (Second Edition)*, Academic Press, San Diego, second edition edition, 2002, pp. 139–163.
- [85] L. Verlet, *Phys. Rev.* 159 (1967) 98–103.
- [86] E. Carter, G. Ciccotti, J. T. Hynes, R. Kapral, *Chem. Phys. Lett.* 156 (1989) 472 – 477.
- [87] T. Bučko, S. Chibani, J.-F. Paul, L. Cantrel, M. Badawi, *Phys. Chem. Chem. Phys.* 19 (2017) 27530–27543.
- [88] M. Sprik, G. Ciccotti, *J. Chem. Phys.* 109 (1998) 7737–7744.
- [89] R. W. Zwanzig, *The Journal of Chemical Physics* 22 (1954) 1420–1426.

- [90] C. Chipot, A. Pohorille, Free energy calculations: Theory and Applications in Chemistry and Biology, Springer, 2016.
- [91] R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. von Lilienfeld, J. Chem. Theory Comput. 11 (2015) 2087–2096.
- [92] M. Rupp, Int. J. Quantum Chem. 115 (2015) 1058–1073.
- [93] S. De, A. P. Bartók, G. Csányi, M. Ceriotti, Phys. Chem. Chem. Phys. 18 (2016) 13754–13769.
- [94] A. P. Bartók, R. Kondor, G. Csányi, Phys. Rev. B 87 (2013) 184115.
- [95] L. Himanen, M. O. Jäger, E. V. Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke, A. S. Foster, Comput. Phys. Commun. 247 (2020) 106949.
- [96] L. Himanen, M. O. Jäger, E. V. Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke, A. S. Foster, Dscribe: Library of descriptors for machine learning in materials science, 2020. <https://github.com/SINGROUP/dscribe>.
- [97] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, J. Mach. Learn. Res. 12 (2011) 2825–2830.
- [98] C. Baerlocher, L. B. McCusker, Database of zeolite structures, 2017. <https://www.iza-structure.org/databases/>.
- [99] X. Yi, Y. Xiao, G. Li, Z. Liu, W. Chen, S.-B. Liu, A. Zheng, Chem. Mater. 32 (2020) 1332–1342.
- [100] K. Momma, F. Izumi, J. Appl. Crystallogr. 44 (2011) 1272–1276.
- [101] A. Alberti, P. Davoli, G. Vezzalini, Z. Kristallogr. Cryst. Mater. 175 (1986) 249–256.

- [102] F. Göttl, P. Sautet, *J. Chem. Phys.* 140 (2014) 154105.
- [103] J. Hermann, R. A. DiStasio, A. Tkatchenko, *Chem. Rev.* 117 (2017) 4714–4758.
- [104] L. M. Woods, D. A. R. Dalvit, A. Tkatchenko, P. Rodriguez-Lopez, A. W. Rodriguez, R. Podgornik, *Rev. Mod. Phys.* 88 (2016) 045003.
- [105] T. Bučko, L. Benco, J. Hafner, J. G. Ángyán, *J. Catal.* 279 (2011) 220–228.
- [106] J. Hajek, J. Van der Mynsbrugge, K. De Wispelaere, P. Cnudde, L. Vanduyfhuys, M. Waroquier, V. Van Speybroeck, *J. Catal.* 340 (2016) 227–235.
- [107] P. Cnudde, K. De Wispelaere, L. Vanduyfhuys, R. Demuyne, J. Van der Mynsbrugge, M. Waroquier, V. Van Speybroeck, *ACS Catal.* 8 (2018) 9579–9595.
- [108] J.-P. Ryckaert, G. Ciccotti, H. J. Berendsen, *J. Comput. Phys.* 23 (1977) 327 – 341.
- [109] N. Shklov, *Am. Math. Mon.* 67 (1960) 1022.
- [110] W. H. Press (Ed.), *Numerical recipes: the art of scientific computing*, Cambridge University Press, Cambridge, UK ; New York, 3rd ed edition, 2007.
- [111] R. Cameron, B., *Math. Enthus.* 11 (2014) 643–648.