FACULTY OF ENGINEERING
AND ARCHITECTURE

Towards State-of-the-Art Molecular Simulations for an Accurate Modeling of Intricate Zeolite-Catalyzed Reactions

Massimo Bocus

Doctoral dissertation submitted to obtain the academic degree of
Doctor of Chemical Engineering

Supervisor
Prof. Veronique Van Speybroeck, PhD
Department of Applied Physics
Faculty of Engineering and Architecture, Ghent University

June 2023

GHENT
UNIVERSITY

ISBN 978-94-6355-732-0
NUR 913, 928
Wettelijk depot: D/2023/10.500/64

Members of the Examination Board

Chair

Prof. Em. Luc Taerwe, PhD, Ghent University

Other members entitled to vote

Mercedes Boronat Zaragozá, PhD, Universitat Politècnica de València, Spain
Prof. Maarten Sabbe, PhD, Ghent University
Prof. Louis Vanduyfhuys, PhD, Ghent University
Jelle Vekeman, PhD, Ghent University

Supervisor

Prof. Veronique Van Speybroeck, PhD, Ghent University

This research has been conducted at the **Center for Molecular Modeling**,
in collaboration with:

- Centre for Membrane separations, Adsorption, Catalysis and Spectroscopy for Sustainable Solutions, KU Leuven (Prof. D. E. De Vos).
- Eenheid Algemene Chemie, Vrije Universiteit Brussel (Prof. F. De Proft).
- Inorganic Chemistry and Catalysis Group, Debye Institute for Nanomaterials Science, Utrecht University (Prof. B. M. Weckhuysen).
- Center for Sustainable Catalysis and Engineering, KU Leuven (Prof. B. F. Sels).
- Organic Synthesis Research Division, University of Antwerp (Prof. B. U. W. Maes)

I am also grateful to Prof. J. Hutter (University of Zurich) for allowing me to spend two weeks in his research group during April-May 2022.

Contents

Preface

These past four years have been quite an adventure, but I am glad I undertook it. I am grateful for all the support I received from many people, without whom I could have never achieved this milestone. While I will do my best to acknowledge everyone, I sincerely hope the reader can forgive me if I inadvertently overlook someone, understanding that it was an unintentional oversight resulting from my known distracted nature.

First, my sincere gratitude goes to my advisor, Veronique. She has been a true inspiration during my PhD and I have always admired her scientific integrity and her ability of managing a group as large as the CMM. She has constantly supported and guided me throughout these years, indulging me whenever I was proposing some new idea to try out. She deserves a lot of credit for the research and the results that are presented throughout this thesis.

I am also grateful to the members of the examination board for taking their time in reading this manuscript. Their comments and the discussion that stemmed therefrom during the internal defense were really thought-provoking and I had quite some fun when formulating my answers. Besides, they allowed me to learn new concepts that I would have not necessarily encountered otherwise.

Many results in this thesis have only been possible thanks to a handful of experimental and theoretical scientists outside Ghent University with whom I had the pleasure and honor to collaborate. I would therefore like to thank Jannick Vercammen and Prof. Dirk De Vos (KU Leuven) for making us part of their outstanding work on palladium-loaded zeolites. Prof. Andrew Beale (UCL) and his group for collaborating with us when unraveling the deactivation mechanism of the MTO process. Prof. Bert Weckhuysen (Utrecht University) and Prof. Frank De Proft (VUB) for their advice and proofreading of the benzene ethylation manuscript. Mathias Bal and Prof. Bert Maes (UAntwerp), Xian Wu and Prof. Bert Sels (KU Leuven) for their valuable experimental input in the guaiacol *O*-demethylation project. Finally, Prof. Juerg Hutter (University of Zurich), Frederick Stein and the rest of his group for welcoming me and sharing their expertise during my visit there, being always willing to help me whenever I had an issue.

I am very grateful to all the CMM members I could meet and work with during the past four years for the interesting discussions, the nice after-work activities and, in general, for creating a pleasant and stimulating work environment. While I will not list everybody by name for the sake of this preface's length, there are some special mentions that must be made. First, I want to thank Chiara and Julianna for being so friendly and supportive in the first months of my PhD, in spite of my twenty-thousand questions per day. Then Alexander, Aran, Elias, Louis, Maarten, Pieter and Sam for the great collaborations throughout these years which, in all cases, have led to results of a quality I could have never hoped to achieve by myself. A big thanks also goes to Leen and Mieke, whose invaluable help and constant kindness makes the life of every finishing PhD student much, much easier.

I am thankful to Bjarne, Bram and Wim, the master's thesis students that I currently have the pleasure to supervise, for their patience towards my inconsistent availability when trying to follow their work while finalizing this PhD thesis. You have all done an outstanding work and I can only wish you the best for your future careers, where I am confident you will achieve amazing successes.

Starting a completely new life abroad can be challenging, but luckily I had the chance of meeting some great friends whose kindness and time together I will hardly ever forget. I therefore want to warmly thank the CMM kids Liesbeth, Karen and Siebe for immediately making me feel a bit Belgian. Also part of the CMM kids are 'the boys' Tom and Juul, who I thank for being the best office mates and west-Flemish teachers one could desire, and Ruben and Sander, who are not only two amazing friends but also exceptional housemates. Unfortunately, almost two years of my time in Ghent were spent mostly in lockdown due to the COVID pandemic. That time would have been terrible without an outstanding group of friends: Leo, Michael, Tomi and YingXing. They made the pandemic time a lot less boring with beautiful cycling trips and game nights. Among all my friends, a thanks also goes to Jenna, Kuber, Loïc and Liselotte, to Roberta, Thomas and Pablo and to Marija and Martijn for all the drinks, dinners, field trips and much more.

Per concludere, mi devo spostare dal Belgio all'Italia per ringraziare alcune persone molto importanti. In primis tutti i miei amici di Robilante, del secondo piano del Collegio e dell'Uni, per esserci sempre nonostante la lontananza. Poi ovviamente la mia famiglia, che mi ha costantemente sostenuto in tutti questi anni. In particolare, un grosso grazie va alla mia sorellona Carol, che é da sempre un saldo punto di riferimento nonché un esempio per me (senza contare che buona parte dell'energia necessaria per completare questa tesi proviene dai suoi manicaretti). Ed ultimi, ma primi per importanza, i miei genitori per avermi sostenuto in maniera costante negli ultimi 29 anni. Sembra scontato a dirsi, ma se oggi sono qui seduto a scrivere questa prefazione, é esclusivamente merito loro.

Massimo Bocus
Ghent, 9 June 2023

Samenvatting

Brønsted-zure zeolieten zijn momenteel één van de belangrijkste industriële katalysatoren voor de conversie van koolwaterstoffen. Zo zijn ze omnipresent in belangrijke processen zoals vloeibaar katalytisch kraken van koolwaterstoffen of conversie van methanol naar lichte olefines. Daarenboven is het recent duidelijk geworden dat ze ook een belangrijke rol zullen blijven spelen in opkomende technologieën zoals CO_2-naar-koolwaterstoffen en biomassa conversie. Hun brede inzetbaarheid kan ondermeer terug gebracht worden naar hun robuustheid. Omdat ze bestaan uit een kristallijn raamwerk van onderling verbonden SiO_4-tetraëders, kunnen ze zeer hoge temperaturen, drukken en een diversiteit aan chemische omgevingen weerstaan. Zoals zal duidelijk worden in deze thesis, kunnen ze ook gebruikt worden in vloeibaar water op hoge temperatuur zonder volledig te desintegreren, wat het geval zou zijn voor de meeste andere materialen. Op de nanoschaal bestaan zeolieten uit een kristallijn netwerk van microporiën en holtes van moleculaire afmetingen. Deze moleculaire dimensies maken hen bijzonder aantrekkelijk voor katalyse en separatie. Om actieve sites te creëren worden sommige Si^{4+} atomen vervangen door Al^{3+} wat nadien ladingsgecompenseerd dient te worden. Dit kan gebeuren door verschillende kationen maar in vele gevallen wordt er een extra proton gehecht aan het aluminum waardoor er een Brønsted-zure site (BAS, Brønsted acid site) wordt gecreëerd. De BAS is de actieve site waar de meeste reacties worden gekatalyseerd. Om een geschikte katalysator te ontwerpen, is het bovendien nodig om een gepast rooster met de goede topologie te selecteren. Op dit moment werden er reeds 240 zeolieten succesvol gesynthetiseerd.

Tot dusver is gemotiveerd ontwerp van zeolieten dikwijls nog een utopie en wordt er vaak gebruik gemaakt van een trial-and-error procedure, waarbij vele zeolieten worden geprobeerd en het best presterende materiaal wordt geselecteerd om de reactie uit te voeren. Bovendien is er nog een extra complexiteit in zeolieten katalyse. De éénvoudige actieve site – hierboven geïdentificeerd als BAS- is in werkelijkheid veel complexer dan enkel een lokaal gesubstitueerde site van silicium door aluminium en een proton. Zo kunnen verschillende niet-conventionele actieve sites voorkomen waarbij er bepaalde defecten optreden zoals bijvoorbeeld extra rooster aluminium species of gesolvateerde H_3O^+-ionen. Het proton is in werkelijkheid veel mobieler dan oorspronkelijk gedacht en kan zich ook hechten aan andere moleculen of reactieve species in de omgeving van de oorspronkelijke actieve site. Rekening houdende met deze complexiteiten is het tot op vandaag de dag nog niet mogelijk om uitgaande van rationele basisprincipes een volledig optimaal zeoliet ontwerp door te voeren dat optimaal is voor een bepaald proces bij bepaalde procescondities.

Voorgaande geeft ogenschijnlijk een negatieve indruk, maar is tezelfdertijd de belangrijkste drijfveer voor het onderwerp van dit doctoraatsproefschrift. Binnen dit proefschrift werd gepoogd om geavanceerde moleculaire modelleringsmethoden toe te passen en te ontwerpen om zeoliet-gekatalyseerde reacties te modelleren in omstandigheden die zo goed als mogelijk de realistische werkingscondities nabootsen. Het doel van dergelijke geavanceerde modelleringssimulaties, is het bekomen van mechanistische details, reactietrends en betrouwbare reactiesnelheidsconstanten. Gezien de merkwaardige complexiteit zoals hierboven beschreven, is het duidelijk dat men op atomair niveau naar het materiaal moet kijken om de beschikbare actieve sites en moleculaire gedragingen te karakteriseren. Dankzij de introductie van de kwantummechanica meer dan een eeuw geleden en de integratie van deze principes binnen de computationele chemie en katalyse, is het vandaag de dag mogelijk om vrij betrouwbaar informatie te bekomen over de gekatalyseerde reactie die wordt bestudeerd. Binnen mijn onderzoek werden inzichten bekomen in een aantal specifieke processen en reacties en in die zin hoop ik dat deze thesis een bijdrage kan leveren aan de vooruitgang waarbij modellering niet enkel kan gezien worden als een verklarende tool waarbij experimentele waarnemingen worden geduid en beter begrepen maar ook meer en meer kan evolueren naar een voorspellende discipline waarbij modellering ook de experimentele uitkomst zou kunnen voorspellen. Het is duidelijk dat we vandaag de dag dit niveau nog niet hebben bereikt, maar met de komst van sterke computers en geavanceerde theoretische en numerieke modellen kan dit einddoel hopelijk ooit gerealiseerd worden.

Deze thesis is opgebouwd uit zes hoofdstukken waarvan de eerste twee inleidende informatie bevatten over het toepassingsgebied en de gebruikte modelleringstechnieken en de overige hoofdstukken de belangrijkste resultaten weergeven alsook de conclusies en perspectieven.

In het eerste resultaathoofdstuk (Hoofdstuk 3), worden statische simulaties geïntroduceerd om zeoliet-gekatalyseerde reacties te modelleren. In deze methodologie worden slechts enkele punten op het potentiële energieoppervlak in beschouwing genomen. Deze aanpak is goed ingeburgerd en kan erg nuttig zijn om interessante mechanistische inzichten te verkrijgen in complexe reactienetwerken. Twee gevalstudies worden grondige uitgewerkt. De eerste gaat over de dealkylering van alkylfenolen gekatalyseerd door H-ZSM-5 zeoliet in aanwezigheid van stoom. Deze reactie is een belangrijke stap in de productie van basischemicaliën uit biomassa. Ik gebruikte statische berekeningen om de verschillen in reactiviteit tussen verschillende substraten te begrijpen en ik kon enkele experimentele waarnemingen die in de literatuur zijn gerapporteerd duiden vanuit moleculair oogpunt. De tweede case-studie betreft de Pd-gekatalyseerde tolueen-homokoppeling. Dit is een gecombineerde experimentele-theoretische studie uitgevoerd met onze experimentele partners van KULeuven (Prof. D. De Vos). Zij ontdekten dat palladiumacetaat geabsorbeerd in H-Beta zeoliet deze reactie met een hoge selectiviteit kan katalyseren. Vanuit experimenteel standpunt alleen was het echter niet duidelijk wat aan de oorsprong lag van de uitzonderlijke selectiviteit. Dit was dan ook de motivering om het reactiemechanisme te bestuderen aan de hand van simulaties. Hoewel het merendeel van de resultaten werden bekomen met statische berekeningsmethodieken, werden er aanvullend ook verkennende moleculaire dynamica simulaties uitgevoerd om de configuratieruimte van actieve sites en reactieve componenten te onderzoeken.

In het tweede resultaathoofdstuk (Hoofdstuk 4), worden de modelleringstechnieken naar een complexer niveau gebracht teneinde ook reacties te simuleren op operationele condities. Geavanceerde moleculaire dynamica simulaties werden gebruikt om de rol van water te onderzoeken bij het moduleren van de reactiesnelheid voor twee interessante gevallen. De eerste is de alkylering van benzeen met etheen of ethanol in H–ZSM–5. Naast een volledig mechanistisch onderzoek tonen we aan dat een kleine hoeveelheid water de snelheid van protonoverdrachtsreacties opmerkelijk kan versnellen. De tweede reactie kadert opnieuw in het kader van biomassaconversie en betreft de *O*-demethylering van guaiacol uitgevoerd in vloeibaar water op hoge temperatuur. In samenwerking met mijn collega E. Van Den Broeck, simuleerden we de katalyse door H_3O^+-ionen zowel in bulkwater als in de beperkte omgeving van het zeoliet. Hierbij hebben we de moleculaire redenen blootgelegd waarom een hogere activiteit wordt waargenomen indien de reactie wordt gekatalyseerd door een zeoliet in vergelijking met een homogene HCl-katalysator. Onze computationele resultaten werden ondersteund door experimenteel onderzoek uitgevoerd door onze experimentele partners aan de Universiteit van Antwerpen (Prof. B.U.W. Maes) en de KU Leuven (Prof. B.F. Sels).

Het derde resultaathoofdstuk (Hoofdstuk 5) worden 'machine learning' potentialen (MLPs) als veelbelovende methode om berekeningen significant te versnellen geïntroduceerd. Dergelijke MLPs kennen recent een grote opgang gezien ze toelaten om op een computationeel veel efficiëntere wijze het potentieel energie-oppervlak te verkennen terwijl ze nog steeds de nauwkeurigheid behouden van onderliggende kwantummechanische trainingsdata. Het onderzoek gerapporteerd in dit hoofdstuk is het resultaat van een samenwerking met mijn collega's R. Goeminne, A. Lamaire en M. Cools-Ceuppens. Door een nauwkeurige MLP te trainen op basis van gestuurde moleculaire dynamica simulaties op hoge temperatuur, werd het mogelijk om het onderliggende potentieel energie-oppervlak met hoge accuraatheid te komen aan een zeer lage computationele kost. Dit gaf ons toegang tot eerder niet berekenbare eigenschappen, zoals het kinetische isotoopeffect van de reactie en de invloed van nucleaire kwantumeffecten over een breed temperatuurbereik.

Ik hoop dat de inhoud van de thesis en de structuur waarbij systematisch meer complexe methodologieën worden gebruikt om zeoliet gekatalyseerde reacties te simuleren, hetzelfde fascinerende gevoel kan geven als ik heb ervaren de afgelopen vier jaar. Het is duidelijk dat het veld van de computationele chemie en de zeolietkatalyse, zeer snel evolueert. Er liggen nog vele boeiende uitdagingen op het pad naar het ultieme doel van de moleculaire modellering waarbij niet enkel experimenten kunnen worden verklaard maar ook kunnen voorspeld en waarbij zeoliet processen kunnen ontworpen worden vanuit moleculaire inzichten.

Summary

Brønsted-acidic zeolites are one of nowadays' most important industrial catalysts for the conversion of hydrocarbons. They are widely applied in consolidated processes such as fluid catalytic cracking or methanol-to-olefins and are also gaining a primary role in emerging technologies among which CO_2-to-hydrocarbons and biomass conversion. This is made possible by their sturdiness. Consisting of a crystalline framework of interconnected SiO_4 tetrahedra, they can withstand very high temperatures and pressures and a variety of chemical environments. Up to a certain extent, zeolites can even be used in hot liquid water without completely disintegrating as most other materials would. At the nanoscale, zeolites are made of a crystalline network of micropores and pockets of molecular dimensions. There, some Si^{4+} atoms are replaced by Al^{3+} and a proton, which is the Brønsted acid site (BAS) where most reactions are catalyzed. Having such kind of confinement around the active site means that the most suited zeolite can be selected to accommodate reactants, transition states and products of the desired reaction from a menu of about 240 frameworks which are currently possible to synthesize.

Obviously, the previous statement is extremely utopian and most often experiments resort to a trial-and-error procedure, where many frameworks are tested and the best performing material is then selected to perform the reaction. Additionally, a further degree of complexity comes into play when working with zeolites. The simple, clean and pristine BAS, often considered as the main active site in the material, is in reality only one representative of a variegated family of unconventional active sites. Years of research have shown that zeolites are, in reality, extremely complex materials at the nanoscale and a proper understanding of all their defective active sites—ranging from extra framework aluminum species to solvated hydronium ions—is more of a hope than a true possibility. So far, it seems therefore impossible to imagine a way of tailoring the right zeolite catalyst and catalytic conditions based on the reaction one wants to perform.

While this sounds like—and in a sense is—a negative statement, it is also one of the main reasons why this thesis exists in the first place. Here, I try to apply state-of-the-art molecular modeling methodologies to study zeolite-catalyzed reactions as close as possible to their true working conditions, with the purpose of deriving mechanistic details, reactivity trends and reliable reaction rates. Given the outstanding complexity outlined above, it is clear that one needs to look at the material at an atomic scale in order to characterize the available active sites and molecular behaviors. With the tools enabled by the introduction of quantum mechanics more than a century ago, we can now obtain quite reliable information on the catalytic reaction under study. While my investigation provided some insights in specific processes and reactions, I hope it can be seen as a small contribution to our stride in going from explanatory modeling (i.e. to explain experimental observations) to predictive modeling (i.e. to predict the experimental outcome).

After two introductory chapters, the thesis is organized in three main sections. In the first, I introduce static simulations to model zeolite-catalyzed reactions, where only few points on the potential energy surface are considered. This approach is well consolidated and can be very useful to obtain interesting mechanistic insights in complex reaction networks. Two case-studies are presented. The first one deals with the dealkylation of alkylphenols catalyzed by the H–ZSM–5 zeolite in the presence of steam. This reaction is an important step in the production of commodity chemicals from biomass. I used static calculations to understand the differences in reactivity between different substrates and I was able to explain some of the experimental observations reported in literature. The second case study concerns the Pd-catalyzed toluene homocoupling. This is a combined experimental-theoretical study performed with our collaborators in KU Leuven (Prof. D. De Vos). They found that palladium acetate adsorbed in the H–Beta zeolite can catalyze the reaction with high selectivity and, therefore, we used simulations to explore the reaction mechanism and explain the products distributions obtained experimentally. Albeit the majority of both studies relies on static simulations, additional exploratory molecular dynamics was performed as well to investigate the configurational space of active sites and reactive species.

The second section goes a step forward towards realistic operando conditions. Enhanced sampling molecular dynamics is used to investigate the role of water in modulating reaction kinetics for two interesting reactions. The first is the alkylation of benzene with ethene or ethanol in H–ZSM–5. Aside from a full mechanistic investigation, we show that trace amount of water can remarkably speed-up the rate of proton transfer reactions. The second reaction, again in the context of biomass conversion, is the *O*-demethylation

of guaiacol performed in hot liquid water. Working together with my colleague E. Van Den Broeck, we simulated the catalysis by hydronium ions both in bulk water and in the zeolite constrained environment, highlighting the molecular reasons underlying the higher activity of the zeolite with respect to the homogeneous HCl catalyst. Our computational results have been supported by the experimental investigation performed by our collaborators in the University of Antwerp (Prof. B. U. W. Maes) and the KU Leuven (Prof. B. F. Sels).

Finally, the last section introduces machine learning potentials (MLPs) as a promising method to speed up calculations by many orders of magnitude while retaining the accuracy of first-principles methodologies. This was a joined effort with my colleagues R. Goeminne, A. Lamaire and M. Cools-Ceuppens. By training an accurate MLP on high-temperature enhanced sampling simulations, we obtained a very cheap estimator of the original DFT potential energy surface. This gave us access to previously unattainable computed quantities, such as the reaction kinetic isotope effect and the influence of nuclear quantum effects over a broad temperature range.

I hope that the structure of the thesis, which is framed as a gradual climb towards more realistic/more efficient computational methodologies, will be able to transmit to the reader the same fascination I felt in the past four years every time that I was learning or trying a new computational methodology, seeing how far I could go before a new advancement was required. Given the speed at which computational chemistry is evolving, I don't think that we are close to get bored yet and many exciting discoveries still separate us from the grail of molecular modeling, i.e. reaching a sufficient predictive power to not only explain but also guide the experiment.

List of Abbreviations

AIMD	Ab initio molecular dynamics
BAS	Brønsted acid site
B3LYP	Becke 3-parameter Lee–Yang–Parr functional
BLYP	Becke–Lee–Yang–Parr functional
CCSD(T)	Couple cluster with singles, doubles and perturbative triples
CMD	Concerted metalation-deprotonation
CN	Coordination number
CV	Collective variable
DFT	Density functional theory
EFAL	Extra-framework aluminum
EXAFS	Extended X-ray Absorption Fine Structure
FCC	Fluid catalytic craking
FEP	Free energy profile
FEPT	Free energy perturbation theory
FES	Free energy surface
GGA	Generalized gradient approximation
GTH	Goedecker–Teter–Hutter pseudopotential
HCP	Hydrocarbon pool
HF	Hartree-Fock
HPW	Hot pressurized water
KIE	Kinetic isotope effect
KRR	Kernel ridge regression
LAS	Lewis acid site
LSDA	Local spin density approximation
MBAR	Multistate Bennett acceptance ratio
MD	Molecular dynamics
MI	Migratory insertion
MLP	Machine learning potential
MP2	Møller–Plesset second order perturbation theory
MPNN	Message passing neural network
MTH	Methanol to hydrocarbons

MTO	Methanol to olefins
NEB	Nudget elastic band
NMA	Normal mode analysis
NMR	Nuclear magnetic resonance
NQE	Nuclear quantum effect
PAW	Projector augmented wave
PBE	Perdew–Burke–Ernzerhof functional
PES	Potential energy surface
PHVA	Partial Hessian vibrational analysis
PIMD	Path integral molecular dynamics
PW	Plane wave
RPA	Random phase approximation
RPMD	Ring polymer molecular dynamics
SAS	Surface alkoxide species
SES	Surface ethoxide species
SBU	Secondary building unit
TS	Transition state
TST	Transition state theory
US	Umbrella sampling
VASP	Vienna ab initio simulation package
WHAM	Weighted histogram analysis method
XC	Exchange-correlation

Part I

Towards state-of-the-art molecular simulations for an accurate modeling of intricate zeolite-catalyzed reactions

1

Introduction

From the common public's perspective, the job of a chemist remains traditionally linked to the laboratory. There, colorful substances are mixed and shaken to produce thick smokes, fire and explosions. Despite being nowadays very far from this stereotypical picture, practical work remains an essential component of the discipline. The question arises as to whether experiments are capable of providing every detail regarding the behavior of molecules and materials, or if something more is needed.

One of my favorite examples in this respect is a paper from Wang and co-workers, titled *"Will any crap we put into graphene increase its electrocatalytic effect?"*[1]. The authors, clearly overwhelmed by the countless reports claiming that this and that dopant were able to enhance the electrocatalytic properties of graphene, conceived a rather interesting experiment: they doped graphene with bird droppings and—without much of a surprise—its electrocatalytic properties resulted to be improved. This observation led the author to propose in a humoristic way the possibility of tuning graphene properties based on the diet of the birds from which the droppings are harvested.

In spite of the comic fashion in which the paper results are framed, they highlight one of the main challenges that experiments have to face. The outstanding complexity of most real materials and chemical systems makes it extremely arduous to directly bridge measured properties with the elementary behavior of the system atoms. If one can pinpoint the source of a certain experimental observation based on the changes occurring at the nanoscale, it becomes then possible to rationally design new reactions and materials that are tailored towards specific applications.

This is the core purpose of **computational chemistry**, the field of study in which this thesis is situated. Since the revolutionary ideas of quantum mechanics were laid down at the dawn of the XXth century by physicists such as Bohr, Planck, Dirac, Schrödinger and Einstein (among others), it became soon evident that the mathematical foundations to describe the behavior of subatomic particles and—by extension—of atoms, molecules and materials were now established. However, the application of these principles leads to equations which are too difficult to be solved, as already pointed out by Dirac in 1929[2]. This changed with the introduction of effective methods to approximately solve the many-body electron problem, like the Hartree–Fock (HF) formalism first and **Density Functional Theory** (DFT) later on.

While initially only properties of small molecules could be obtained in a reasonable amount of time, the explosion in computational power of the past few decades has allowed to consider systems with hundreds—if not thousands—of atoms and accurately characterize them with first-principles methodologies[3, 4]. Nowadays, computing basic properties of small molecules with decent accuracy has become a trivial task that any theoretical chemistry student can perform on their laptop in a matter of minutes. In spite of requiring study and experience to properly assert whether the result of a calculation is trustworthy or just plain garbage, it can be said that computations have now become an integral part of the mainstream chemistry research.

In spite of these advancements, many hurdles must still be overcome to model molecules and materials at a length- and timescale which is representative of the experiment. Realistic materials, for instance, are astonishingly complex and extremely difficult to describe from a theoretical standpoint. Luckily, during these four years of PhD, I was able to witness what (according to me) can be considered as the most significant revolution in the field of computational chemistry since the development of DFT, i.e. the extensive application of **machine learning methods** to speed-up and enhance theoretical predictions. Based on the data we were able to collect and that are being reported in literature, it appears that an increase in accessible time and length scales of the systems that are possible to study with quantum accuracy has to be expected. This revolution could be as significant as the one associated with the passage from wavefunction-based methods to DFT. While this is—of course—just speculation, I am very curious to see how the field is going to change in the upcoming 10–20 years.

Throughout this thesis, I will try to resume my attempt during the past four years of applying the aforementioned methods to the study of zeolite-catalyzed reactions, highlighting the challenges and opportunities that—I hope—allowed me to contribute to our current understanding of these fascinating materials.

1.1 Thesis goal and outline

The goal of this thesis is two-fold, as schematically illustrated in Figure 1.1. On the one hand, it aims to show how different methodologies are suited to model heterogeneous catalysts and derive various properties. This corresponds to go from left to right in Figure 1.1, where **enhanced sampling techniques** are used to overcome the limitations of **static methods** when investigating complex reactive systems. In particular, a large portion of the thesis aims to gain a deeper understanding on the behavior of water when co-adsorbed in the zeolite catalyst, acting either as spectator molecule or as reactant. On the other hand, the second goal of the thesis is to go beyond one of the major current bottlenecks in computational chemistry, namely the relatively large cost of DFT calculations. Ideally, one would like to have a method capable of simulating the potential energy surface (PES) of complex systems with a reduced computational cost, while maintaining quantum accuracy. Within this thesis, we showcase how **machine learning potentials** (MLPs) can be trained on a limited set of DFT calculations and then be used to cost-effectively compute a wide range of properties for the system under study. Accomplishing this second step is fundamental in order to go towards longer length- and timescales (from top to bottom in Figure 1.1), which will hopefully allow to model zeolite-catalyzed reactions in a more realistic fashion.

Overall, the thesis is organized in a more-or-less (chrono)logical fashion, where each chapter builds on top of the previous one and introduces a methodological advancement that allowed me to study more complex reactive systems in a more reliable way. Each chapter proposes moreover some examples of zeolite-catalyzed reactions I investigated with the respective methodology, highlighting the advantages and limitations that—at the end—allowed me to move forward in the research (Figure 1.1).

This thesis is organized as follows: in the **Introduction**, a general overview is given on what a catalyst is and why zeolites are actively used as catalysts in industry. We then examine a couple of interesting processes where zeolites play a pivotal role as catalysts, namely the (de)alkylation of aromatic compounds and the valorization of lignin-containing biomass. Both processes are extensively investigated in both academia and industry. These two examples were chosen as they are the broad family of processes to which most of the reactions I investigated belong, therefore recurring various times throughout the thesis. Finally, I will introduce a topic which I personally find extremely intriguing, namely the intricate nature of real zeolites. In spite of being often thought as perfectly crystalline materials, zeolites are very responsive to the reactive environment and the framework can undergo

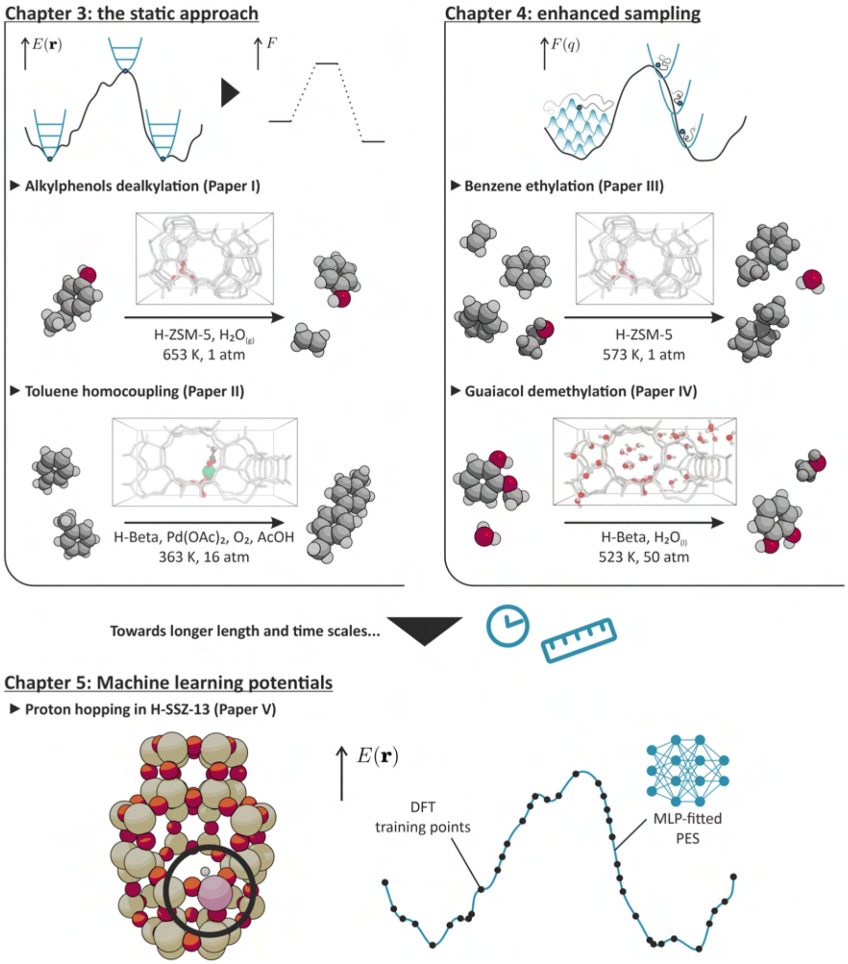


Figure 1.1: A schematic depiction of this thesis organization, where each chapter introduces a methodological advancement and presents one or more examples of zeolite-catalyzed reactions investigated with these methods. Chapter 3 introduces the static approach, where a reaction is investigated based on a limited set of interesting points on the PES. A better exploration of the PES, based on enhanced sampling techniques, is the main topic of Chapter 4. Finally, Chapter 5 showcases how a fast and accurate evaluation of the DFT PES can be achieved using machine learning potentials (MLPs).

extensive modifications. The impact of such modifications on reactivity is still poorly understood and, in my opinion, should be one of the main driving forces to move towards improved methodologies in the near future.

Chapter 2 lays the foundations of key methodological concepts that are relevant throughout the whole thesis. As previously mentioned, extensive details about the computational methods specifically used in the results-oriented chapters is provided at the beginning of the chapters themselves. Irrespective thereof, all the results still rely on the initial choice made to solve the many-body electron problem which, throughout this thesis, is always DFT. At the end of the chapter, the available strategies to model a zeolite-catalyzed reaction are briefly reviewed, from the atomistic model to the choice of the level of theory.

Chapters 3–5 are the core of this thesis. They all start by presenting a methodology that can be used to model zeolite-catalyzed reactions and that I have extensively applied in the past four years. Afterward, a summary of the main results I obtained is presented. An extensive presentation of these results can be found in the papers that, in their published or submitted form, are appended in the second part of the thesis. As outlined in Figure 1.1, **Chapter 3** covers the principle of static methodologies, which rely on optimized configurations of reactants, transition states and products. Sometimes these methods are also referred to as 'local' in literature. **Chapter 4** represents a leap forward towards realistic operating conditions, where molecular motions are taken into account using (enhanced sampling) molecular dynamics techniques, allowing for complex reaction environments. Finally, **Chapter 5** introduces the use of MLPs to replace the expensive DFT calculations and access longer length- and timescales. Differently from the previous chapters, the focus is not on the applications of the methodology but rather to the benefits that using an MLP can bring with respect to standard DFT methods within the field of zeolite catalysis.

To conclude, the final **Conclusions and Perspectives** chapter summarizes the main contributions of this thesis in the field of computational zeolite catalysis. Moreover, I will present what according to me are the main challenges that have to be tackled in the upcoming years in order to make computations truly valuable for experimentalists, by vastly improving their predictive power.

1.2 The catalyst concept

As previously explained, this thesis focuses on zeolite catalysis and, therefore, this paragraph aims to introduce the reader to the general concept of catalysis and heterogeneous catalysts. Broadly speaking, a chemical reaction comprises one or more rearrangements of valence electrons. The frontier molecular orbitals change in shape and location and—consequently—new chemical species are formed. Aside from sporadic exceptions, like radical combination reactions, this orbital rearrangement is energetically demanding as the electrons are perturbed from an equilibrium configuration. Most chemical reactions are thus associated with a significant barrier, that has to be overcome in order to obtain the products. 'Significant' means, in this case, not likely to be overcome through pure thermal energy. In numbers this corresponds to at least 3–4 times the thermal energy $k_\mathrm{B}T$, with k_B Boltzmann's constant. In layman's terms, one can imagine a chemical reaction as the crossing of a mountain pass connecting two valleys. The exploring hiker is, however, extremely lazy and moves randomly—preferentially towards lower altitudes. It is then easily to imagine how getting the hiker to cross the pass is a very unlikely event.

The unlikely occurrence of chemical reactions can be either a blessing or a curse. For instance, we luckily do not spontaneously combust in atmospheric oxygen because of the very high activation barrier. A handful of energy (for instance a spark or an electric arc) is necessary to trigger the reaction. On the other hand, there are plenty of reactions that would be extremely useful, but are unfortunately also associated with very high barriers. A textbook example is the conversion of nitrogen and hydrogen to ammonia for fertilizers production, which is ideal to introduce the concept of catalyst. Ammonia synthesis represents one of the greatest scientific achievements in the last century and led to a Nobel prize award twice, to Fritz Haber in 1918[5] Carl Bosch in 1931[6].

Ammonia synthesis through nitrogen fixation can be described by the following chemical equation:

$$\mathrm{N_2 + 3\,H_2 \longrightarrow 2\,NH_3}$$

If hydrogen and nitrogen are mixed at normal conditions (20 °C, 1 atm) nothing happens because, as previously stated, crossing the reaction barrier is extremely unlikely despite ammonia being thermodynamically more stable than the reactants. To make the reaction possible, an intermediate is needed. Such intermediate—the catalyst—is a third substance which actively interacts with the reagents. It can pre-activate them and/or facilitate their interaction, lowering the reaction barrier while being possible to recover unchanged at the end of the reaction. In the Haber-Bosch process, such catalyst is a metallic surface, traditionally iron. N_2 and H_2 can be chemisorbed on the metal surface as N and H atoms under high temperatures and pressures and can then easily recombine and desorb as ammonia.

The idea of a material capable of enhancing reaction rates is not restricted to the ammonia synthesis, but can be easily generalized to any third-party substance that actively takes part in a reaction. The catalyst should reduce the reaction activation energy while remaining formally unaltered at its end (‘unaltered’ is quite an oversimplification, see more details in Section 1.4). In the following section, we will see how the class of nanoporous materials known as zeolites can also take the role of an active catalyst and facilitate the occurrence of a broad range of chemical reactions.

In general, catalysts are divided in three major classes, namely homogeneous, heterogeneous and enzymatic. An homogeneous catalyst exists in the same phase as the reactants. Typical examples are small organic molecules dissolved in the same solvent as the reactants, solvated hydronium ions for Brønsted-acid catalysis and transition metal complexes. The 2021 Nobel Prize in Chemistry was awarded to List and MacMillan "for the development of asymmetric organocatalysis"[7], in which chiral proline-derived molecules are used to catalyze organic reactions with a high enantiomeric excess. In heterogeneous catalysis, the catalyst and the reactants are not present in the same phase. Zeolites are a typical example of heterogeneous catalyst. Indeed, they are microporous solids (with pores diameters < 2 nm) and are commonly used for reactions conducted with gas-phase or solvated reactants. Finally, enzymes are proteins which, thanks to their intricate 3-dimensional structure and—in some cases—to additional co-factors, catalyze with outstanding selectivity most of the reactive processes in living organisms.

Since our focus is on heterogeneous catalysis, some additional details can be provided concerning the quantities of interest that can be computed to characterize an heterogeneously catalyzed reaction. We assume that the reader is familiar with the general thermodynamic concepts of enthalpy, entropy, free energy, etc.[8] When one or more reactive gas-phase species interact with a solid micropourous material, they are initially physisorbed in its pores. This process is generally exothermic, but associated with a negative entropy variation. We will refer to the free energy of adsorption as $\Delta F_{\mathrm{ads,R}}$ (Figure 1.2). Afterwards, the system has to cross the reaction barrier $\Delta F^{\ddagger}_{\mathrm{cat}}$. While in Figure 1.2 a single barrier is shown for the sake of simplicity, the conversion from reactants to products can involve multiple reaction steps and/or reaction pathways. Moreover, active sites in microporous solids like zeolites can be separated by pores that are of molecular dimensions, often making diffusion barriers as important as the reactive ones. It goes without saying that taking this complexity into account is very challenging and it is one of the main tasks of methods such as microkinetic modeling or kinetic monte carlo analysis, which are however only marginally touched on in this thesis. The energy variation of the adsorption step ($\Delta F_{\mathrm{ads,R}}$) adds to $\Delta F^{\ddagger}_{\mathrm{cat}}$. This is particularly relevant when comparing computational results with the experiment, as the activation energies derived from basic kinetic analysis like Arrhenius plots are always apparent, meaning that they include the adsorption energy (unless the reaction is zero order in the reactants concentration, in that case the true activation energy should be retrieved[9]). We will refer to the overall activation energy, including the adsorption step, as 'apparent' ($\Delta F^{\ddagger}_{\mathrm{cat,app}}$, see Figure 1.2). Finally, the reaction products must desorb from the catalyst, with an associated free energy of desorption $\Delta F_{\mathrm{des,P}}$. Note that the difference between adsorption and desorption energies is purely practical, as they are exactly the same process seen in two opposite directions. Throughout this thesis, we will use the convention:

$$\Delta F_{\mathrm{ads,A}} = F_{\mathrm{A+B}} - F_{\mathrm{A,gas}} - F_{\mathrm{B}} = -\Delta F_{\mathrm{des,A}}, \tag{1.1}$$

in which A is the molecule being adsorbed/desorbed and B the solid catalyst.

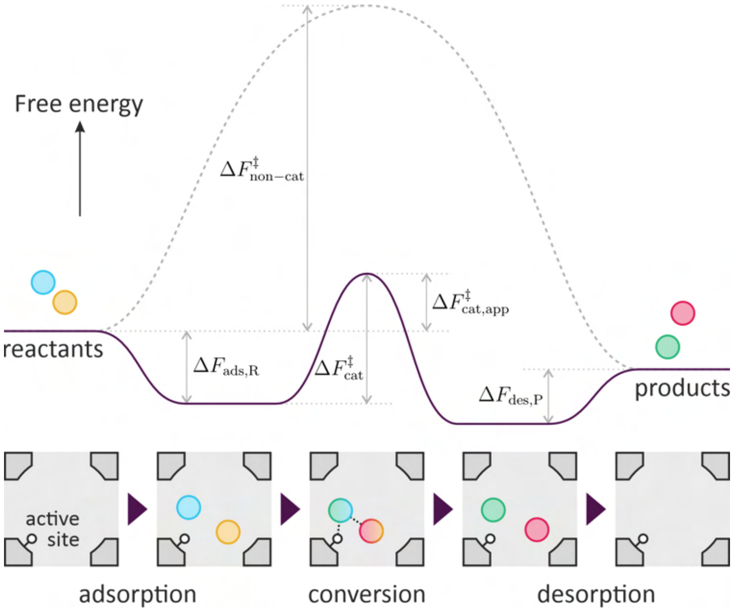


Figure 1.2: An heterogeneous catalyst can promote a chemical reaction by adsorbing the reacting species and allowing for low-energy transition states thanks to spatial confinement and the presence of active sites.

1.3 Zeolites in research and industry

Having outlined the main characteristic of an heterogeneous catalyst, we can describe more in detail the physico-chemical properties of zeolites and highlight some of their main applications as catalysts for industrial processes. A zeotype material is a crystalline, microporous network of interconnected corner-sharing XO_4 tetrahedra, where X is in general a 3+–5+ cation.[10, 11] In this thesis, we focus on Brønsted-acidic zeolites. They are aluminosilicates in which X is mostly Si^{4+} with occasional Al^{3+} substitutions. Obviously, the Al substitutions create a charge imbalance in the system, that must be compensated. To this end, it is very common for a proton to be the charge-compensating cation, which lies on one of the oxygens in the first coordination sphere of the Al defect (Figure 1.3*a*). This proton is rather acidic and it is commonly referred to as Brønsted acid site (BAS). The BAS is the traditional active site for Brønsted acid-catalyzed reactions in zeolites, which are among the most common industrial processes in which zeolites are involved.

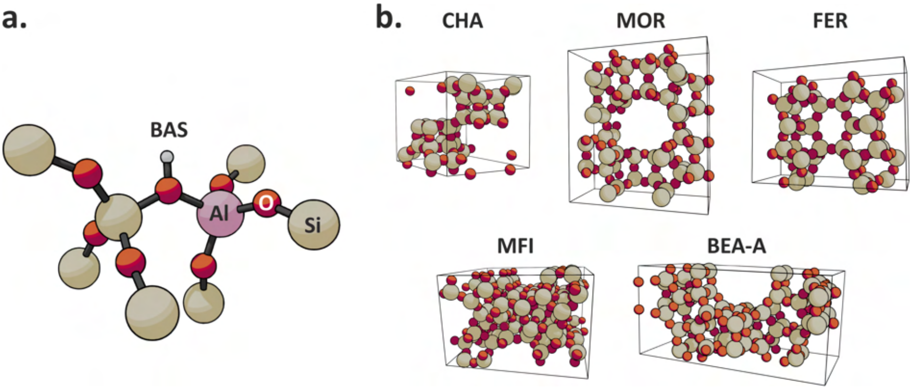


Figure 1.3: *a.* Detail of the zeolite BAS up to four bonds from it (extracted from the MFI framework with the Al on the T12 location). *b.* Crystallographic unit cells of some zeolites commonly used in research and industry. The MFI, BEA and CHA topologies are also extensively used throughout this thesis.

Interestingly, there exists a plethora of possible zeolite frameworks. This is due to the high flexibility of the Si–O–Si bond, which can be adapted to many different topologies. Currently, there are 240 synthesized frameworks listed in the IZA (international zeolite association) database[12], but over 2 millions could theoretically exist[13]. Some relevant examples of commonly used topologies in research and industry are shown in Figure 1.3*b*. The size and shape of the pores and channels in a zeolite is an essential aspect of its catalytic activity. Indeed, proton transfer reactions that trigger the acid-catalyzed process are enhanced by the pores of molecular dimensions, which tend to remarkably stabilize charged species[14]. Moreover, such pores can give rise to three different types of reaction selectivity (Figure 1.4), i.e. they favor the formation of a specific product when many could be obtained from different reaction pathways[15]:

1. **Reactants shape selectivity:** if one reactant is hindered enough, its entrance in the zeolite pore system and/or its diffusion therein will be

severely hampered, making it impossible to reach the active site and further react.

2. **Transition state shape selectivity:** only transition states that can be easily accommodated in the pockets of the zeolite channel network are strongly stabilized by the interactions with the pores walls and active sites. This actually makes zeolites—in a sense—similar to enzymes, where the reactants and active site pockets are perfectly suited for each other in a key-lock mechanism. Of course, while enzymes have evolved their active pockets to perfectly accommodate the reacting substrate, it is the chemist's task to find a zeolite suited to host the desired transition states.

3. **Products shape selectivity:** a reaction network with many possible outcomes can lead to products that are way too hindered to diffuse away from the pockets in the zeolite channels. This means that they will either revert back to the reactants and undergo another reaction path or they will lead to catalyst deactivation (see Section 1.4.2). In any case, the hindered product will not be observed in the products stream.

With this basic principles in mind, we can proceed and discuss in greater detail a couple of relevant industrial applications in which zeolites are widely employed.

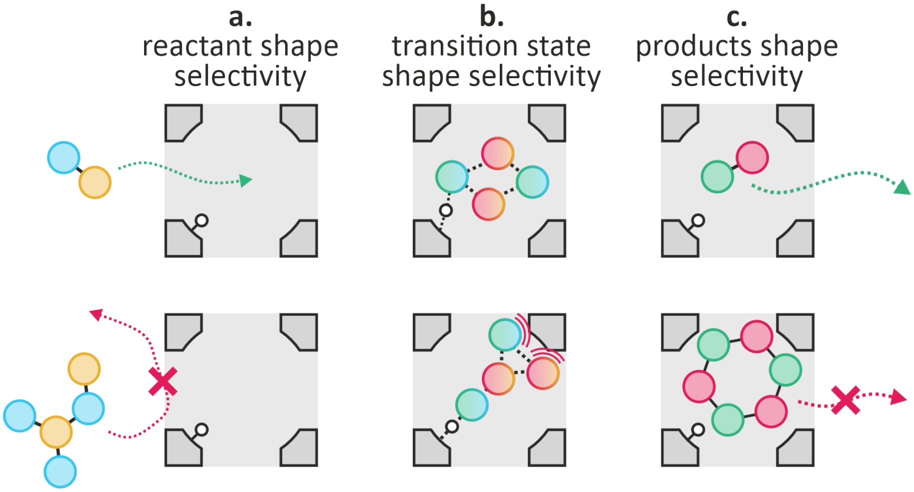


Figure 1.4: A schematic depiction of the shape selectivity concept, when concerning the reaction reactants (*a*), transition state (*b*) or products (*c*).

1.3.1 (De)alkylation or aromatics

A common reaction for which zeolites can be employed is the (de)alkylation of aromatics[16]. As one might remember from any basic organic chemistry course, alkylation of aromatics proceeds through an **electrophilic substitution** (S_EAr) mechanism. In this textbook mechanism, the electrophile (activated by a Brønsted or Lewis acid) attacks the aromatic ring forming a positively-charged arenium ion intermediate, commonly known as Wheland complex. The latter, stabilized by resonance, is then deprotonated by a Brønsted base present in the environment, giving the final neutral product. In some cases, it has been proposed that the Wheland complex might not be a reaction intermediate at all and the S_EAr proceeds in a concerted manner[17].

In the context of zeolite catalysis, the alkylation reaction is most often performed using alkenes or alcohols[16]. The most common industrial example is represented by the ethylation of benzene with ethene in the Brønsted acidic H-ZSM-5 zeolite (MFI topology, see Figure 1.3*b*), which since the patent deposit in 1976 is known as Mobil-Badger process[18, 19]. Ethylbenzene is a very important platform chemical for the polymer industry, as its dehydrogenation leads to styrene for polystyrene production. Ethanol can also be used to replace ethene as ethylating agent, with the advantage that it can be produced from renewable sources. Currently, this alternative has a reduced industrial relevance because of the low cost of oil-derived hydrocarbons[20].

In recent years, the opposite process with respect to alkylation—i.e. dealkylation—has also gained popularity in the context of biomass conversion (see further details in the following section). Indeed, alkylated aromatics can be obtained from biomass and, in the path towards drop-in chemicals, defunctionalization through zeolite catalysis is a promising strategy[21–25]. As final S_EAr example, the continuous methylation of aromatics by methanol can be cited. This reaction is at the base of the hydrocarbon pool mechanism on which the methanol-to-hydrocarbons (MTH) processes are based[26] (see Figure 1.9).

The purpose of this paragraph is to discuss in details the general mechanistic features of aromatic alkylation with alkenes and alcohols in Brønsted acidic zeolites. This discussion is important as we are going to re-encounter this mechanism twice throughout the thesis: in **Paper I**, dealing with the dealkylation of alkylphenols, and in **Paper III**, in which the ethylation of benzene is studied. The presented mechanistic details are well established in literature and supported by a plethora of experimental[27] and theoretical[28–34] studies. We will assume the S_EAr reaction to proceed in the 'usual' direction, namely as an alkylation process, but everything said applies to dealkylation as well, simply reverted. For convenience, benzene and ethene or benzene and ethanol will be chosen as model reactants to discuss the mechanism.

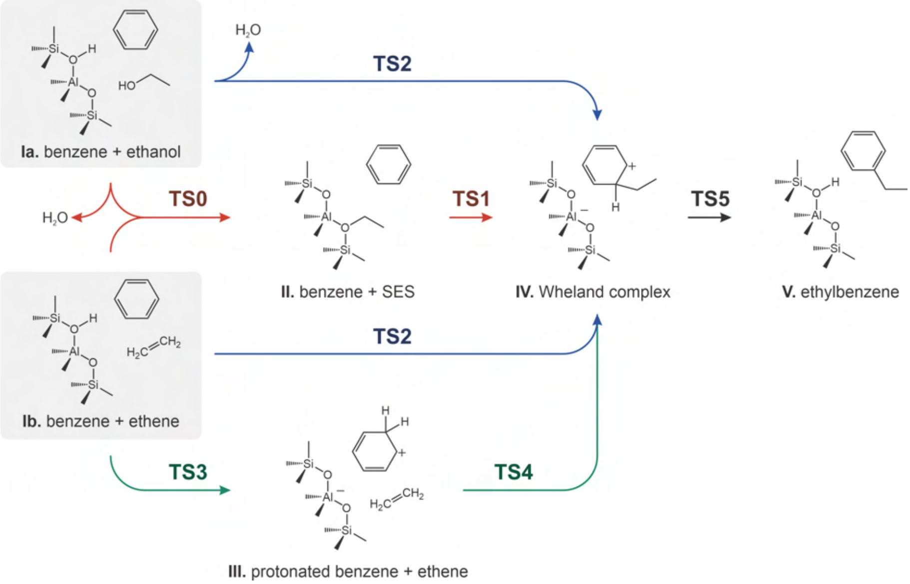


Figure 1.5: A general overview of the possible mechanistic pathways through which the alkylation of benzene with ethanol (**Ia**) or with ethene (**Ib**) can proceed.

A full overview of the possible reaction paths is shown in Figure 1.5. First, the two reactants have to be adsorbed on the zeolite BAS, which is required to activate the electrophile. In the case of ethanol (**Ia**), this activation is produced through a partial protonation of the hydroxyl group. The incipient partially positive water molecule is a good leaving group, making the alkyl carbon susceptible to undergo nucleophilic attack from an electron-rich species. For ethene (**Ib**), the interaction between the π bond electronic cloud and the BAS is sufficient to strongly polarize the former, making one of the carbon atoms quite electrophilic. When it attacks the aromatic ring, the BAS is completely transferred to the carbon, becoming part of the product.

The two main mechanisms reported in literature are the step-wise one, which proceeds through TS0 and TS1, and the concerted one, via TS2. In the former, the activated electrophile reacts with one of the zeolite framework oxygens in the first coordination sphere of the Al defect, forming a surface ethoxide species (SES, **II**). The SES bond remains quite polar and the carbon is still a good electrophile, thereby being capable of attacking the aromatic ring with formation of the Wheland complex **IV**. Alternatively, the activated electrophile can directly attack the aromatic ring, immediately forming **IV** through TS2.

A final, less investigated mechanism has been proposed by Kolboe et al.[35] It starts with a protonation of the benzene ring by the BAS (TS3). The protonated benzene (**III**) acts as the electrophile in the reaction, attacking the ethene molecule in a sort of intramolecular TS2 (TS4) which leads again to **IV**. In all the considered mechanisms, the final step is the deprotonation of the Wheland intermediate **IV** to form the final neutral ethylbenzene product **V**.

As mentioned, the concerted mechanism (TS2) and the step-wise one (TS0, TS1) are the only ones commonly considered in the literature. When investigating the ethylation of benzene with ethanol, Chowdhury and co-workers[27] were able to detect the formation of Wheland-type intermediates (such as **III** and **IV** in Figure 1.5) with operando UV-vis spectroscopy. Moreover, they used nuclear magnetic resonance (NMR) spectroscopy to further investigate the reaction intermediates, and found that the SES formation also seems to be of importance at realistic reaction conditions.

Most of the theoretical studies thus far have focused on the ethylation of benzene with ethanol or ethene using solely static methodologies. Hence there are two interesting questions that remain unclear: first, how is the reaction affected by different substrates? **Paper I** tries to answer this question (Chapter 3), investigating the dealkylation reaction of lignin-derived alkylphenols. Second, how are the mechanistic details derived from static simulations affected by realistic reaction conditions? Such question is tackled by **Paper III** (Chapter 4), where enhanced sampling molecular dynamics techniques are used to obtain mechanistic insights in the ethylation of benzene with ethene and ethanol at operando conditions.

1.3.2 Biomass conversion

After having been used for decades as workhorses in the petrochemical industry, zeolites are far from becoming obsolete catalysts. In recent years, indeed, the growing concern towards the planet's climate and environment has shifted the interest of the mainstream chemistry research towards sustainable industrial processes and the usage of non-fossil based feedstocks. In this context, zeolites have been shown to be very effective in catalyzing reactions for the conversion of biomass[21]. While preserving some similarities with oil-derived hydrocarbons, biomass has normally a high content of oxygenated functionalities and water, leading to reactive conditions that can drastically change the function of the zeolite catalyst. In this paragraph, we will discuss in general terms the processes used to convert lignocellulosic biomass into platform chemicals, as this is the application area in which both **Paper I** and **Paper IV** are situated.

Lignocellulosic biomass, such as wood, has two main polymeric components, namely a (hemi)cellulosic fraction and lignin. These two types of macromolecules have extremely different composition. While (hemi)cellulose is made of long carbohydrate chains, lignin is mainly composed of oxygenated aromatic units, tightly connected in a reticulated network. In the paper production process, lignin is normally disintegrated with aggressive chemicals to purify the cellulosic fraction. The remaining slurry can then be used as low-value fuel. This is however a remarkable waste of potentially useful chemicals, as lignin contains the highest percentage by weight of aromatic units. Nowadays, aromatic compounds are greatly used in the production of plastics but are almost exclusively derived from oil fractionation. It appears therefore natural to imagine that, if lignin could be efficiently depolymerized while leaving the cellulosic fraction (almost) unaltered for further usage, the chemical industry could have an infinite supply of bio-aromatics from a renewable source. We will not dive too deep in the outstandingly vast subject of lignin valorization as a detailed discussion would require its own chapter. We address therefore the interested reader to specialized reviews on the topic[36–39].

An efficient depolymerization of lignin is, in principle, the core goal of the **lignin-first approach**, extensively studied and developed during the past decade[40, 41]. With improved catalytic processes, lignin can now be depolymerized with high efficiency in a mixture containing mostly alkyl-guaiacols and alkyl-pyrogallols (A_1–A_2 in Figure 1.6)[42, 43]. Despite being an amazing achievement, these compounds are still too complex to be considered drop-in molecules for the chemical industry and a further *C*- and *O*-dealkylation is desirable[41].

As first option, the methoxy functionalities can be fully removed with a Ni-catalyzed hydrogenation, which converts the intermediates of Figure 1.6 in simpler alkylphenols. These alkylphenols can be *C*-dealkylated towards phenol and olefins in the H–ZSM–5 zeolite at high temperature and in the presence of steam. The whole process has been shown to be an economically valuable route in the future biorefinery[41]. The mechanism of alkylphenols dealkylation has some interesting substrate-dependent features, which are going to be the main focus of Section 3.2.1 (summarizing the results of **Paper I**). Unfortunately, the process is not suited to dealkylate alkyl chains carrying an hydroxyl group (A_2, B_2 in Figure 1.6). Indeed, the double bonds produced from dehydration reactions quickly polymerize deactivating the catalyst.

Another option consists in using homogeneous Brønsted acids (HCl in this case) in hot pressurized water (HPW) to perform the demethoxylation reaction[44–47]. HPW is an interesting green solvent as, while being in subcritical conditions, the reduced dielectric constant of water at high temperatures

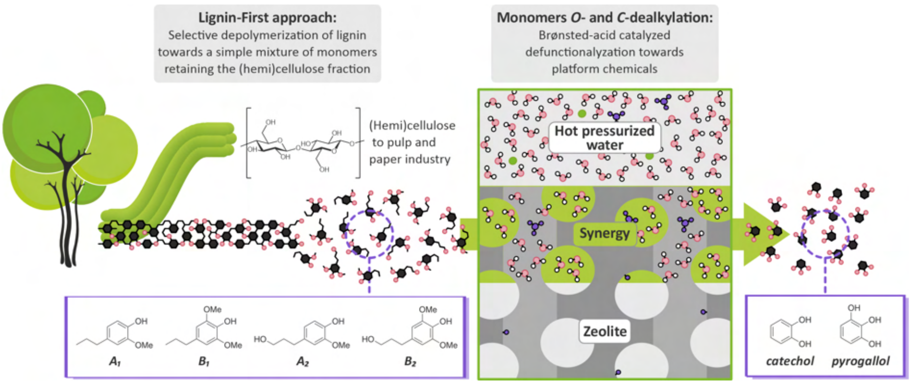


Figure 1.6: With the recently developed lignin-first approach, the lignin fraction of wood can be efficiently depolymerized in a mixture of mainly four monomers (A_1, A_2, B_1 and B_2). Subsequently, they can be *C*- and *O*-dealkylated towards catechol and pyrogallol with a Brønsted acid catalyst. *C*-dealkylation can be conducted in gas-phase at high temperatures with a zeolite catalyst. *O*-dealkylation can be performed in hot pressurized water using either an homogeneous Brønsted acid or a zeolite. The two processes are the main topic of **Paper I** and **Paper IV**, respectively.

makes it somehow similar to an organic solvent. Moreover, it has been shown that the homogeneous catalyst can be replaced with a Brønsted acidic zeolite, with all the advantages associated with heterogeneous catalysts (easy separation, reduced corrosion, etc.)[48]. Previously, it has been recognized that hydronium ions can present an enhanced catalytic activity when constrained in pores of molecular dimensions, such as the zeolite ones[49, 50]. While this effect has been deeply studied in terms of ionic environment for reactions involving charged intermediates[51], it remained unclear whether the rate enhancement persists also for reactions possibly not involving carbocations. In **Paper IV**, we explore into details the mechanism of guaiacol *O*-demethylation in HPW using an homo- (HCl) and heterogeneous (H–Beta zeolite with BEA topology, see Figure 1.3*b*) catalyst, highlighting the different behavior and reactivity of hydronium ions when going from bulk water to the constrained environment. The results are summarized in Section 4.3.2.

1.4 Zeolites are more intricate than they appear

In Section 1.2, we stated that a catalyst is a third-party substance that interacts with the reagents, increasing the reaction rate, and that it can be recovered unaltered at the end of the process. This is however not always completely correct, as the reactants and reaction conditions themselves can have a profound impact on the catalyst structure and stability. With the reaction progressing, there can be a dynamic evolution in the type of active sites that were present in the original material and each type of new active site can have a different concentration and a different activity[52]. Moreover, undesired side-processes can lead to catalyst poisoning and loss of active sites, in what is commonly referred to as deactivation. In this paragraph, I will try to summarize what in my opinion is a fascinating aspect of zeolite catalysis, i.e. their hidden complexity. It is indeed now well understood that BASs are by far not the only type of active site present in the crystalline material and that, in spite of their remarkable thermal stability, zeolite frameworks can undergo extensive chemical modifications at the nanoscale, where the framework bonds are at least partially disrupted. This makes them intricate materials and much more challenging to accurately model with respect to what is commonly thought. Modeling this kind of complexity is going to be—in my opinion—one of the main challenges for computational chemists working in the field during the upcoming years, as nicely pointed out in a recent perspective article by Chizallet[53].

1.4.1 Framework lability and unconventional active sites

As previously discussed, the prototypical active site of aluminosilicate zeolites is the BAS (Figure 1.3*a*). In most computational studies of zeolite-catalyzed reactions, a pristine BAS is indeed the only type of active site that is taken into account (see Section 2.3). It is becoming clear, however, that the picture is much more complicated. Our purpose here is not to provide a full review on the dynamic nature and complexity of realistic zeolite active sites, as this topic has already been the subject of excellent review papers[54] and we also contributed a book chapter on the subject[55]. The main goal is to introduce the reader to a more realistic depiction of zeolite active sites, which is going to be relevant when exposing the results of dynamic simulations in Chapter 4 and in the conclusion, when the current challenges in zeolite modeling will be outlined.

One of the main culprits for the formation of unconventional active sites is water, whose interaction with Brønsted acidic zeolites has become a central topic in recent literature[56–58]. We can divide the interactions between water and Brønsted acidic zeolites in three main stages, based on the timescale on which they occur (Figure 1.7). First, one or more water molecules can come into contact with a pristine BAS. The interaction between the two is extremely strong. With one adsorbed molecule, the proton shuttles almost freely between the framework and the water (Figure 1.7*a*). If one additional water molecule is present, the proton can be completely abstracted from the framework for a significant amount of time, leading to the formation of (partially) solvated hydronium ions (H_3O^+). When the water loading becomes higher than 2 molecules per BAS, the fraction of time during which the proton resides on the zeolite framework becomes fundamentally insignificant. The easy formation of hydronium ions in zeolites exposed to water is nowadays well accepted and documented by experimental[59] and theoretical[60, 61] studies (remark that the listed references are by far non-exhaustive, we refer the interested reader to the many recent review papers on the topic[56–58] for a complete overview).

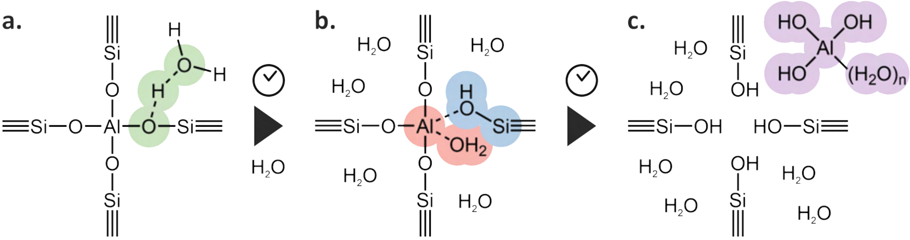


Figure 1.7: When water interacts with a BAS, there can be an initial fast proton exchange between the zeolite and the water molecules (*a*). At high water concentrations, the framework bonds become labile and can undergo partial and reversible hydrolysis (*b*). With longer exposure times and/or higher temperatures, water can lead to an irreversible breaking of the framework bonds, leading to the formation of EFAL species (*c*).

Water can have a stronger impact on the framework at high temperatures and/or high concentrations. Indeed, it is now established that the framework bonds can undergo reversible hydrolysis when exposed to water. This has been shown based on theoretical simulations[62] as well as experiments. For instance, Pugh and co-workers showed that zeolite frameworks can be enriched with ^{17}O by simple exposure to $H_2{}^{17}O$ at room temperature[63]. This is caused by the low activation energy for the nucleophilic attack of water on the tetrahedral framework atoms (Figure 1.7*b*). While Al atoms undergo easily partial hydrolysis, it is accepted that the process, while slower, occurs for the silicon atoms as well[62, 63].

On a longer timescale, water can have an even deeper effect on the zeolite structure and lead to a complete hydrolysis of the tetrahedral atoms. In the case of Al, this leads to the formation of extra-framework Al (EFAL) species, which are (partially) hydrated aluminum hydroxide complexes (Figure 1.7*c*)[64–72]. EFAL formation is not a new concept in zeolite catalysis. Indeed, high-temperature water vapor is commonly used in the steaming pretreatment procedure[56]. There, the catalyst is exposed to steam in order to (partially) dealuminate the material. This is done to tweak its catalytic properties. At high temperatures, indeed, EFALs can loose some of the coordinating water molecules and become Lewis acid sites (LASs), allowing for additional reaction paths that a pristine BAS could not catalyze[73]. Under severe steaming conditions, the framework can start to loose silicon atoms as well, with the consequent formation of mesopores. These also have an important role in modulating the catalytic properties of the material, mainly for their role in facilitating diffusion and increasing the fraction of surface-catalyzed reactions. As an example, ref. 25 extensively investigates the effect of mesoporosity and EFAL concentration on the dealkylation of alkylphenols, which we introduced in Section 1.3. Since the study of mesopores with first-principles modeling is still at its infancy and of scarce interest for the main topics treated in this thesis, it will not be further discussed.

In spite of their ubiquity and importance, EFAL remains relatively poorly understood. A plethora of structures with varying coordination, nuclearity and solvation levels have been proposed based on spectroscopic and/or theoretical evidences[74]. For instance, Pidko and co-workers used thermodynamic modeling based on ab initio calculations to determine the most representative EFAL species in the FAU topology and determined that a polynuclear cluster composed of 3 Al atoms was the preferred species in this large-cage zeolite[75]. Understanding the type and amount of EFAL species in a working zeolite catalyst is essential to obtain a clear reactivity picture. Indeed, it is now accepted that EFAL species can migrate and interact with pristine BASs present in the material, creating new BAS/LAS synergies with a potentially higher catalytic activity than the two isolated components[76–78]. The polynuclear clusters identified by Pidko, for instance, were also predicted to enhance the catalytic activity of neighboring BASs[79, 80] and NMR experiments have shown that highly-deshielded ^{1}H NMR signals are potentially caused by an EFAL in interaction with a BAS.[81, 82]

Thus, in conclusion, it appears clear that modeling a zeolite as a perfect crystal with one, well-defined active BAS is not necessarily representative for the true complexity of the material[53]. A working zeolite used, for instance, in the MTH process might look more like the one depicted in Figure 1.8. Aside from pristine BASs, the non uniform water distribution in the material[83] could lead to some protons to be solvated as hydronium ions. The consequences of this solvation on reactivity are extensively discussed in Section 4.3.1 and 4.3.2. Many of the aluminum sites might present various degrees of abstraction from the framework, going from framework-associated sites to fully free EFAL species[74]. The latter, after migration, can interact with each other and with the BAS creating synergistic effects whose understanding is still at its infancy. Extended dealumination and desilication might lead to the formation of mesopores covered in silanol groups and surface BASs, with catalytic properties can be quite different from the active sites in the constrained micropores[84, 85]. Finally, extended carbonaceous deposits can either block the micropores or poison the catalyst surface, leading to deactivation (see next section). It is clear that a proper understanding of this complexity is going to be a true challenge for molecular modeling in the upcoming years, as the size and number of systems we can investigate is still severely hampered by the computational cost of ab initio simulations (Section 2.3). In Chapter 5 and in the conclusions, we will discuss how machine learning methods could be the key to go beyond this limitation and towards a more realistic modeling of complex catalytic systems.

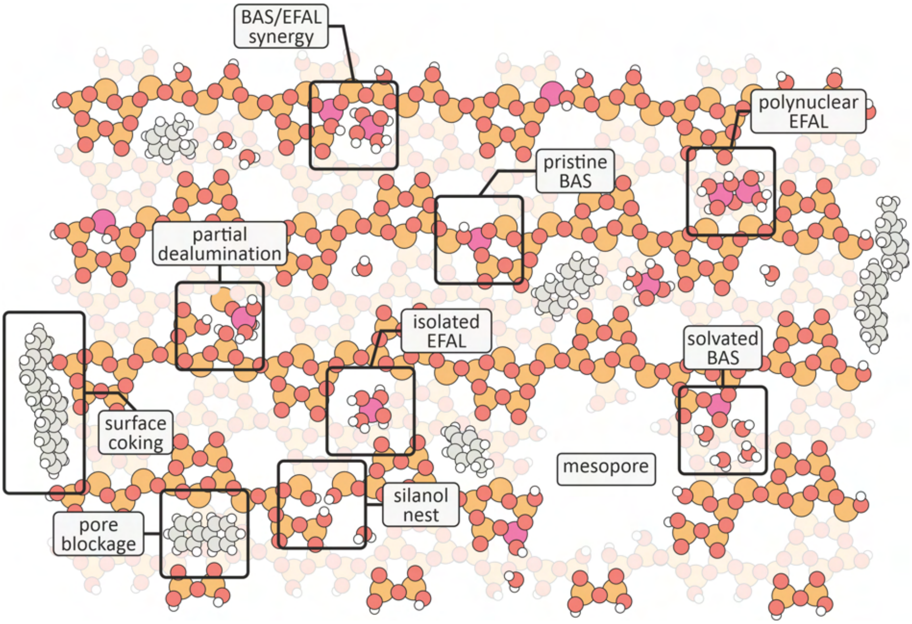


Figure 1.8: Schematic, non-exhaustive depiction of how a H-ZSM-5 crystal (here cut along the sinusoidal channels) might appear after catalyzing the MTO reaction.

1.4.2 Catalyst deactivation

The term "deactivation" commonly refers to the loss of activity and/or selectivity of a catalyst. This definition is very broad and can refer to a plethora of mechanisms and causes that vary a lot depending on the catalyst type (homo-, hetero- and bio-) and on the reaction driving force (thermal-, photo- and electro-)[86]. Hereafter, we will focus on the deactivation of zeolites in the methanol-to-olefins process (MTO), as this is a very active field of study in which I had the pleasure to be involved[87]. MTO is a widely studied zeolite-catalyzed process in the scientific literature and molecular modeling[26]. Since methanol can be obtained from renewable feedstocks, converting it to highly demanded hydrocarbons like ethene and propene would be a desirable alternative to the currently used fossil sources[88].

Without going into excessive detail, after an induction period where the initial C-C bonds are formed the hydrocarbon pool (HCP) mechanism takes over. This is a very fascinating and unique mechanism, where large hydrocarbon molecules that cannot diffuse away from the zeolite pores are subjected to continuous methylation reactions, releasing olefins through craking and dealkylation reactions. Such hydrocarbons are either aromatic compounds or large olefins and, depending on this, two catalytic cycles are possible (dual cycle concept), as seen in Figure 1.9. While a lot of effort has been done to elucidate the mechanistic features of the dual HCP cycle and of the initial C-C bond formation (among the most recent investigations some of the works by Studt and co-workers[89, 90] can be mentioned), the key features of the deactivation process are less explored. Deactivation in the MTO process mainly occurs via coking, where large aromatic molecules can either cover the catalyst surface, preventing the reactants to enter, or get stuck in the zeolite pores, jamming the channels and hindering the reaction. Following the deactivation process on a molecular scale is challenging for experiments. Atom probe tomography has for instance been used to obtain sub-nm information on coke distribution in a large spent H−ZSM−5 crystal[91]. It was found that a strong correlation exists between coking and Al distribution, where regions of the crystal with low Si/Al ratio were also the more susceptible to coke formation. This is in line with previous evidences from UV-vis and confocal fluorescence microscopy[92, 93].

Tracing the exact molecular species that are present in the zeolite catalyst under operando conditions is even more challenging and requires a technique such as Raman spectroscopy. Indeed, the adsorption peaks in a Raman spectrum provide excellent fingerprints to distinguish between families of molecules. The problem is that the working zeolite emits a very strong fluorescence signal when irradiated with the laser pulse. Because of its high intensity, the fluorescence overwhelms the detector and makes the Raman signal undetectable. For this reasons, our collaborators at the University College in London (Prof. A. Beale) used for the first time Kerr-gated Raman spectroscopy to collect spectroscopic evidences on a working H−SSZ−13 (CHA topology, see Figure 1.3*b*) catalyst during the MTO reaction. Kerr-gated Raman spectroscopy (Figure 1.10*a*) exploits the time delay between the Raman signal, which is basically instantaneous with respect to the excitation pulse, and the fluorescence, which is slightly delayed. The Kerr medium (CS_2) can be used to selectively change the polarization of the Raman signal with a well-timed pulse of light and a polarizer is then used to get rid of the fluorescence signal.

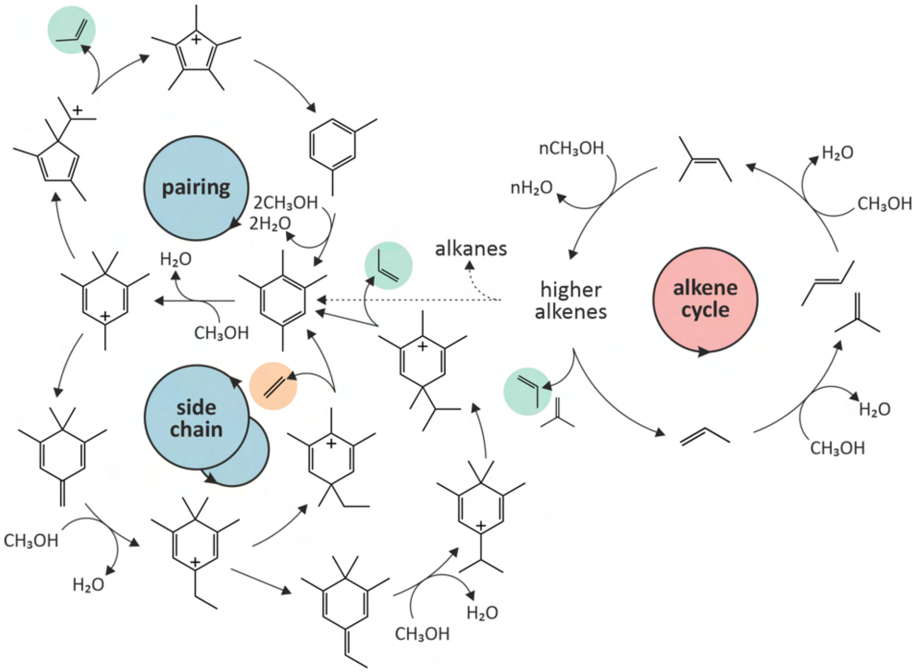


Figure 1.9: The dual cycle HCP mechanism to convert methanol into ethene and propene during the steady-state regime of the MTO process. Adapted from ref. 26, with permission from the Royal Society of Chemistry.

While Kerr-gated Raman allows to obtain clean spectra, attributing the peaks to specific molecular species remains a challenging task. For this reason, Alexander Hoffman from our group used molecular dynamics simulations (see more technical details in Chapter 4) to compute the Raman spectra of important intermediates at various MTO stages, going from the induction period to the fully deactivated catalyst. The results are summarized in Figure 1.10*b*, where it can be seen how the theoretical predictions allow to follow the gradual formation of HCP-active aromatics and polyolefins from the initial C_3–C_4 short chain alkenes (260–280 °C). With further polymerization, large and no longer mobile branched polyolefins are formed. Interestingly, Alexander noticed that the highly branched polyolefins present at the onset of deactivation, after protonation, underwent spontaneous cyclization.

We then decided to perform additional simulations to better understand the role of branched olefins in the formation of cyclic hydrocarbons. First, we noted that the branching does not play a special role in facilitating the cyclization step, as MD simulations performed on protonated 1,2,3-heptatriene also resulted in spontaneous ring closure within 50 ps (Figure 1.11). We then focused our attention on the fact that, in order to cycle, the molecule has to

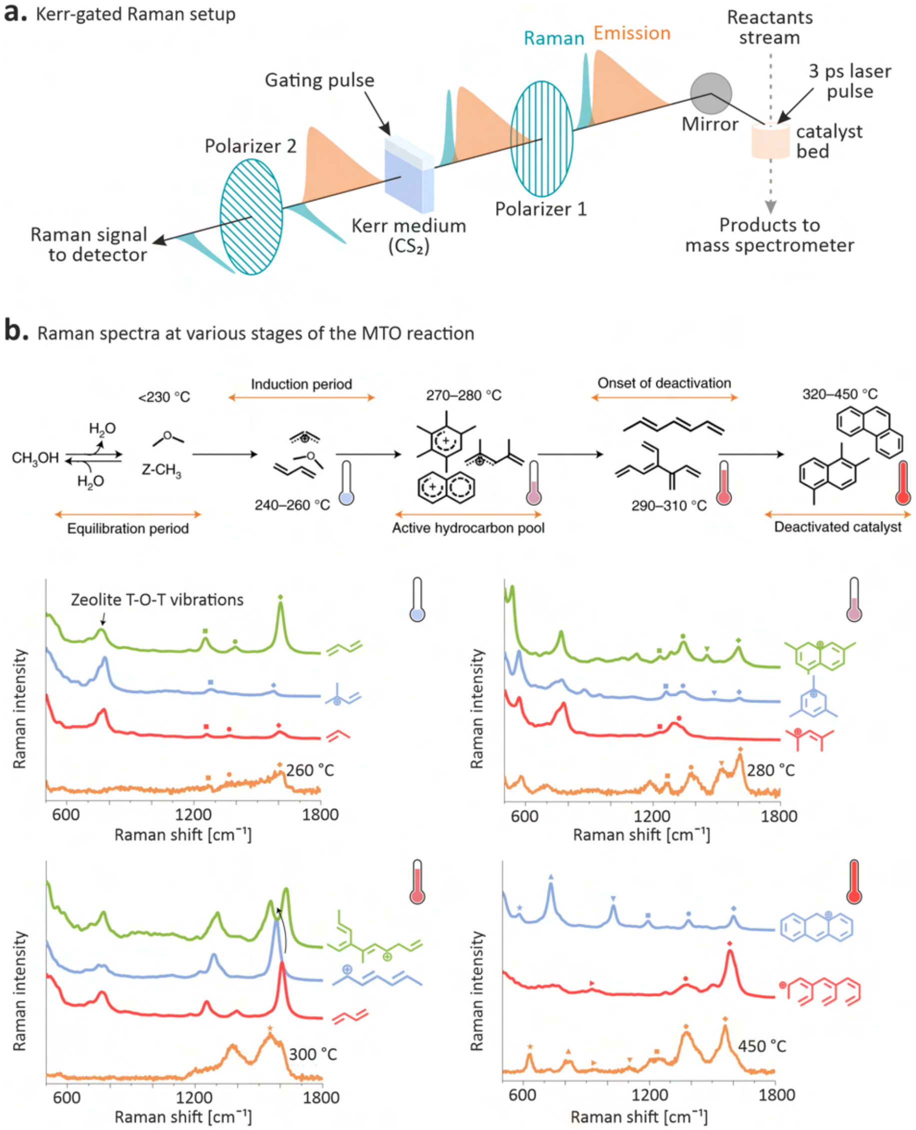


Figure 1.10: *a.* Schematic depiction of a Kerr-gated Raman setup for the *operando* study of the MTO process. *b.* Raman spectra of the MTO reaction in the H–SSZ–13 zeolite collected at various reaction temperatures and their comparison with the computed ones for specific intermediates. Adapted from ref. 87, with permission from Springer Nature.

curl, so that carbon 1 and 5 are close to each other. Curling free energies obtained from static simulations (Figure 1.12, see Chapter 3 for additional technical details) highlighted how, for linear molecules, the process is highly endoergonic ($\Delta F \sim 55$ kJ · mol^{-1}) while the situation is exactly reverted for the branched 3,4-divinylhexa-1,3,5-triene. Therefore, we believe that—once formed—branched species will undergo very fast cyclization. Multiple cyclizations finally lead to polyaromatic molecules that, at high temperature (400 °C), clog the zeolite pores leading to a fully deactivated material. Thanks to this mechanistic insights we were able for the first time to link the formation of branched polyolefins with catalyst deactivation in the MTO process.

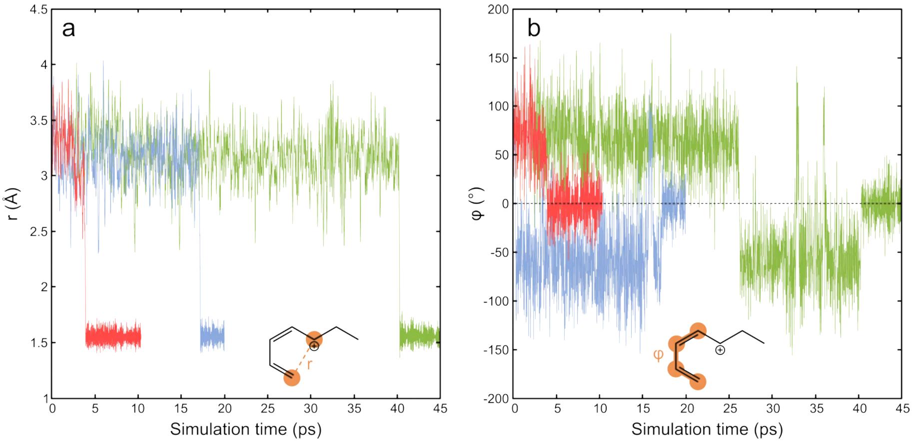


Figure 1.11: C1-C5 distance (*a*) and C1-C2-C3-C4 dihedral angle (*b*) during 3 MD simulations of 6-protonated 1,3,5-heptatriene at 350 °C. To ensure that the reaction always happens spontaneously and we did not randomly sample a rare event, we performed the three simulations starting from different initial dispositions of the adsorbate in the zeolite cage. After the cyclization had taken place, each simulation was allowed to continue for a few ps and then stopped.

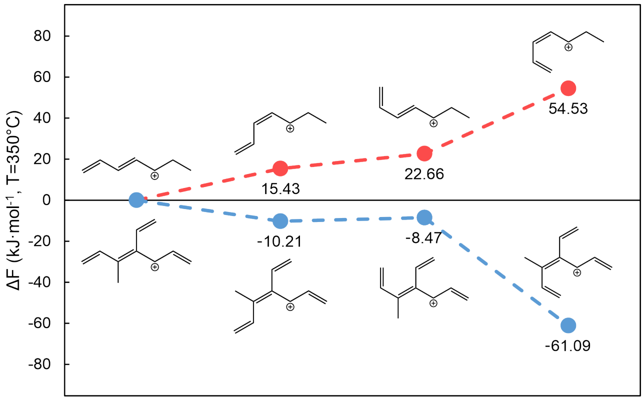


Figure 1.12: Relative free energies for the isomerization of protonated 1,3,5-heptatriene (blue dots) and 3,4-divinylhexa-1,3,5-triene (red dots).

2

Some key methodological concepts

After introducing zeolites from an industrial/technological standpoint, we are now going to outline some basic theoretical concepts that are necessary in order to understand how this type of materials can be modeled. First, we will briefly introduce the concept of DFT as a tool to solve the many-body electronic structure problem and derive system-related properties of interest. Then, we will present the main software that has been employed in this thesis to solve the DFT equations. Finally, some methodological challenges specifically related to zeolite catalysis modeling will be summarized. It should be noted that this chapter only collects a set of methods that are of importance throughout the thesis and that more specific methodologies are reported with greater detail in each of the subsequent chapters.

2.1 The DFT PES

The central quantity within any molecular modeling exercise is the PES of the system under study, which we are therefore now going to introduce. The concept of PES has its roots in the Born–Oppenheimer approximation. Since nuclei are much heavier than electrons, it can be assumed that the latter adjust instantaneously to the slow movements of the former. This means that the electronic and nuclear motions can be decoupled and the energy of the system depends only parametrically on the nuclear coordinates. In this way, we can write the functional relation $f : \{Z_\alpha, \mathbf{R}_\alpha\}_{\alpha=1}^{P} \rightarrow E$. This functional it the PES and it maps the nuclear charges (Z_α) and nuclear coordinates ($\mathbf{R}_\alpha$)

of P atoms to the potential energy (E). E, in first-principles simulations, is computed by solving the many-body electron problem encoded in the deceivingly simple-looking time-independent Schrödinger equation:

$$\hat{\mathcal{H}}\psi_e\left(\{\mathbf{r}_i\}_{i=1}^{N};\{\mathbf{R}_\alpha\}_{\alpha=1}^{P}\right) = E\psi_e\left(\{\mathbf{r}_i\}_{i=1}^{N};\{\mathbf{R}_\alpha\}_{\alpha=1}^{P}\right) . \tag{2.1}$$

The Hamiltonian operator $\hat{\mathcal{H}}$ includes the kinetic energy of the electrons and the electrostatic interactions between charged particles or, in formula,

$$\hat{\mathcal{H}} = -\sum_{i}^{N}\frac{1}{2}\nabla_i^2-\sum_{i}^{N}\sum_{\alpha}^{P}\frac{Z_\alpha}{|\mathbf{r}_i-\mathbf{R}_\alpha|}+\sum_{i}^{N}\sum_{j>i}^{N}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|}+\sum_{\alpha}^{P}\sum_{\beta>\alpha}^{P}\frac{Z_\alpha Z_\beta}{|\mathbf{R}_\alpha-\mathbf{R}_\beta|} , \tag{2.2}$$

where i and j run over the N electrons in the system while α and β over the P nuclei. Note that in the formula above $\hat{H}$ is expressed in atomic units. Unfortunately, the Schrödinger equation can be solved exactly only for systems with one electron, which are not particularly common in nature. Fortunately, there exist methods to approximately solve many-body electronic problems. Among the first was the Hartree–Fock method (HF), where Pauli's exclusion principle is taken into account by approximating the many-body wavefunction by one Slater determinant. While being a reasonable approximation, HF does not capture electron correlation but only the exchange part of the two-body interaction. Thus, HF is almost never used in current computational chemistry. Hereafter, we will focus on a second very common method to solve the many-body electronic problem, i.e. Density Functional Theory (DFT). DFT is basically the only method used throughout this thesis to compute first-principles PESs, being a good compromise between accuracy and computational efficiency. While the electronic wavefunction $\psi_e\left(\{\mathbf{r}_i\}_{i=1}^{N};\{\mathbf{R}_\alpha\}_{\alpha=1}^{P}\right)$ (eq. 2.1) is an exceedingly complex mathematical object, the electron density is a quantity much easier to interpret. Aside from being positive and real by definition, it cusps atop the atomic nuclei with a peak height that is proportional to the underlying atomic charge. If one also considers that integrating the electron density over space gives the total number of electrons, it appears quite reasonable that all the information contained in the Hamiltonian (eq. 2.2) is actually also encoded in the electron density itself.

These concepts were rigorously formulated in the two theorems by Hohenberg and Kohn, where it was proven that the energy of a system is a functional of the electron density (hence density functional theory). Later Kohn and Sham proposed a scheme where our real system density, made of real interacting electrons, is connected with the density of a fictitious system made of non-interacting electrons. This led to the Kohn-Sham equations, that can be solved in a self-consistent way—a knack of modern computers. While being conceptually very interesting, the crucial weakness of DFT lies in the fact that the exact functional is not known. Indeed, for non-interacting electrons, the electronic kinetic energy and coulomb repulsion terms can be easily written as functional of the electron density. In doing so, we are left with an additional unknown contribution to the total energy called **exchange-correlation functional** (XC), which should give the final tweak to the non-interacting electron density and make it identical to the real one. Thus, it can be stated that while HF is an approximate method with an exact solution, DFT is an exact method with an approximate solution due to the ad hoc introduction of the XC functional.

Exchange correlation functionals are commonly classified following the clever analogy with Jacob's ladder proposed by Perdew[94], where we start from the Hartree world (which is exchange-correlation free) and we gradually climb towards the heaven of chemical accuracy (Figure 2.1). Throughout this thesis, the results are most often obtained using GGA (generalized gradient approximation) methods. In particular, the Perdew-Burke-Ernzerhof[95] (PBE) functional is the method of choice to investigate zeolite-catalyzed reaction, because of its relatively low cost and its generally good accuracy in capturing reactive trends[96, 97]. Two exceptions are represented by **Paper II** and **Paper IV**; in the latter, we compare the zeolite catalyst with an aqueous Brønsted acid. The GGA BLYP functional[98, 99] was used in that case, to allow comparison with previous results obtained within our group[46]. For **Paper II**, we improved the accuracy of the PBE results with the hybrid B3LYP functional[100, 101], which introduces a fraction of exact HF exchange to the BLYP functional (Eq. 2.3). This decision was made because the paper focuses on Pd complexes, for which B3LYP is a common choice in computational investigations[102]. The B3LYP XC functional is defined as:

$$E_{\mathrm{XC}}^{\mathrm{B3LYP}} = (1-a)E_{\mathrm{X}}^{\mathrm{LSDA}} + aE_{\mathrm{X}}^{\mathrm{HF}} + b\Delta E_{\mathrm{X}}^{\mathrm{B}} + (1-c)E_{\mathrm{C}}^{\mathrm{LSDA}} + cE_{\mathrm{C}}^{\mathrm{LYP}} \; , \quad (2.3)$$

Where $a = 0.2$, $b = 0.72$ and $c = 0.81$ and LSDA stands for local spin density approximation.

It must be noticed that part of the electron correlation is not correctly captured by basically any DFT method currently available. The main deficiency this creates is the (partial) lack of Van Der Waals interactions. These are weak attractive interactions between atoms caused by instantaneous fluctuations in the electron density. While they are indeed weak compared to other interaction types, it is also true that their purely attractive nature can produce very large effects, especially when adsorption phenomena are considered. Currently, one of the most successful and widely used methods to include

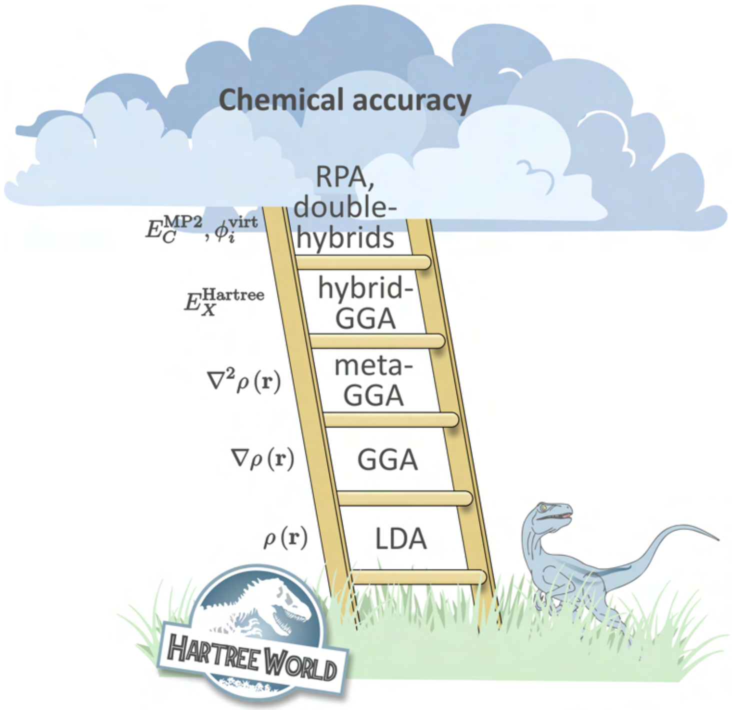


Figure 2.1: Jacob's ladder is an analogy to rank DFT functionals, going from the ground level non-interacting electrons to the heaven of chemical accuracy.

dispersive interaction in DFT functionals is the D3 scheme proposed by Grimme[103]. This cheap but effective method simply consists in applying an a posteriori pair-wise correction to the DFT energy, which has the form

$$E_{\mathrm{disp}} = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(f_{d,6}(r_{ij})\frac{C_{6ij}}{r_{ij}^{6}} + f_{d,8}(r_{ij})\frac{C_{8ij}}{r_{ij}^{8}}\right), \tag{2.4}$$

in which the dispersion coefficients C_{6ij} and C_{8ij} are geometry- and functional-dependent and $f_d(r_{ij})$ is a damping function. For the latter, both a zero-damping and a Becke–Johnson damping[104] variants exist. In this thesis, all calculations employ the D3 dispersion scheme, with one of the two damping variants (the differences in energy are, based on our tests, marginal).

2.2 Employed software for DFT simulations

Nowadays, tens of programs are capable of performing DFT calculations. Throughout this thesis, all simulations have been performed with either the Vienna Ab initio Simulation Package[105–107] (VASP) or with CP2K[108, 109].

VASP is a software to perform periodic DFT simulations, where the electron density is expanded with a plane-waves (PWs) basis set. A plane-wave can satisfy the periodicity of a crystalline lattice if defined as

$$\phi_{\mathbf{K}}\left(\mathbf{r}\right) = \frac{1}{\sqrt{\Omega}} e^{i\mathbf{K}\cdot\mathbf{r}}, \tag{2.5}$$

where Ω is the unit cell volume and **K** a reciprocal lattice vector. Using a linear combination of PWs offers several advantages, as they are orthonormal by definition and, if an infinite number is considered, they form a basis for any function with lattice periodicity (or, in other words, increasing the amount of PW used as basis set for a DFT calculation systematically improves results accuracy for the given DFT functional). Alas, the PW basis set used in VASP calculations must be truncated to perform the simulations in a reasonable amount of time. This truncation is mostly problematic atop the nuclei, where the electron density tends to cusp as explained before. For this reason, all VASP simulations presented in this thesis resort to the projected-augmented waves method (PAW)[110, 111]. It is also important to remember that the final electron density for a periodic material should be obtained by performing an integration over the Brillouin zone in the reciprocal lattice. Yet, zeolites have normally quite large unit cells (meaning a very small reciprocal cell and Brillouin zone) and, being insulators, their band structure is mostly flat. Thus, accurate results can be obtained by considering only the Γ point of the Brillouin zone, which remarkably speeds up the computations. All periodic simulations performed throughout this thesis limit the sample of the Brillouin zone to the Γ point.

CP2K is also used to perform periodic DFT simulations and the main difference with VASP lies in the choice of basis set. Indeed, CP2K uses a combination of atom-centered gaussian basis functions and PWs[112, 113], performing different tasks on the most suited basis set in order to speed-up the calculations. This makes CP2K—in general—faster than VASP. While being much more ‘chemical intuitive’ than PWs, atom-centered orbitals are not intrinsically orthonormal and they are less robust in convergence with respect to PWs. As in VASP, care must be taken when employing PWs in proximity of the atomic cores. For CP2k, the electron density is smoothed using Goedecker–Teter–Hutter (GTH) pseudopotentials[114].

In general, we have used VASP when performing static simulations (Chapter 3) because of its easily improvable accuracy and robustness, while we resorted to CP2K when performing MD simulations (Chapter 4) because of its remarkable speed while maintaining a good accuracy.

2.3 Modeling a zeolite catalyst

As explained in Section 1.4, zeolites are quite complex materials at the microscopic scale. Nonetheless, if one expects the reactivity to be mostly influenced by the local environment around the active site, it is relatively easy to construct a model for—in the simplest case—the pristine BAS in interaction with some reactive molecule. This is one of the main reasons for the popularity of zeolites in computational catalysis starting from the early 80s, when ab initio calculations on more than a couple atoms were becoming possible[115].

Initially, small cluster models were used, comprising the BAS, the Al defect and few surrounding Si atoms (Figure 2.2*a*)[116]. The dangling valences of the terminal atoms were capped with hydrogen, whose positions were sometimes fixed to mimic the restrain due to the missing framework atoms[115]. While certainly useful to obtain initial insights in the reactivity and behavior of the BAS in interaction with simple organic molecules, it soon became evident that many important effects were neglected by the lack of the surrounding zeolite walls[31]. Indeed, it is now known that the high acidity of zeolites does not only come from the intrinsic high electrophilicity of the BAS itself (whose acidity is actually quite moderate[117]) but also from the large stabilizing effect that the surrounding framework has on protonated organic molecules due to non-covalent interactions[118, 119]. By using larger and larger cluster models it was found that a major effect had to be expected on the reaction barriers and on the stability of charged species[31].

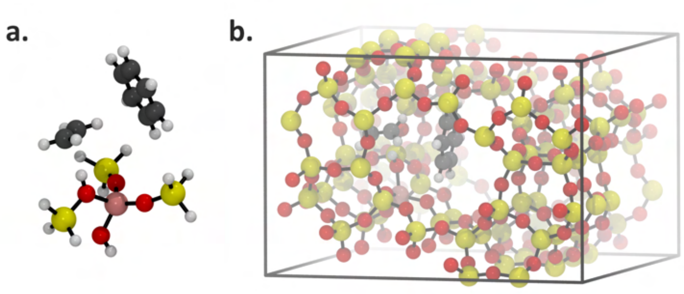


Figure 2.2: *a.* An ethene and a benzene molecules adsorbed on a T4 cluster representation of the H–ZSM–5 zeolite, as optimized by Vos et al.[28] *b.* Same as in *a*, but this time with a fully periodic H–ZSM–5 unit cell, as optimized by Hansen et al.[31] The large difference in the chemical environment experienced by the adsorbed molecules between the two cases is evident.

To solve the problem a periodic model can be used, where a portion of the system (usually one or few more unit cells) is repeated indefinitely in space

(Figure 2.2*b*). This allows to take into account the crystalline nature of the framework. Starting from 1996, full DFT MD on small periodic zeolites became possible and were first used to study the behavior of methanol adsorbed in the acidic SOD and CHA frameworks[120, 121]. Nowadays, almost all theoretical investigations of zeolite-catalyzed reactions are performed using periodic simulations with a flavor of GGA-DFT, which is also the technique we primarily adopted throughout this thesis. The general strategy to get an initial model starts with the atomic coordinates of an all-silica framework, which are readily available from the IZA database of zeolite structures[12]. A plethora of programs exists to modify the pristine framework and add the Al defect, the BAS, adsorbed organic molecules, etc. Here, we mainly relied on our in-house developed ZEOBUILDER software[122].

Having discussed how a model of the zeolite can be constructed, a little more can be said on the appropriate level of theory for the simulations—i.e. the method with which energy and forces are evaluated. In most studies, periodic zeolites are characterized with GGA-DFT methods (very commonly PBE–D3, see Section 2.1). But is GGA-DFT sufficient to obtain accurate results? The answer, according to literature, is *qualitatively* yes but *quantitatively* not always. Starting from the pioneering work of Tuma and Sauer[123, 124], it soon became obvious that GGA-DFT systematically underestimates reaction barriers. The proposed methodology to counteract this effect consists in extracting a relatively large cluster from the periodic model and compute its energy at an MP2 level of theory. MP2[125] (Møller–Plesset second-order perturbation theory) is a post-HF method, meaning that electron correlation is introduced—in this case—through perturbative corrections. Then CCSD(T)[126] is used on an even smaller cluster around the active site to recover the remaining electron correlation. CC (couple cluster) with single, double and triple electron excitations is a post-HF method as well, and it is currently accepted as the golden standard for chemical accuracy. This means that its predictions are expected to be within 1 $\text{kcal}\cdot\text{mol}^{-1}$ from the real value. A graphical depiction of this cluster-in-periodic approach is given in Figure 2.3. Additional corrections can be used to account for the basis set incompleteness. With the development of approximate and cheaper flavors of CCSD(T), like DLPNO-CCSD(T)[127, 128], it has now become possible to skip the MP2 step and immediately use the CC calculation on a large enough cluster[96].

In spite of their accuracy merits, the above mentioned composite approaches that use post-HF methods are very expensive and not suited for dynamic simulations (see Chapter 4). Luckily, extensive benchmark studies conducted by Studt and co-workers[96, 97] have shown that GGA-DFT, despite systematically underestimating reaction barriers, mostly preserves reactivity trends and can

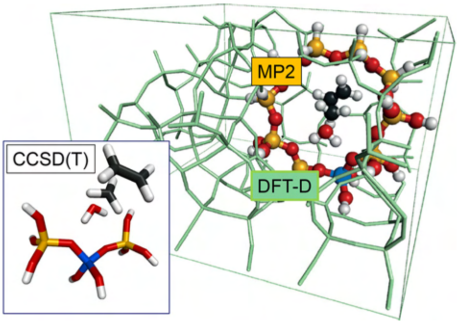


Figure 2.3: To improve the energy estimate of a zeolite-catalyzed reaction, a correction can be made by considering a relative large cluster computed with MP2 and a smaller cluster computed with CCSD(T) accuracy. Reprinted with permission from ref. 129. Copyright 2019, American Chemical Society. The figure was adapted with permission from ref. 130. Copyright 2016, John Wiley and Sons.

therefore be confidently used to determine reaction mechanisms and obtain molecular insights in the process under study. For this reason, throughout this thesis we mainly adopted GGA functionals to characterize the chemistry of the system, being well-aware that the obtained reaction barriers are in general underestimated.

Before concluding, it is interesting—in my opinion—to mention a case study from the literature on computational zeolite catalysis that even nowadays remains particularly troublesome. The reaction in question is the formation of carbocations upon alkenes adsorption on the zeolite BAS. This is a relevant process for industry, as zeolites are widely used in conversion reactions involving alkenes. For instance, fluid catalytic cracking (FCC) is used in oil refineries to crack heavy hydrocarbon fractions to higher added-value products, such as gasoline[131–133]. When an unsaturated compound adsorbs in a Brønsted acidic zeolite, its double bond strongly interacts with the electrophilic BAS, forming a so-called π-complex (Figure 2.4*a*). The BAS can then attack the double bond, creating a positive charge on one of the two carbon atoms. This carbenium ion (Figure 2.4*b*) can then be chemisorbed back on one of the framework oxygen atoms to form a surface alkoxide species (SAS, Figure 2.4*c*). Note that this conversion is fundamentally the same as TS0 in the benzene ethylation reaction, that we discussed extensively in Section 1.3.1.

The interesting challenge in computing the alkene adsorption process is that the different intermediates are stabilized by very different factors. Indeed, the alkene and the carbenium ion can be highly mobile, strongly benefiting from higher temperatures because of the higher entropy with respect to the immobile SAS. On the other hand, the carbenium ion is disfavored

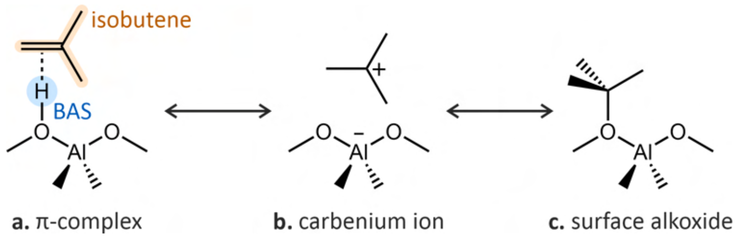


Figure 2.4: Upon adsorption in a Brønsted-acidic zeolite, an alkene (isobutene is here taken as example) is firstly physisorbed on the BAS, forming a π-complex. The molecule can then be protonated by the BAS, forming a carbenium ion. If the latter chemisorbs on the framework, a surface alkoxide species is formed.

enthalpically because of its charged nature. Moreover, increasing the degree of substitution on the carbocation stabilizes it, while disfavoring the SAS because of steric hindrance. Obviously this enthalpy-entropy balance is also strongly dependent on the magnitude of the confinement effect due to the volume of the zeolite voids[134]. Hereafter we will focus on studies performed on isobutene which, aside from being a typical FCC intermediate, is the smallest alkene containing a tertiary carbon and has represented a central case study for computational chemists since the early 2000[124, 135–141].

To fully understand the state in which isobutene preferentially exists in Brønsted acidic zeolites, it is essential to accurately determine which of the three species prevails at reaction conditions. Unfortunately, an accurate estimate of the entropic stabilization for the π-complex and the carbenium ion naturally requires the use of operando modeling techniques that explicitly take temperature effects into account (more details are given in Chapter 4). The problem is that this type of calculation is very expensive and can only be performed with GGA-DFT in a reasonable time. However, GGA-DFT has the tendency to artificially spread electron density due to the lack of correlation[142] (Section 2.1), delocalizing charges and hence massively overestimating the stability of carbenium ions[143] (Figure 2.4*b*).

Currently, there are two major 'schools of thought' in literature. On the one hand, enhanced sampling MD—which explicitly takes into account temperature effects—indicates carbenium ions (at least tertiary ones) to be a predominant species in zeolite-catalyzed reactions, because of the large entropic stabilization at the high catalytic temperature of processes like FCC[144, 145]. On the other hand, groups that resort to static calculations with high-level corrections to the PES claim that their importance is limited, as enthalpy is artificially lowered by tens of $\mathrm{kJ \cdot mol^{-1}}$ because of the GGA functionals used in MD[89, 143, 146]. While the formation of protonated species has been unequivocally observed experimentally only for more electrophilic molecules with (extended) aromatic systems[27, 118, 147, 148] the current evidences for the presence of alkyl carbenium ions[149–152] are still considered ambiguous[153]. Indeed, it is experimentally very difficult to observe these metastable intermediates, as with their pronounced reactivity they tend to rapidly rearrange forming other species.

Recently, there have been attempts to reconcile the two worlds by performing higher level corrections directly on the free energy profiles obtained from enhanced sampling MD. For instance, Bučko and co-workers have used a kernel ridge regression[154] (KRR) machine learning algorithm along with the REMatch kernel[155] to learn the energy difference between GGA and RPA (random phase approximation, one of the last steps in Jacob's ladder, see Figure 2.1) calculations from a limited set of structures extracted from a GGA Blue-Moon simulation[156–158]. The model evaluates then the same difference for all structures from the simulation, allowing to correct the expectation value of an observable (the free energy in this case) thanks to free energy perturbation theory (FEPT).

$$\langle O \rangle_{\mathcal{H}'} = \frac{\langle O \exp\left(-\beta \Delta E\right) \rangle_{\mathcal{H}}}{\langle \exp\left(-\beta \Delta E\right) \rangle_{\mathcal{H}}} \,. \tag{2.6}$$

In eq. 2.6, the ensemble average of an observable O can be evaluated based on a high-level hamiltonian $\mathcal{H}'$ if the ensemble average obtained with the low-level hamiltonian $\mathcal{H}$ is reweighted with the potential energy difference between the two methods ΔE. Of course, as many evaluations of ΔE as possible have to be performed to converge the average and that is precisely the task that the machine learning model helps to speed up. It is also important to notice that care has to be taken to ensure that both the low- and high-level methods explore a compatible phase space, otherwise the obtained results will not be accurate[159]. FEPT can be very effective to improve the results of a simulation but, unfortunately, it is readily applicable only to the Blue-Moon sampling method[160]. This is because, in Blue-Moon sampling, MD simulations are performed with a fixed collective variable (CV) value, producing discrete free energy points along the CV path. For the moment, the CV can be imagined as a simple variable describing the progress of a reaction, more details are provided in the dedicated Section 4.2.1 of Chapter 4. Since all MD samples of a simulation relates to a single CV value, very good statistics is obtained for each of the considered points, which is necessary to properly converge the properties derived from FEPT. For methods that provide a continuous free energy surfaces (FESs), like metadynamics (Section 4.2.2) or umbrella sampling (Section 4.2.3) an unrealistic simulation time would be required to obtain converged results. An alternative method has been proposed by De Wispelaere et al.[161], in which a single point calculation at a high level of theory is used to compute the enthalpic correction, after a constrained optimization at a fixed CV value. This approach could be valid only if the correction is consistent across all states visited during the MD simulation, which makes the method somehow less robust.

While both these approaches have certainly contributed to enhance our understanding of how the results of MD simulations can be pushed towards chemical accuracy, I believe that the final answer will come from the world of MLPs (Chapter 5). Indeed, it has now become possible to use a very limited amount of single point calculations to obtain a fast and accurate estimate of the PES for a chemical reaction. Since recently robust implementations of fully periodic MP2[162, 163] and double-hybrid functionals[164, 165] have become available in mainstream computational chemistry softwares like CP2K, it seems only natural to use them to train an MLP which can then be used to directly compute high-level FESs without restrictions in terms of sampling time or problems related to the phase space coverage of different methods. We will further elaborate on this in Section 5.4, at the end of Chapter 5, which is fully dedicated to machine learning potentials.

3

Exploring zeolite-catalyzed reaction mechanisms with static simulations

*In Section 2.1 we discussed how a DFT flavor can be used to retrieve the energy and forces of a chemical system composed of P atoms with coordinates $\{\boldsymbol{R}_1, \boldsymbol{R}_2, ..., \boldsymbol{R}_P\}$. Excluding the rotational and translational degrees of freedom, we are left with a $3P - 6$ dimensional surface that contains all relevant information about the system. In this section, we will explore how stationary points on this surface represent stable structures and transition states of a chemical reaction. We will compute the second derivative of the potential energy at these points to determine their normal vibrational modes and use them to compute reaction and activation free energy differences. This approach has been applied successfully to explore the reactivity and selectivity of zeolite-catalyzed dealkylation of alkylphenols (**Paper I**) and the toluene homocoupling reaction in Pd-loaded zeolites (**Paper II**).*

3.1 Vibrational analysis to characterize stationary points

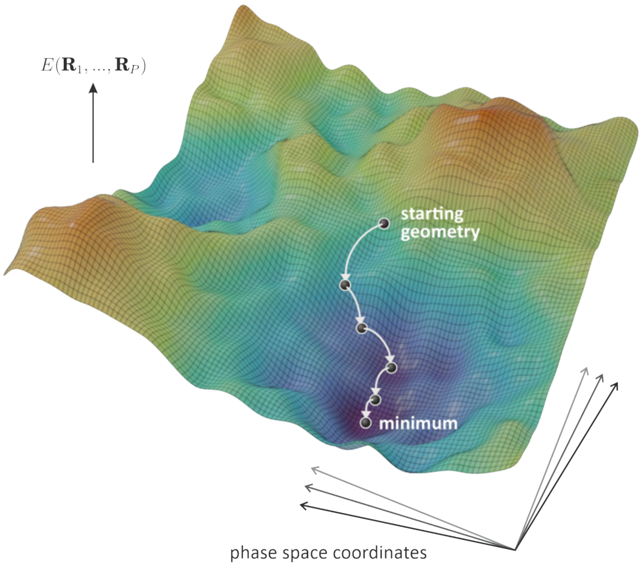
In Section 2.3 we illustrated how an atomistic model for a zeolite catalyzed reaction can be built and in Section 2.1 how DFT can be used to compute the PES of such system. Here, we will discuss how a chemical reaction can be characterized based on a limited set of significant points on such PES, how these points can be retrieved and how quantities of interest such as free energy differences and reaction rates can be computed. In good approximation, it is reasonable to think that three points are sufficient to describe a chemical reaction, namely reactants, transition state and products. Reactants and products correspond to an energetically stable geometry of the system—i.e. a minimum on the PES. In classical harmonic **transition state theory** (TST), the transition state is the highest point in energy on the path connecting reactants and products, while being a minimum in any other direction. In other words, if we imagine the reaction as the crossing of a mountain range, the likeliest transition state to be crossed will correspond to the mountain pass with the lowest altitude.

When using DFT to compute the energy of a certain atomic configuration $E(\mathbf{R}_1, ..., \mathbf{R}_P)$, one can also obtain analytically the forces acting on the atoms of the system, i.e. the inverse of the energy gradient with respect to the atomic coordinates $\mathbf{F}_\alpha = -\partial E/\partial \mathbf{R}_\alpha$. Knowing the gradients is essential to reach stationary points, as any common minimization algorithm can be used to reach a minimum on the PES. The process is called **geometry optimization**, and it is schematically depicted for an hypothetical PES (reduced to only two dimensions for the sake of clarity) in Figure 3.1*a*. A handful of efficient algorithms is available to perform this task. For the results presented hereafter, we normally resorted to the conjugate gradient method as implemented in VASP. Optimizing a transition state is somewhat more difficult and requires to introduce the concept of frequency calculation beforehand, therefore it will be discussed in more detail further in this section.

A stationary point on the PES is considered to be reached when all forces acting on the system's atoms are lower than a certain user-defined threshold or, alternatively, when the energy difference between two subsequent optimization steps decreases below a certain limit. While the first approach is theoretically more robust, the second is often used as forces are more susceptible to numerical noise on the PES, hampering convergence. As one could imagine, in unlucky instances it is possible that the system ends up in a stationary point which is not a true PES minimum, meaning that along one or more internal degrees of freedom the optimization has stopped on a local maximum. To check this, a normal mode analysis (NMA) has to be performed. This corresponds to compute and diagonalize the Hessian matrix, with elements $H_{\alpha\beta} = \partial^2 E/\partial \mathbf{R}_\alpha \partial \mathbf{R}_\beta$, which returns the system's vibrational frequencies in the harmonic approximation. A schematic depiction of a NMA analysis is shown in Figure 3.1*b*. A true minimum in the PES will be characterized by all-positive frequencies, while a transition state by a single imaginary frequency (i.e. with a negative force constant) along the direction connecting the two minima of interest (Figure 3.1*b*).

Computing the normal modes of the system is not only important to deter-

a. Geometry optimization



b. Frequency calculation

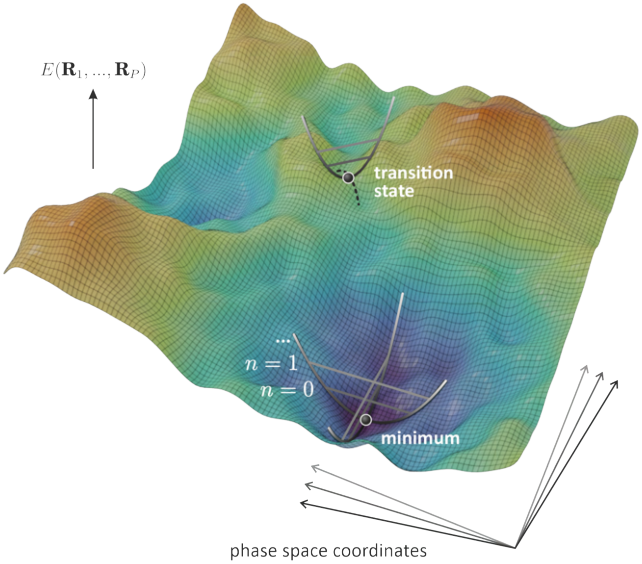


Figure 3.1: *a.* A geometry optimization allows to reduce the energy of a system until a stationary point is reached (a PES minimum in the figure). *b.* In the harmonic approximation, the PES in the neighborhood of a stationary point is approximated with a parabola, leading to a set of harmonic vibrational modes. For a transition state, one of these modes is imaginary and corresponds to the barrier crossing (dotted line in the figure).

mine the type of stationary point that has been reached, but also to compute thermodynamic quantities of interest such as entropy, enthalpy and free energy differences. This makes intuitive sense, as by approximating the neighborhood of the stationary point with a parabola one can assess how flat the PES is in that direction. A flat PES (i.e. associated with a low force constant of the parabola or, equivalently, with a low vibrational frequency) means that a lot of low-energy states are present, heavily increasing the entropic likelihood of the stationary point. Based on the principles of statistical physics, the molecular **partition function** of a system can be expressed as a product of translational, rotational, vibrational and electronic components[8]:

$$Q = q_{\mathrm{trans}} q_{\mathrm{rot}} q_{\mathrm{vib}} q_{\mathrm{elec}} \,. \tag{3.1}$$

For molecules in gas-phase, q_{trans} and q_{rot} can be easily approximated with the particle in a box and the rigid rotor quantum mechanical models. For solids, their contribution is of course absent. The electronic part q_{elec} is also of limited importance, as the energy difference between the fundamental electronic state and the first excited state is normally large enough so that only the former results to be occupied. The part we are most interested in is the vibrational contribution. In the harmonic approximation, it can be expressed as

$$q_{\mathrm{vib}} = \prod_{i=0}^{3P-6} \frac{\exp\left(-\frac{1}{2}\beta h\nu_i\right)}{1-\exp\left(-\beta h\nu_i\right)}\,, \tag{3.2}$$

where $\beta = (k_B T)^{-1}$ and the potential energy of the stationary point is taken as reference zero. ν_i are the normal modes of the system. The partition function can then be used to compute any thermodynamic quantity of interest:

$$S = \beta^{-1} \left(\frac{\partial \ln Q}{\partial T}\right)_{N,V} + k_B \ln Q \,. \tag{3.3}$$

$$U = -\left(\frac{\partial \ln Q}{\partial \beta}\right)_{N,V} \,. \tag{3.4}$$

$$F = U - TS \,. \tag{3.5}$$

S, U and F are the entropy, internal energy and Helmholtz free energy of an isochore system, respectively (to move towards an isobaric system and obtain the Gibbs free energy, a factor pV is still required). By computing the partition function of reactants and transition state, moreover, the reaction rate can be retrieved using Eyring's TST formalism[166]:

$$k = \frac{1}{\beta h}\frac{Q_{\mathrm{TS}}}{Q_{\mathrm{R}}}e^{-\beta\Delta E^{\ddagger}} , \tag{3.6}$$

with $\Delta E^{\ddagger}$ the energy difference between reactants and transition state. Since the hessian cannot be computed analytically, it is obtained numerically with finite displacements of the atoms. This means that $2P$ single point calculations are required, where P is the number of atoms, making a frequency calculation rather expensive. For this reason, we adopted throughout this thesis a partial hessian vibrational analysis[167, 168] (PHVA) approach to characterize stationary points, in which only the adsorbate and a T8 cluster around the Al defect (see Figure 1.3*a*) are considered for the frequency calculation while all other atoms in the system are kept frozen. Limiting the frequency calculation to the T8 cluster has been shown to provide sufficiently converged thermodynamic properties, as the contribution from the remaining framework vibrations will mostly cancel out when computing free energy differences[169].

It is important to notice that the harmonic approximation can often be inadequate to produce reliable free energy values. If a small molecule is inserted in a zeolite with large pores, for instance, it will have almost free rotations and vibrations along the channels directions. These are clearly very anharmonic in nature and, given the extremely low magnitude of the force constant, subjected to heavy numerical noise. There are nowadays a handful of strategies to overcome this problem and obtain more reliable thermodynamic quantities, for instance including anharmonicity or treating some of the adsorbate modes as (partially) free rotations and translations[169–172]. This is particularly relevant to compute adsorption energies, as the harmonic approximation can severely overestimate the entropic loss associated with the passage from gas-phase to the constrained zeolite pore[173, 174]. As we will not particularly focus on the adsorption step throughout this thesis, we opted in general to accept the limitations of the harmonic approximation and replace small amplitude frequencies with an arbitrary value (in general 60 cm^{-1}). This choice is relatively common in literature[175–177] as one reduces the risk of having a high numerical noise on the lowest frequencies, relying on error cancellation when evaluating intra-zeolite free energy differences.

While optimizing towards a minimum is relatively easy, reaching a transition state is somewhat trickier as one would like to optimize along all degrees of freedom except the one connecting reactants and products, which should be maximized. There exist many algorithms that allow for this, among which the nudged elastic band (NEB) method[178] and its climbing image variant[179] are quite commonly used. In the following results, we rely on the improved dimer method[180]. This is cheaper than NEB but tends to work only if the initial geometry is already very close to the TS. Once a guess for the TS has been constructed with reasonable lengths of the bonds involved in the reaction, an initial optimization is performed either with frozen atoms or with frozen bond lengths. The latter is preferred, allowing the molecule to reorient in the zeolite pore, but requires an external interface (the atomic simulation environment, ASE[181], in this case) to perform the optimization, not being natively available in VASP. After a NMA is performed, the imaginary mode is used to nudge the system in the TS direction until the forces magnitude goes below the convergence threshold. We found that the equilibrium geometry obtained with the improved dimer method can be further relaxed without escaping the transition state region if a quasi-Newton algorithm[182] is used.

This concludes the list of ingredients needed to characterize a chemical reaction using static simulations. In the following section, we will see how the methodology can be applied to zeolite-catalyzed reactions to obtain mechanistic insights on their activity and selectivity.

3.2 Application to zeolite catalysis

3.2.1 Alkylphenols dealkylation

In **Paper I**, static simulations were used to obtain mechanistic insights on the alkylphenol dealkylation reaction. As outlined in Section 1.3.2, this reaction is of high relevance in the field of biomass conversion, as it allows to convert lignin-derived alkylphenols in more valuable molecules, namely phenol and olefins[21–25, 41]. Interestingly, there were some experimental observations still difficult to rationalize. Among them, was the significant difference in reactivity between substrates. Experiments focused mainly on three different reactants, 4-ethylphenol (4-EP), 4-*n*-propylphenol (4-*n*-PP) and 4-isopropylphenol (4-*iso*-PP)[23, 24]. It was found that, using H−ZSM−5 with Si/Al=40, 4-EP reaches 50% conversion at about 650 K. 4-*n*-PP is more reactive and reaches 50% conversion at 575 K. Finally, the reactivity of 4-*iso*-PP is remarkably higher and 50% conversion is reached below 500 K.

To better understand the reaction mechanism and these differences in reactivity, we performed a set of static simulations that covered all the reaction pathways discussed in Section 1.3.1 (Figure 1.5). The results are shown in Figure 3.2. The reaction barriers show quite a nice agreement with the experimental trends. Indeed, 4-EP dealkylation has the highest barriers for the dealkylation step, followed by 4-*n*-PP with differences that are, however, not particularly marked. 4-*iso*-PP on the other hand is clearly much more reactive and exhibits a sharp decrease in all the computed barriers. We attributed this effect to the methyl substitution on the alkyl tail. Indeed, substituted alkenes tend to be more stable thanks to hyperconjugation with the C-H σ bonds. This effect is minor in the case of 4-*n*-PP, but it becomes important for 4-*iso*-PP, where conjugation occurs directly with the carbon atom carrying the partial positive charge. The stabilization is sufficient to make an isopropylcation intermediate appear as shallow minimum on the PES (Figure 3.2*c*), although this might be an artifact caused by the adopted level of theory (Section 2.3).

Mechanistically, the calculations suggest that the reaction proceeds preferentially through the concerted mechanism (TS2 in Figure 3.2). Its energetic requirement is indeed always slightly lower than the alkoxide deprotonation (TS4) and the intramolecular dealkylation (TS5) presents the highest activation energy. In conclusion, we have briefly illustrated some of the results of **Paper I**, highlighting how static calculations were able to provide some interesting reactivity trends and mechanistic insights in a potentially relevant zeolite-catalyzed process. For additional details, we refer the interested reader to the published version of **Paper I**, appended in the second part of this thesis.

3.2.2 Pd-catalyzed toluene homocoupling

Creating carbon-carbon bonds is a pivotal task for any organic chemist. With this purpose, transition metals catalyzed cross-coupling reactions are among the most commonly used. Indeed, Heck, Negishi and Suzuki received the 2010 Nobel Prize in chemistry for their development[184]. In the Suzuki reaction, an homogeneous Pd complex is used to catalyze the C-C coupling of an arylboronic acid and an aryl halide to link the two aromatic rings (Figure 3.3*a*). Mechanistically, the reaction starts with an oxydative insertion of the aryl alide on the $Pd^{(0)}$ complex (Figure 3.3*b*). Then, the so-formed $Pd^{(II)}$ complex transmetallate with the arylboronic acid, meaning that Pd displaces the boron and bonds with the aryl fragment. The two aryl fragments, now both bonded to Pd, can then form a new σ bond between them, leaving two electrons on the Pd atom. This leads to the final cross-coupling product and regenerates the initial $Pd^{(0)}$ catalyst.

In spite of its importance, the Suzuki reaction suffers from some inconvenient problems: first, the two coupling reagents must be preactivated by introducing an halogen atom and a $B(OH)_2$ group on the carbon atoms whose coupling is desired. This requires extra syntetic steps and produces additional waste at the end of the coupling reaction. For this reason, the focus in recent

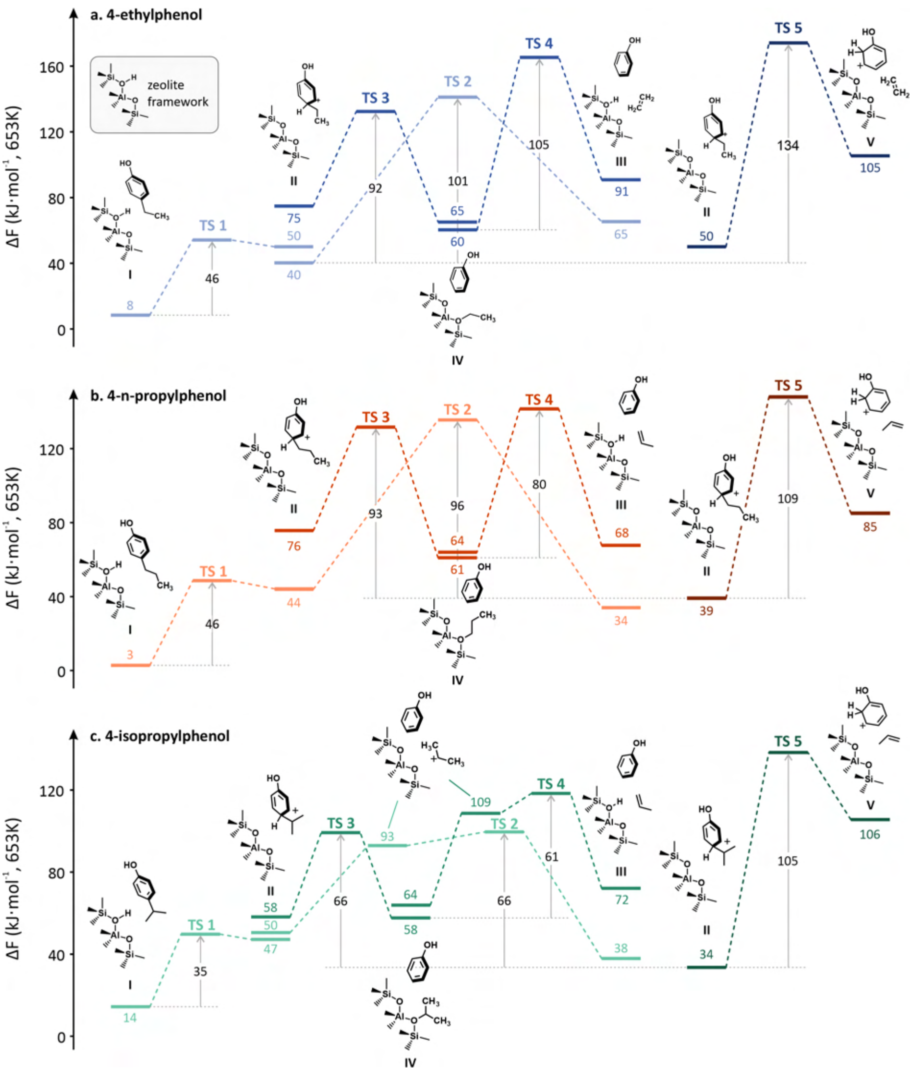


Figure 3.2: Free energy profiles for the three possible dealkylation pathways (compare with Figure 1.5) of 4-EP (*a*), 4-*n*-PP (*b*) and 4-*iso*-PP (*c*). The values come from static calculations performed with VASP at a PBE-D3 level of theory. Reprinted with permission from Ref. 183.

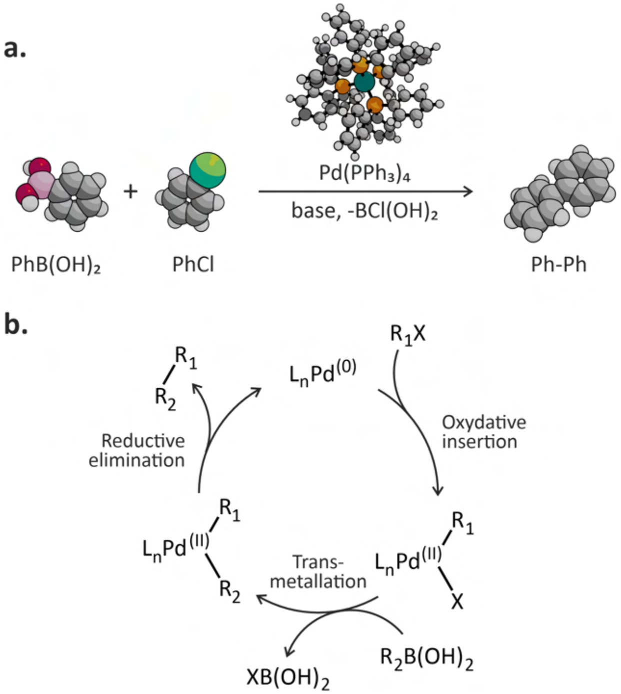


Figure 3.3: *a.* A protototypical example of Suzuki coupling, where chlorobenzene and phenylboronic acid are converted to biphenyl in the presence of a base and $Pd^{(0)}$ tetrakis(triphenylphosphine) as catalyst. *b.* The accepted homogeneous cycle of a cross-coupling reaction, where the organohalide compound undergoes oxydative addition on the $Pd^{(0)}$ complex. The boronic acid then transmetallates, creating a Pd^{2+} center coordinated to both the organic residues. A σ bond is formed between the latter, leaving the Pd center in a reductive elimination step and regenerating the initial $Pd^{(0)}$ complex.

years has shifted towards coupling reactions that can directly activate C–H bonds. This, however, triggers a second problem. Indeed, for relatively simple aromatic compounds and simple Pd complexes the reaction is poorly selective, meaning that basically all aromatic C-H bonds can be activated.

Our experimental partners from Prof. D. De Vos' group at the KU Leuven proposed an interesting solution to both issues, which is extensively presented in **Paper II**. They showed that $Pd(OAc)_2$ can be chemisorbed in medium-pores zeolites (H–Beta, in particular) and catalyzes with remarkable activity the homocoupling between two toluene molecules (Figure 3.4). Molecular oxygen is used to reactivate the catalyst, simply producing water as reaction by-product. The very interesting advantage related to the use of a zeolite is that the restrained environment induces selectivity in the reaction. While in

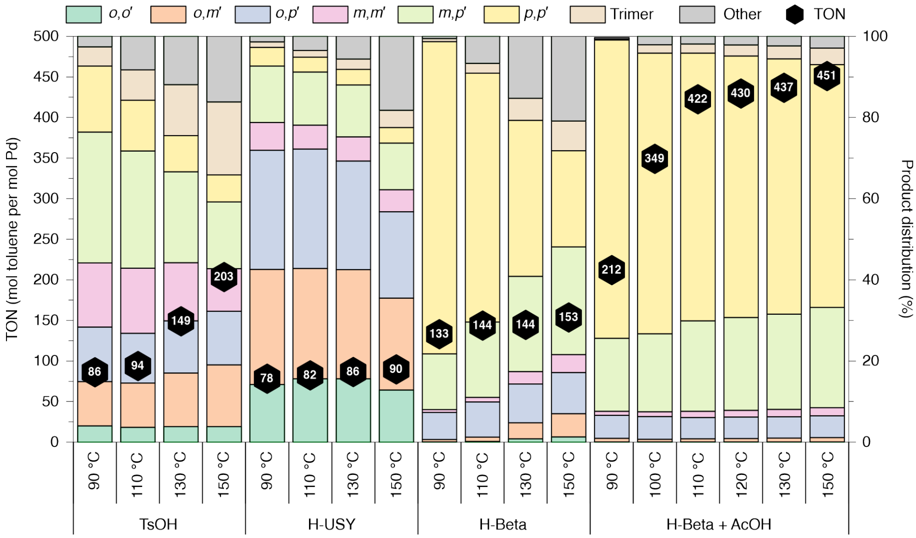


Figure 3.4: The H–Beta zeolite in the presence of additional acetic acid outperforms the homogeneous catalyst and the H–USY zeolite both in terms of selectivity and reaction TON (number of reactive events per active site). Reaction conditions: 15 μmol $Pd(OAc)_2$, 2 ml toluene, 16 bar O_2, 16 h. Following amounts were employed: TsOH (0.3 mmol) and 1 ml AcOH, H-USY (Si/Al = 40, 50 mg), H-Beta (Si/Al = 75, 100 mg), H-Beta + AcOH (200 equiv. AcOH to Pd). Adapted from Ref. 185, with permission from Springer Nature.

homogeneous phase an unbiased distribution of bitolyl isomers is obtained (using tosylic acid as co-catalyst, TsOH in Figure 3.4), the H–Beta framework remarkably favors the formation of the *para,para'* isomer (i.e. both methyl groups are in *para* with respect to the newly formed C–C bond).

To understand which reaction mechanism leads to the observed selectivity, we performed a set of static simulations in a periodic H–Beta zeolite model. The experimental reaction environment is quite complex, as the reaction is performed in liquid toluene with a high O_2 pressure. Nonetheless, the size of toluene molecules is comparable with the H–Beta channels, meaning that the toluene-toluene interactions should have a limited effect. This was also proven by an additional set of MD simulations at reaction conditions we performed on either an H–Beta unit cell containing only one toluene molecule adsorbed on the Pd complex or an additional 4 toluene molecules, to fill the remaining pore volume. We saw that, independently on the loading, the toluene molecule adsorbed on the Pd complex interacts with it mainly through its *para* carbon, providing some initial insights into the zeolite role in tweaking the reaction selectivity.

NMR and EXAFS (extended X-ray absorption fine structure) measurements confirmed that the adsorbed Pd is most likely present as homogeneous-like isolated atoms. The hypothesis is that the BAS is transferred to one of the acetate ligands, which is released and substituted by the zeolite framework oxygens (step 0 in Figure 3.5).When studying the interaction of an adsorbed $Pd(OAc)_2$ complex with the pristine BAS through MD simulations, we indeed noticed that the BAS is in close interaction with one of the acetate ligands, weakening the original coordination bonds. After docking of Pd on the zeolite framework, we are left with isolated, single-atom active sites which make the reaction local in nature and, therefore, static calculations should be suited to derive reliable mechanistic features.

The first important difference between the reaction under study and a traditional cross coupling is that the initial Pd atom is in a +2 oxydation state. Moreover, toluene is not preactivated and, therefore, a direct oxydative insertion is unlikely. We then proposed that the residual acetate ligand of the chemisorbed Pd complex acts as a base and deprotonates one toluene molecule adsorbed on the Pd atom, with concerted formation of a Pd-C bond (step 1 in Figure 3.5). This step is well-known in literature for C-H activation and it is commonly referred to as concerted metallation-deprotonation (CMD)[186]. After the first CMD step, we tried to consider a second Pd-C bond formation (as in the regular cross-coupling mechanism). However, the lack of a free base to assist a second CMD requires the proton to be transferred to the zeolite framework. This was found to be highly endoergonic, making this path unrealistic. Another option would be a transmetallation between two Pd-R complexes, which are however suggested to be quite isolated based on NMR measurements. We then proposed that the second toluene molecule is directly attacked by the first one in a migratory insertion step (MI, step 2 in Figure 3.5)[102, 187]. The computed reaction barrier is quite low and corroborates this type of mechanism. Finally, a proton transfer reaction can occur to form an intermediate Pd hydride moiety, which is then reoxidized to the initial active species.

Concerning the reaction regioselectivity, theory predicts the CMD step to be highly *para* selective. Indeed, a free energy barrier of 91 $\text{kJ} \cdot \text{mol}^{-1}$ is predicted for the formation of the *para* product, while 101 and 103 $\text{kJ} \cdot \text{mol}^{-1}$ are required to form the *meta* and *ortho* products, respectively (Figure 3.5*b*). The selectivity of the MI step, on the other hand, is less defined. While the formation of the *ortho,para'* product is clearly unfavorable, the *meta* and *para* activations have analogous activation energies. This is not very surprising. Experimentally, indeed, a large predominance of products with at least one *para* tolyl group is observed (90%), in line with a highly selective CMD step (Figure 3.4). On the other hand, the *para,para'* vs. *meta,para'* selectivity is less prominent (60–70% and 20–30%, respectively). At the modest reaction

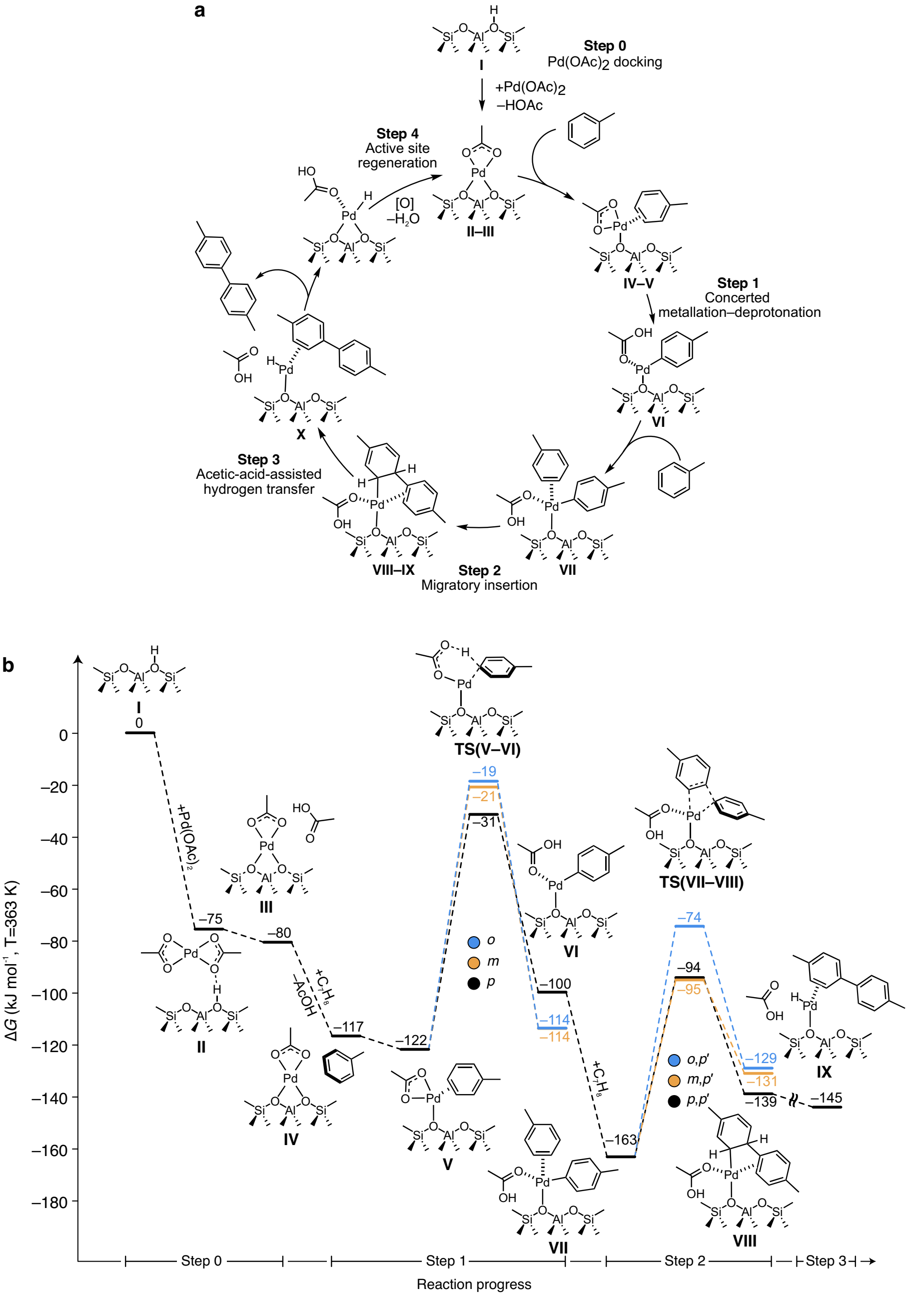


Figure 3.5: *a*. The proposed catalytic cycle for toluene homocoupling in Pd-loaded Beta zeolite. Its key features are an initial concerted metallation-deprotonation step followed by a low-activated migratory insertion. *b*. Free energy profile (B3LYP-D3//PBE-D3) for the first 3 steps of the cycle in *a*, where the reaction regioselectivity emerges. Adapted from Ref. 185, with permission from Springer Nature.

temperatures (90–150 °C) this would indicate a difference in activation energy for the MI step of $5\ \mathrm{kJ \cdot mol^{-1}}$, which is not significant when static calculations based on a single point on the PES are performed. Therefore, it has to be expected that our simulations could not perfectly capture the selectivity of the second coupling step. On the other hand, the proposed DFT profile for the homocoupling of toluene in Pd-loaded Beta zeolite provides some interesting insights on the most likely reaction mechanism (which is very different from the homogeneous case) and agrees reasonably well with the experimentally-observed selectivity.

Before concluding, it can be pointed out that the second part of the cycle, namely the formation of Pd hydride and its reoxidation with molecular oxygen, are not included in our periodic DFT study. This is due to the fact that molecular oxygen is used as oxidant and, therefore, the oxidation reaction requires a change in spin multiplicity (from triplet to singlet). To get an idea about the feasibility of the reoxidation step, Samuel E. Neale from our group performed additional non-periodic simulations based on cluster models. This was necessary as non-periodic codes (like `Gaussian 09` used here) have efficient implementations to find the minimum energy crossing point[188], i.e. the lowest point in energy where the triplet and singles surfaces cross. With this additional calculations we could show that the oxidation step does not present unreasonably high barriers and appears therefore to be a reasonable path for the reaction to follow.

3.3 Outlook

In this chapter, we saw how static simulations can be used to derive reactive trends capable of elucidating the most likely reaction mechanisms in a complex reaction network. Moreover, they can help to explain on which step(s) the experimentally-observed reactivity and selectivity depends. But if static calculations can have this type of predictive/explanatory power, one can ask the question whether it is really necessary to move towards more expensive methodologies. To answer this question, it is interesting to highlight the type of substrates investigated in the presented results. In both **Paper I** and **Paper II**, indeed, relatively hindered aromatic molecules are studied, with similar dimension as the zeolite pores. This is important since, as explained before, small molecules can undergo large anharmonic motions which are difficult to capture. On the other hand, the motion of a hindered aromatic ring will be mostly hampered by the framework atoms and error cancellation should allow for reasonable trends. Moreover, while not being the focus of this chapter, in both studies we complemented the static results with additional

MD simulations to study the behavior of reactants and intermediates at realistic operando conditions. In most cases, MD simulations can provide additional insights in the mobility of the adsorbed species, strengthening the results obtained from static simulations.

As an example of the shortcomings in static simulations, the free energy profiles of Figure 3.2 can be considered. While comparing similar quantities can be fine because of error cancellation (e.g. comparing the magnitude of the TS2 barrier across substrates), the overall free energies can be quite wrong. First, it can be seen that none of the substrates shows a negative adsorption free energy (+8, +3 and +14 $\mathrm{kJ \cdot mol^{-1}}$ for 4-EP, 4-*n*-PP and 4-*iso*-PP, respectively). While the temperature is of course high (653 K), it seems unlikely that the reactant adsorption in the catalyst could become disfavored. Here, the entropy penalty of adsorption is obviously overestimated due to the fact that not all the possible conformations that the molecule can assume in the zeolite are taken into account. Second, for all substrates the reaction free energy is also largely positive (**III** is always at higher free energies than **I**), which would mean that the reaction does not occur at 653 K in spite of that being the case for all substrates. The problem is, in this case, the formation of ethene. This product is very small and mobile in the catalyst and, consequently, the static approach strongly underestimates the entropic gain of the reaction.

Therefore, if one wants to obtain more reliable free energy differences for reactions involving highly mobile molecules, the clear choice is to move towards methods where the PES is explored taking more properly into account the temperature-induced motions of the atoms. This is one of the main purposes of (enhanced sampling) molecular dynamics, which is going to be the focus of the next chapter.

4

Enhanced sampling molecular dynamics to capture complex reactive environments

*After discussing the strengths and weaknesses of static simulations, we will stride towards a more realistic description of reactive events using molecular dynamics. This method allows to follow the atomic motions based on the underlying PES. As such, thermal effects can be included more properly. Unfortunately, crossing free energy barriers greater than few times $k_\mathrm{B}T$ is extremely unlikely within the accessible timescales of current simulation methods. Therefore, we will introduce enhanced sampling techniques, where a bias along important collective variables is used to force the reaction and derive free energy barriers and kinetic constants. This approach has been used to derive fundamental mechanistic insights in the ethylation of benzene in H–ZSM–5 (**Paper III**) and the O-demethylation of guaiacol in H–Beta with hot liquid water (**Paper IV**).*

4.1 Let the atoms dance: an overview of molecular dynamics

In Chapter 3, we discussed how DFT can be used to compute the forces acting on the system's atoms and perform, for instance, a geometry optimization. Here, we explain how the same forces can be used to integrate Newton's equations of motion and run an MD simulation. It is not the purpose to extensively review the topic, but rather to summarize the methods employed in this thesis—which are all part of the CP2K package (see Section 2.2). After the seminal work of Car and Parrinello[189], ab initio molecular dynamics soon became a very popular simulation technique[190]. In this thesis we make use of Born–Oppenheimer MD. As the name suggests, this type of MD relies on the Born–Oppenheimer approximation, where the nuclear and electronic degrees of freedom are assumed to be fully decoupled (i.e. any change in the nuclear coordinates is associated with an instantaneous adaption of the electron density).

A typical MD simulation starts by attributing to the atoms in the system velocities with a random orientation and a magnitude extracted from a standard Boltzmann's distribution at the given temperature. The sum of all velocities is restrained to be ~ 0 in order to avoid translations of the system's unit cell in space. The atomic coordinates are then propagated in time with finite time steps. This can be done with many algorithms, one of the most commons being the velocity Verlet:

$$\mathbf{r}\left(t+\Delta t\right)=\mathbf{r}\left(t\right)+\mathbf{v}\left(t\right)\Delta t+\frac{\mathbf{f}\left(t\right)}{2m}\Delta t^{2} \tag{4.1}$$

$$\mathbf{v}\left(t+\Delta t\right)=\mathbf{v}\left(t\right)+\frac{\mathbf{f}\left(t+\Delta t\right)+\mathbf{f}\left(t\right)}{2m}\Delta t \tag{4.2}$$

Where **r**, **v** and **f** are the atoms' positions, velocities and forces, respectively. Δt is the simulation time step, which has to be chosen. In our case, all systems contain H–X bonds (X = C, O), whose high stretching frequency requires a time step of about 0.5 fs.

With the MD algorithm described so far, one would sample the microcanonical (N,V,E) ensemble, meaning that the total energy of the system (the sum of the kinetic and potential energy) is preserved. To sample more interesting ensembles, like the isothermal (N,V,T) or isothermal-isobaric (N,P,T) ones, a thermostat and/or a barostat have to be used. In our simulations, we always adopted a Nosé–Hoover chain of thermostats[191, 192] for temperature control and a Martyna–Tobias–Klein barostat[193] for pressure control.

4.2 Enhanced sampling techniques to overcome reaction barriers

When sampling an isothermal ensemble with MD algorithms (the same applies of course to Monte Carlo[194] simulations as well), one explores states with canonical distribution. This means that the likelihood of observing a certain state decreases exponentially with the state energy. Therefore, high-energy transition states have a negligible probability of being sampled in the accessible time scale of a typical DFT MD simulation which is generally in the order of 10^1–10^2 ps (Figure 4.1*a*). In order to derive activation free energies associated with barriers larger than few times $k_\mathrm{B}T$, a typical solution consists in defining a bias that acts on (some of) the atoms in the system. This forces the latter to overcome the barrier and directs it towards another region of interest on the PES (typically the reaction products, see Figure 4.1*b*). Software-wise, in all enhanced sampling simulations presented in this thesis the bias has been applied by coupling the MD engine with PLUMED[195, 196]. Many different enhanced sampling techniques exist. For a more detailed overview, we refer the interested reader to specialized reviews on the topic[197]. Within this thesis, we primarily used metadynamics and umbrella sampling, which are briefly introduced hereafter.

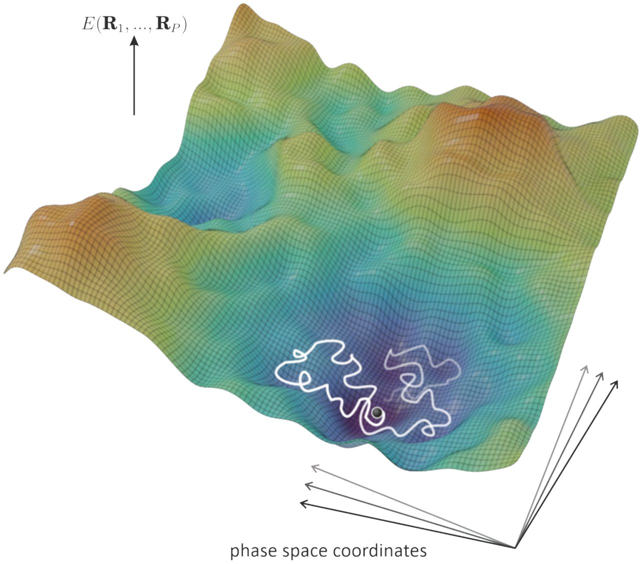
4.2.1 Collective variables

The first challenge one has to face when using enhanced sampling methods is to define a proper reaction coordinate along which the bias can be applied. Such reaction coordinate should be a function of the atomic coordinates and vary monotonically when the system evolves from reactants to products. The most common way to define a reaction coordinate is to use one or more collective variables (CVs), $Q\left(\mathbf{R}\right) = q$. While nowadays quite some research is being performed to find (semi)automated methods to define ‘good’ CVs for a certain reactive process[198], the most frequent selection criteria remains based on one’s chemical intuition. For instance, to study the bond-breaking reaction between two atoms a and b one could simply use the distance d_{ab} between them as CV. In almost all the work perform in this thesis, the CV of choice is normally a linear combination of coordination numbers (CNs). The CN between two groups of atoms A and B is defined as the switching function

$$\mathrm{CN}\left(\mathrm{A};\mathrm{B}\right) = \sum_{a\in\mathrm{A}}\sum_{b\in\mathrm{B}} \frac{1-\left(\frac{d_{ab}}{d_0}\right)^n}{1-\left(\frac{d_{ab}}{d_0}\right)^m}. \tag{4.3}$$

In general, n ranges from 4 to 6 and $m = 2n$. For most chemical reactions, the parameter d_0 should be chosen in order to approximately equal the length that a bond between two atoms a and b would take around transition state. In this way, the value of CN(A;B) is going to be roughly equal to the number of bonds between the atoms of group A and B.

a. Molecular dynamics



b. Enhanced sampling

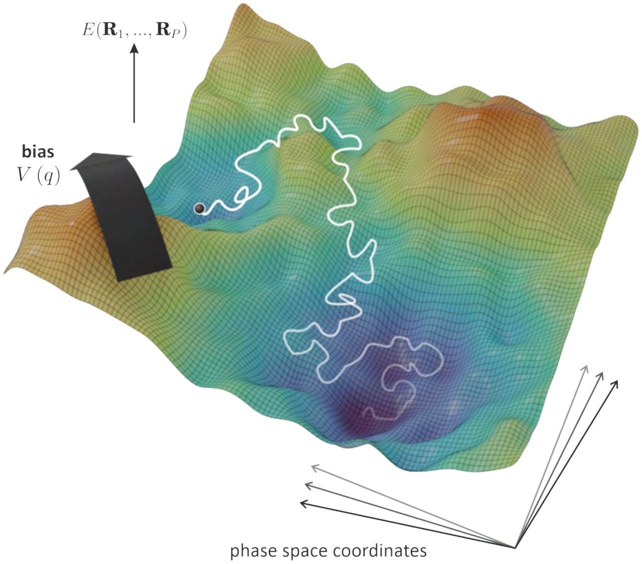


Figure 4.1: *a.* During a regular MD simulation, the system tends to visit low-energy regions of the PES. *b.* To overcome a large energy barrier, one has to apply an external bias to the system and force the transition.

4.2.2 Metadynamics

Metadynamics[199] was originally proposed by Laio and Parrinello. The idea behind this technique is quite clever: a gaussian hill of bias potential is spawned at the average CV value visited by the system during a regular time interval τ (Figure 4.2*b*). This creates an history-dependent bias, which is given by the hills sum:

$$V_b\left(q,t\right)=\sum_{i\tau<t}h\left(i\tau\right)e^{-\frac{\left(q-q_0\left(i\tau\right)\right)^2}{2\sigma^2}}. \tag{4.4}$$

The frequency of hills deposition (τ) and the hills width (σ) are user-determined parameters, while $q_0\left(i\tau\right)$ is the average value of q visited by the system in the time interval $\left[\left(i-1\right)\tau, i\tau\right)$. The simulation can normally be considered converged if, in the CV interval of interest, the system shows diffusive behavior or, in other words, if it freely goes back and forth throughout the CV range of interest.

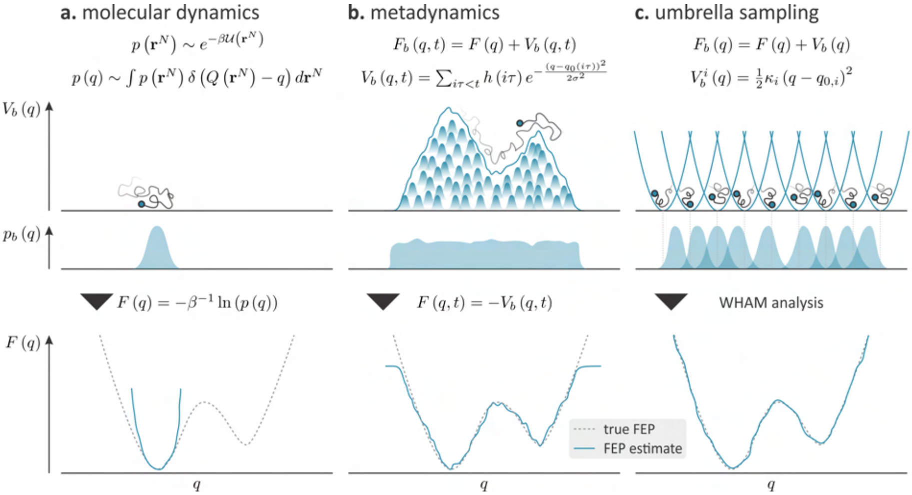


Figure 4.2: *a.* During an unbiased MD simulation, the free energy as function of a CV can be retrieved from the probability distribution of the explored CV values. *b.* Metadynamics adds history-dependent gaussian potential hills along the CV with the objective of flattening the observed probability distribution. The free energy is retrieved by reverting the final potential given by the sum of the hills. *c.* In umbrella sampling, multiple simulations are run in parallel within quadratic potentials distributed along the CV range of interest. The biased probability distributions are combined through a WHAM analysis to give a final free energy estimate.

Metadynamics is a powerful tool to explore a complex CV space. If one needs, for instance, 2 CVs to describe a reaction, the system is allowed to spontaneously explore the most favorable regions of the phase space, which is gradually filled by the bias hills. Thus one does not have to define the bias a priori like in 2-dimensional umbrella sampling (vide infra) and the most favorable reaction path in the 2-dimensional CV space is going to be naturally sampled. On the other hand, the technique also has a remarkable disadvantage: it is indeed very difficult to converge, as the addition of discrete potential hills means that one cannot possibly converge systematically towards the true free energy. Moreover, differently from umbrella sampling a single, long MD simulation has to explore the whole phase space. Luckily, there are—among many others[200]—two clever variants of metadynamics that allow to solve both issues.

In well-tempered metadynamics[201] the hills height is rescaled by a factor that is inversely proportional to the bias already present at the CV value where the hills is spawned:

$$V_b\left(q, i\tau+1\right) = h e^{-\frac{V_b(q,i\tau)}{\Delta T}} e^{-\frac{(q-q_0(i\tau))^2}{2\sigma^2}}. \tag{4.5}$$

In which the ΔT parameter controls the rate at which the hills height is decreased. Often this term is expressed as a bias factor $\gamma = \frac{T+\Delta T}{T}$. Note that in this case the bias does not converge to the free energy but to $-\frac{\Delta T}{T+\Delta T} F\left(q\right)$. Equation 4.5 indicates that in regions of the CV space that have not been explored yet high hills will be spawned, while in regions that have already been explored the spawned hills are going to be much smaller, accelerating convergence.

Finally, the simulation time can be drastically reduced using multiple-walkers metadynamics[202], where N metadynamics simulations are run in parallel and each walker feels a bias which is given by the sum of the hills spawned by all walkers. The well-tempered recipe can be easily combined with the multiple-walker method to quickly explore free energy landscapes, which is what we did in **Paper III** to explore the stability of Wheland intermediates during the ethylation of benzene in H–ZSM–5 and in **Paper IV** to produce an initial free energy estimate for the various reaction steps of the guaiacol *O*-demethylation reaction.

4.2.3 Umbrella sampling

In umbrella sampling[203, 204] (US) the reaction path along the CV is divided in windows (also called umbrellas). On each window, a potential is applied to restrain the system in the region of interest. In its most common formulation, the bias has a quadratic form characterized by a force constant κ and a center q_0 (Figure 4.2*c*):

$$V_b^i\left(q\right) = \frac{1}{2}\kappa_i\left(q - q_{0,i}\right)^2. \tag{4.6}$$

With a tight enough κ and regularly distributed umbrellas along the reaction path, one can sample every CV value of interest. As schematically depicted in

Figure 4.2c, the overall potential felt by the system will be given by $F_b(q) = F(q) + V_b(q)$ so that, if the underlying $F(q)$ is increasing, one will observe a biased distribution $p_b(q)$ shifted towards the left with respect to the umbrella center q_0. If the distributions in the various umbrellas are overlapping, the information on the underlying free energy for each of them can be combined through the weighted histogram analysis method (WHAM), obtaining the final reaction FES. There are of course other possible estimators to obtain free energies from biased simulations, as an example the multistate Bennett acceptance ratio (MBAR) can be cited[205].

Interestingly, there is nothing special about the choice of quadratic biases and, in principle, any functional form can be used. In **Paper IV**, for instance, we used US with the purpose of improving the convergence of a metadynamics-derived FEP[206]. In a nutshell, a well-tempered multiple-walkers metadynamics simulation is used to sample the reaction path, allowing to determine if the chosen CV is suited to describe the reactive event under study and to obtain a rough initial estimate of the reaction FEP. The latter is then inverted and used as additional bias in a subsequent US simulation, which is much easier to converge. The advantage is that the metadynamics-derived bias counteracts the underlying FEP, allowing to use wider umbrellas with a reduced risk of undersampling steep regions of the FEP. A summary of the idea is depicted in Figure 4.3

4.2.4 Retrieving accurate kinetic constants from enhanced sampling simulations

Up to now, we have seen how a free energy profile can be derived as function of a certain CV (q). One could then think that the reaction free energy barrier should be the difference between the lowest point in the reactant basin and the highest point in the TS region. While still being a common practice in literature[207, 208], this is however not the best way to tackle the problem. Indeed, as one might imagine, the same reactive process can be effectively described by many different CVs, for instance a distance, a coordination number, etc. Different CV will lead to very different FEP shapes and—as a consequence—to very different raw free energy barriers (Figure 4.4).

On the other hand, the final property that we would like to obtain is the kinetic constant of the reaction. Being a measurable property, it is clear that it should not depend at all on the arbitrary CV choice made to perform the enhanced sampling simulation.[209–212] Based on the theoretical foundations laid by Anderson, Bennett and Chandler 50 years ago[213–215], it has become apparent that the kinetic constant of a reaction can be derived from enhanced sampling simulations. In the past four years, I contributed to the testing

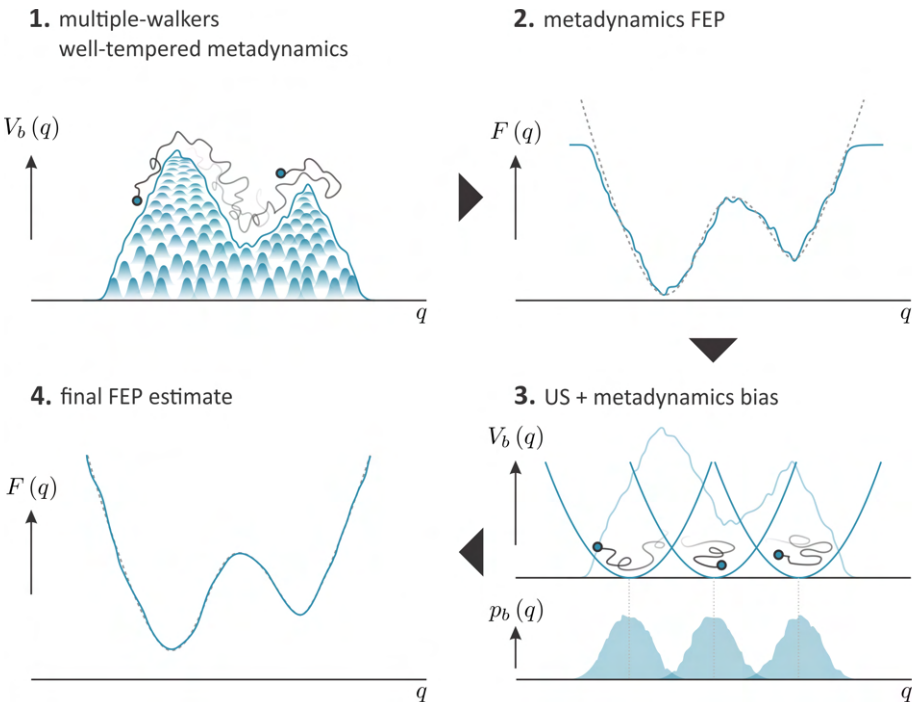


Figure 4.3: The free energy profile obtained from metadynamics can be used as additional bias in a subsequent US simulation to improve convergence of the final FEP estimate.

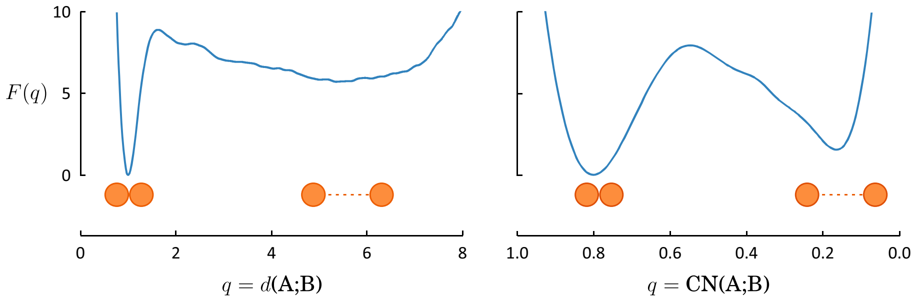


Figure 4.4: Free energy profiles obtained from metadynamics for the dissociation of two Lennard-Jones particles. On the left, the process is described using the distance between the two particles (d(A;B)); on the right, using the coordination number between the two particles (CN(A;B)). The obtained FEPs are clearly very different from each other. Adapted from Ref. 209, with the permission of AIP Publishing.

and usage of the `ThermoLib` library[216], mainly developed by Prof. L. Vanduyfhuys in our group. `ThermoLib` is a `Python`/`Cython` library written to analyze and manipulate FESs from enhanced sampling simulations. Aside from allowing to compute or read 1- and 2-dimensional FESs, it can be used to derive accurate reaction rates in the Bennett-Chandler transition state theory (TST) formalism and to convert a 1-D FEP $F\left(q_1\right)$ to a new collective variable $F\left(q_2\right)$ or to a new set of two CVs $F\left(q_3, q_4\right)$. The opposite operation is of course also possible, where a 2-D FES is projected on a single CV, function of the two original ones.

In the classical Bennet–Chandler approach to transition state theory (TST), the kinetic constant of a reaction can be written as[217] (Figure 4.5):

$$k = \langle \dot{q}(0)\theta\left(q(t) - q^*\right)\rangle_{q(0)=q^*} \frac{e^{-\beta F(q^*)}}{\int_R e^{-\beta F(q)} dq}. \tag{4.7}$$

In spite of its apparent difficulty, this equation simply shows that the rate of an activated event is inversely proportional to the height of the raw free energy barrier ($F\left(q^*\right)$) and to the likelihood of finding the system in the reactant basin ($\int_R e^{-\beta F(q)} dq$). On the other hand, it is directly proportional to the average 'speed' at which the system crosses the transition state region, provided that it ends up in the product basin. This is encoded by the heavyside function $\theta\left(q(t) - q^*\right)$, which is equal to 1 if $q(t) > q^*$ and 0 otherwise (bottom part of Figure 4.5).

The main problem in this formulation is computing the pre-exponential factor, i.e. the ensemble average $\langle \dot{q}(0)\theta\left(q(t) - q^*\right)\rangle_{q(0)=q^*}$. Indeed, to accurately determine the chances of barrier recrossing, one needs to explicitly initialize a very large number of trajectories atop the transition state with random initial velocities ($\dot{q}(0)$) and let them evolve in time to assess whether they end up or not in the products region. This approach is commonly known as **reactive flux formalism**[218]. On the other hand, an approximate solution can be found by applying one of the key approximations of TST, namely setting the chances of barrier recrossing to zero. In practice, this means that every trajectory with a positive initial velocity is assumed to end up in the product state. It can be shown that applying the approximation allows to rewrite eq. 4.7 as

$$k_{\mathrm{TST}} = \sqrt{\frac{1}{2\pi\beta}} \langle \left\|\nabla_r Q(\mathbf{r}^N)\right\| \rangle_{q^*} \frac{e^{-\beta F(q^*)}}{\int_R e^{-\beta F(q)} dq}, \tag{4.8}$$

where the difference is, as expected, in the initial factor. In practical implementations, an ensemble of TS structures is extracted from an US simulation in a small interval around q^*. From this ensemble, the mass-weighted derivative of the CV with respect to the atomic coordinates ($\nabla_r Q(\mathbf{r}^N)$) is computed to obtain the desired ensemble average. Remark that this formulation holds if the simulation is constrained around q^* (as in US)[210], but not if it is restrained

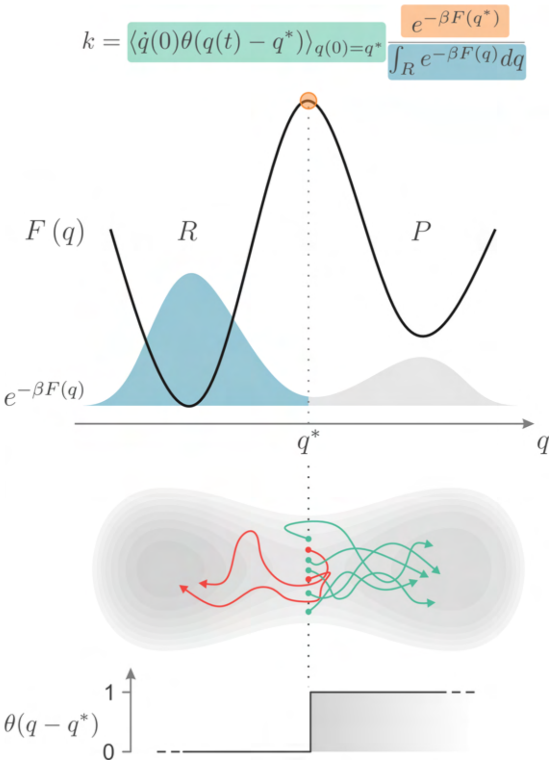


Figure 4.5: In the Bennet–Chandler approach to transition state theory, the kinetic constant of a reaction is inversely proportional to the height of the reaction barrier ($F(q^*)$), to the chance of finding the system in the reactant basin ($\int_R e^{-\beta F(q)}dq$) and directly proportional to the 'speed' at which the transition state region is crossed, provided that the reactive trajectory ends up in the product basin ($\langle \dot{q}(0)\theta(q(t)-q^*)\rangle_{q(0)=q^*}$).

on q^* (as in Blue–Moon sampling). In that case, a slight change in the formula to account for the restrain is needed[219]. It must also be pointed out that CV-free methods exist to compute reaction rates, i.e. methods that do not require an initial localization of the transition state region. Among them, one of the best-known is transition path sampling[220]. In spite of its undoubted elegance, this methods requires the generation of a reactive trajectories ensemble, which is still computationally too demanding for DFT with medium/large systems.

Working with kinetic constants can be quite cumbersome, as they normally span multiple orders of magnitude and they are not easy to interpret. In standard literature one often refers to free energy barriers, which are more 'chemically intuitive' (at least in the author's opinion) than a rate constant. This is because a free energy profile often allows to identify at a glance the steps that could be rate determining and the intermediates that could be the most stable. Therefore, we will often make use of a **phenomenological barrier**[219], which is simply obtained by plugging the kinetic constant into Eyring's classical TST equation:

$$\Delta F^{\ddagger}_{\text{phen}} = -\frac{1}{\beta}\ln\left(k\beta h\right). \tag{4.9}$$

Eq. 4.8 provides a straightforward method to derive reliable kinetic constant out of enhanced sampling MD simulations, which are no longer dependent on the choice of CV[210] provided that the latter is well suited to describe the reactive event under study[212]. Of course, to obtain converged results, it is important that the reaction path sampling is as complete as possible, which can become an hurdle when complex linear combinations of simpler CVs are used. In this case, it is interesting to use some statistical tool to present our original FES in function of new collective variables. For instance, if a difference in coordination numbers is used as CV it might be useful to generate a 2-dimensional FES expressed as function of the two single CNs, to ensure that the reaction path can indeed be easily described by their difference. Alternatively, one might be interested in understanding how a third coordination number evolves during the reaction and see how the free energy depends on this new CV. It is known that a 2-dimensional FESs $F(q_1, q_2)$ can be projected along a single CV to give a new 1-dimensional FEP $F(q)$ with the following equation[†]:

$$F(q) = -\beta^{-1}\ln\left(\int_{-\infty}^{+\infty}\delta\left(q - Q\left(q_1, q_2\right)\right)e^{-\beta F(q_1,q_2)}dq_1 dq_2\right). \tag{4.10}$$

Where we make use of Dirac's delta function δ. This formula holds if a deterministic relation exists between the two original CVs and the new one (i.e. $q = Q(q_1, q_2)$) but it can be made more general by the use of conditional probabilities. In the simple case where one wants to project the 2-D FES $F\left(q_1, q_2\right)$ onto either q_1 or q_2, eq. 4.10 reduces to:

$$F(q_1) = -\beta^{-1}\ln\left(\int_{-\infty}^{+\infty}e^{-\beta F(q_1,q_2)}dq_2\right). \tag{4.11}$$

What according to me is more surprising, is that the opposite operation is also possible, i.e. deprojecting a 1-dimensional FEP on 2 or more new CVs. In

---

[†] Since they do not represent the same object, $F(q)$ and $F(q_1, q_2)$ should be given different names, for instance $F_q(q)$ and $F_{12}(q_1, q_2)$. A similar argument holds for regular probabilities and conditional probabilities. To present the equations in a cleaner form, we will refrain from using all the appropriate subscripts and ask mathematicians to forgive this notation shortcut.

this case, the following formula is used:

$$F(q_1, q_2) = -\beta^{-1} \ln \left(\int_{-\infty}^{+\infty} p\left(q_1, q_2 | q\right) e^{-\beta F(q)} dq \right). \tag{4.12}$$

The key ingredient of this equation is the conditional probability $p\left(q_1, q_2|q\right)$, which encodes how likely it is to observe a certain couple of values (q_1, q_2) given a certain value of q. This is graphically depicted in Figure 4.6. As it can be seen, regions in the (q_1, q_2) space that have been visited often during the simulation will correspond to low free energy values (e.g. blue and yellow dots in Figure 4.6*b*). Of course, this information alone is not sufficient. Indeed, since the data comes from an enhanced sampling simulation, all regions along the biased reaction path should be visited with similar probability, (mostly) independent on their actual free energy. Therefore, the conditional probability is weighted by the free energy value at the original CV ($e^{-\beta F(q)}$, Figure 4.6*a*). Overall, the free energy in the new 2-D space $F\left(q_1, q_2\right)$ (Figure 4.6*c*) is going to be low if (q_1, q_2) are visited often and the corresponding q in the 1-dimensional profile is also low in free energy (e.g. blue dots in Figure 4.6). In the other cases, the final 2-D free energy is going to depend on the balance between the two factors. As a final note, it must be pointed out that expanding a 1-dimensional FEP $F\left(q_1\right)$ derived from US along a new CV to obtain $F\left(q_1, q_2\right)$ is conceptually analogous as performing a 2-dimensional WHAM analysis setting the force constant of the umbrellas along q_2 to zero.

4.3 Application to zeolite catalysis

Having introduced all the main concepts related to enhanced sampling techniques, we will now present some of the results obtained in **Paper III** and **Paper IV**. In particular, we will focus on the insights we were able to obtain on the role of water on the reactivity of zeolite-catalyzed reactions. As explained extensively in Section 1.4.1, water can fully solvate the BAS as hydronium ion (H_3O^+), drastically changing its behavior and reactivity. In **Paper III**, we showed how changing the water loading in H–ZSM–5 can drastically change the kinetics of protonation reactions during the ethylation of benzene with ethene and ethanol. In **Paper IV**, we considered a fully solvated environment and elucidated the reactivity differences between free and confined hydronium ions during the *O*-demethylation of guaiacol in hot pressurized water (HPW) with either HCl or H–Beta as catalysts.

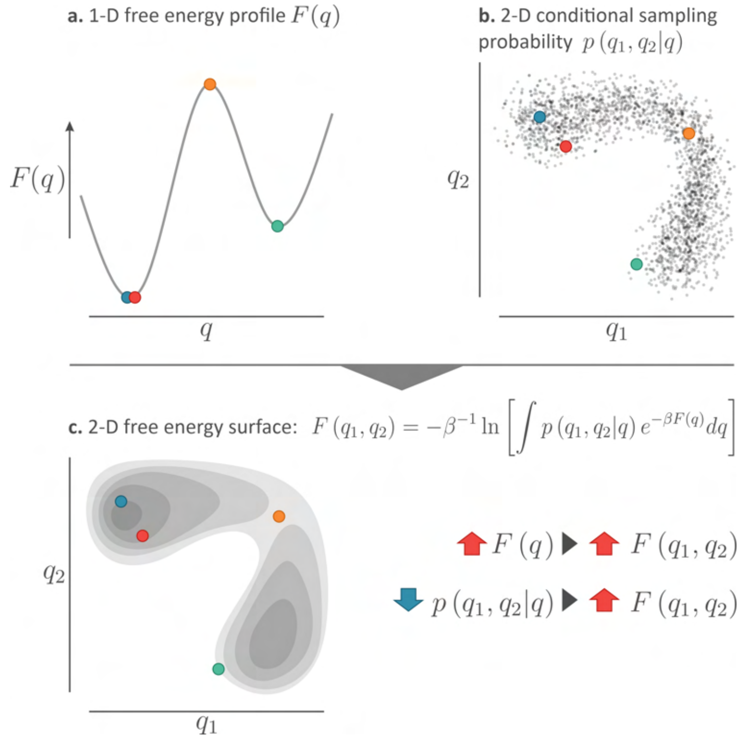


Figure 4.6: A 1-dimensional FEP (*a*) can be deprojected along a set of two new CVs based on the conditional probability of observing a couple of values (q_1, q_2) for a determined value of q (*b*). If the free energy $F(q)$ increases, then the corresponding $F(q_1, q_2)$ will also increase. At the same time, if the probability of sampling a couple of values (q_1, q_2) at the same q value decreases, $F(q_1, q_2)$ will again increase and vice-versa. The final 2-dimensional FES is shown in *c*.

4.3.1 Water modulates the kinetics of protonation reactions

In Section 1.3.1 we extensively discussed the importance of benzene ethylation and its mechanism. While in **Paper III** an extensive mechanistic investigation of the reaction is performed, we will here focus on the role of water in modulating the kinetics of protonation reactions. Indeed, the spectroscopic evidences from Chowdhury and co-workers[27] pointing to the existence of Wheland-type intermediates were obtained using ethanol as ethylating agent, where water is a reaction by-product. Moreover, it was shown that trace amount of water can affect the rate of H/D exchange for benzene molecules in H–ZSM–5[221] and water is, in general, known to influence the rate of key reaction steps in zeolite catalysis as pointed out in several theoretical investigations[222–225].

Therefore, we investigated the *para* protonation of ethylbenzene in the presence of 0, 1, 3 and 6 water molecules co-adsorbed in the zeolite unit cell (Figure 4.7). These coverages were chosen to represent a broad range of water partial pressures in the system[226]. The forward and backward phenomenological barriers for the reaction (eq. 4.9) are summarized in Table 4.1, while an overview of the raw FEPs at various water loadings is shown in Figure 4.8*a–c*. As it can be seen, the forward barrier (i.e. from neutral to protonated ethylbenzene) drops by more than 10 $\mathrm{kJ \cdot mol^{-1}}$ when going from 0 to 1 water molecule per unit cell. Interestingly, at higher loadings the barrier increases again, going back to similar values as the anhydrous case with 6 water molecules.

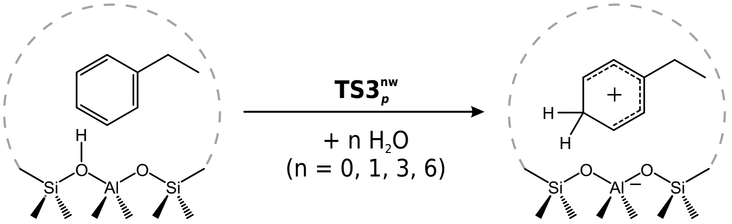


Figure 4.7: Schematic depiction of the ethylbenzene *para* protonation in the presence of various amount of water. Adapted with permission from Ref. 227. Copyright 2022, American Chemical Society.

Table 4.1: Forward and backward phenomenological barriers ($\mathrm{kJ \cdot mol^{-1}}$) at 573 K for the *para* protonation of ethylbenzene at various water loadings.

	water molecules per unit cell			
	0 H_2O	1 H_2O	3 H_2O	6 H_2O
$\Delta F^{\ddagger}_{\mathrm{forward}}$	76 ± 2	62 ± 3	68 ± 2	72 ± 2
$\Delta F^{\ddagger}_{\mathrm{backward}}$	18 ± 2	18 ± 3	11 ± 2	14 ± 2

To better understand the reasons underlying such peculiar behavior, we expanded the original FEPs in 2-D FESs (eq. 4.12), as shown in Figure 4.8*d–f*. One of the two new CVs represents the minimum distance between the zeolite oxygens in the first coordination sphere of the Al defect and the atom carrying the extra proton in the system. The latter is defined according to a combination of definitions proposed in the literature[228, 229] and a schematic depiction of some key values is shown in Figure 4.8*g*. The second new CV is the distance between the same proton atom and the ethylbenzene *para* carbon.

The resulting 2-D FESs, shown in Figure 4.8*d–f*, depict a clear and general picture concerning the influence of water on the kinetics of protonation reaction, which we believe is quite general and not restricted to the specific

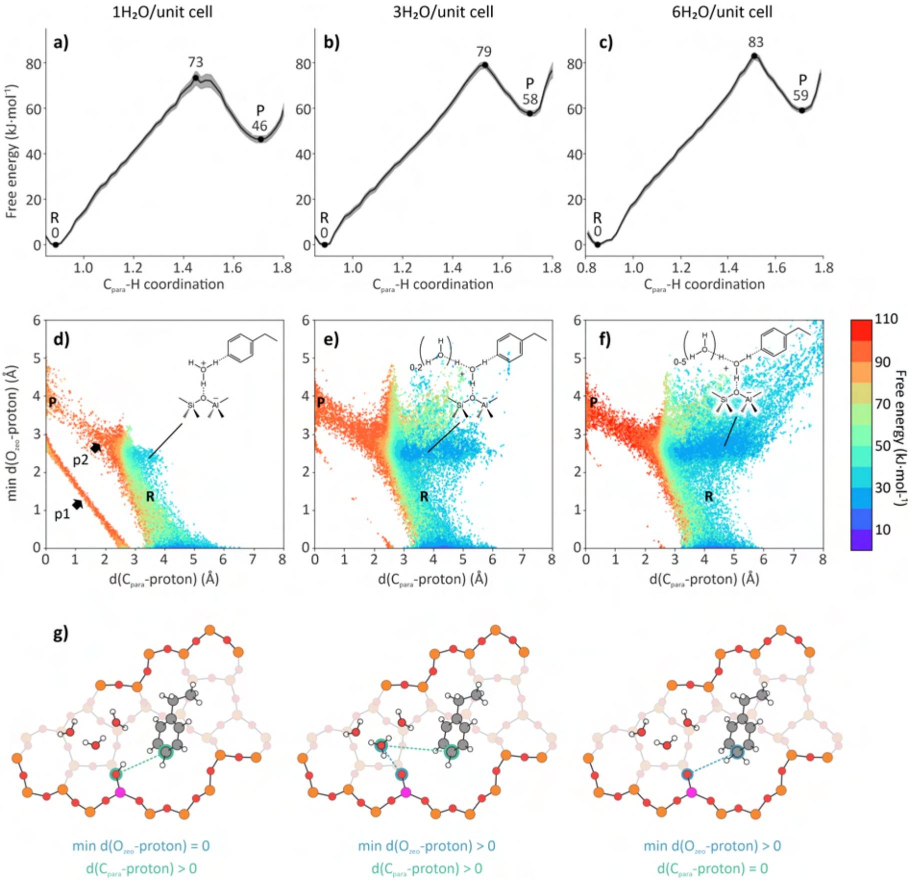


Figure 4.8: *a–c.* FEPs for the *para* protonation of ethylbenzene in the presence of 1, 3 and 6 water molecules in the zeolite unit cell. *d–f.* Expansion of the profiles along the distances between the atom carrying the excess proton and the zeolite framework oxygens and the *para* carbon, as graphically depicted in *g*. Adapted with permission from Ref. 227. Copyright 2022, American Chemical Society.

reaction under study. When 1 water molecule is present in the unit cell, it can actively take part in the protonation reaction and act as proton shuttling agent. Next to the direct proton transfer between the zeolite and the organic molecule (**p1** in Figure 4.8*d*), a brand new water-mediated path is opened (**p2**) with a subsequent acceleration of the reaction. When more water is introduced in the catalyst, on the other hand, full solvation of the BAS as hydronium ion becomes prevalent (see Section 1.4.1). The latter can diffuse relatively easily in the zeolite, as it can be seen from the large distances explored by the proton atom in Figure 4.8*e* and *f*. Given this plethora of new accessible states, it becomes more unlikely for the hydronium ion to find itself in proximity of the ethylbenzene molecule, resulting in a slower protonation rate compared with the lowest coverage of 1 molecule per unit cell.

Overall, these results provided unprecedented atomistic insights on the role of water as proton-transferring agent. While it is suggested that at very low coverages a beneficial effect on the rate of protonation reactions is present, it is also known that water adsorbs non-homogeneously on the zeolite BASs[83]. Experimental measurements suggest indeed that the adsorption of more than one molecule per BAS is associated with a larger heat of adsorption[226] because of the formation of hydronium ions. Therefore, the rate enhancement derived from our simulations might not be measurable experimentally, unless extremely low water coverages would be used.

4.3.2 On the enhanced activity of confined hydronium ions

In the introduction, we explained how modern catalytic technologies have allowed to depolymerize lignin in a mixture of few alkylguaiacols and alkylpirogallols (Section 1.3.2). We also stated that the relevance of these compounds in the market is somehow limited, and they should be ideally further defunctionalized to obtain simpler drop-in molecules for the chemical companies. **Paper IV** is located in this context, where we investigated the *O*-demethylation of guaiacol in HPW using either an homogeneous (HCl) or heterogeneous (H–Beta) Brønsted acid. Both catalysts have been shown to be effective in catalyzing the reaction by our experimental partners in the University of Antwerp (Prof. B. U. W. Maes) and in the KU Leuven (Prof. B. F. Sels)[47, 48]. In **Paper IV**, we performed a combined theoretical/experimental investigation to understand the reaction mechanism and the reactivity differences between hydronium ions in bulk water and in a confined environment. The theoretical investigation of the reaction in homogeneous phase was performed by Elias Van den Broeck (CMM), who also had the idea of comparing it with the heterogeneous catalyst.

It is nowadays well-known that the microporous environment of zeolites has a deep impact on the properties of water molecules, that when exposed to the hydrophobic pore walls behave in a surface-like manner[230]. This is quite important for the catalytic activity of hydronium ions that, as pointed out by seminar studies from Lercher and co-workers, can become significantly more active than in homogeneous phase[49, 50]. Subsequent studies performed on cyclohexanol dehydration pointed out that the charge separation between water clusters where the H_3O^+ ion is located and the framework aluminum creates a strong ionic environment. This will affect the energy of transition states and the adsorption properties of reaction intermediates differently, since the reaction proceeds through an ionic E_1 mechanism[51].

What we were able to show in our study, is that the rate-enhancement effect is not exclusively related to the ionic environment, but persists also for reactions that do not go through ionic intermediates. Indeed, our enhanced sampling based investigation of the reaction both in homo- and heterogeneous phase (Figure 4.9*b,c*) showed that the *O*-demethylation of guaiacol preferentially proceeds through a direct S_N2 mechanism (Figure 4.9*a*). Differently from previous reports in literature[44, 45], we did not observe the formation of a (meta)stable oxonium ion intermediate. This can be seen in the 2-dimensional FESs of the S_N2 reaction shown in Figure 4.10. There, q_2 follows the protonation state of the guaiacol methoxy oxygen while q_1 the actual S_N2 reaction. If the oxonium ion would be a stable intermediate, it would appear as a minimum on the FES in the top left corner, at the onset of the S_N2 reaction. What we did observe, on the other hand, is the formation of an ionic contact pair, where the proton remains somehow shared between the guaiacol oxygen and the hydronium ion. It is important to notice, however, that the S_N2 reaction does not proceed until this contact pair is formed and therefore the protonation step is not part of the rate limiting TS. This was confirmed experimentally by measuring the reaction kinetic isotope effect (KIE), which resulted to be strongly inverse (~ 0.6) when going from H_2O to D_2O, incompatible with a rate-determining step involving hydrogen atoms.

While the reaction features described thus far are shared by both the homo- and heterogeneous systems, there are of course some key differences. The most important one is that the reaction proceeds remarkably faster in the zeolite. This was measured experimentally and also emerged clearly from our simulations, where the phenomenological barrier of the S_N2 step decreases from 111 $kJ \cdot mol^{-1}$ in the homogeneous case to 98 $kJ \cdot mol^{-1}$ in the heterogeneous one. To explain this interesting observation, we focused on the hydronium ion state in the two environments. We found that the restrained environment of the zeolite does not allow for a full solvation of H_3O^+. Indeed, the average number of water molecules within a 3.6 Å radius decreases from ~ 6 in the homogeneous case to ~ 4 in the heterogeneous one. This, according to our calculations, enhances the electrophilicity of H_3O^+. The average Mulliken's charge on the hydronium ion oxygen increases from -0.379 to -0.357 when going from bulk to the restrained environment.

The enhanced electrophilicity has a strongly beneficial effect concerning the association between the guaiacol substrate and the hydronium ion. To show this, we report in Figure 4.11 a set of free energy profiles showing the preferred protonation state of the closest water molecule to the methoxy oxygen of guaiacol (quantified as coordination number between the water oxygen and all hydrogen atoms in the system). The bottom profile in both cases comes from simulations performed in proximity of the reactant state

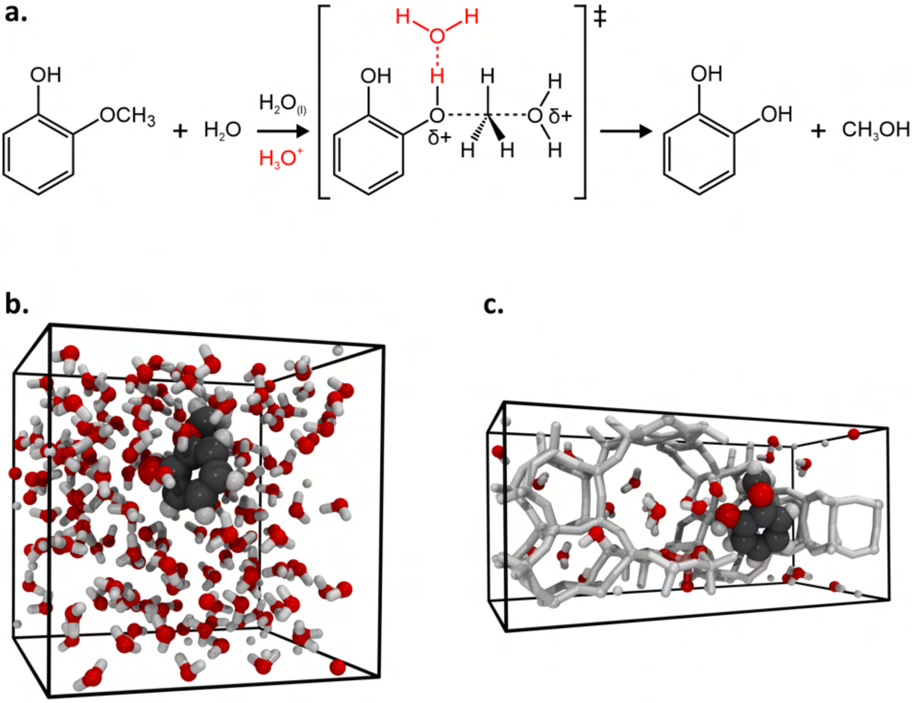


Figure 4.9: *a.* Schematic depiction of the acid-catalyzed S$_{\text{N}}$2 reaction of guaiacol with water. In **Paper IV**, more pathways were investigated but the S$_{\text{N}}$2 mechanism was found to be the lowest activated one. *b.* Unit cell of used to simulate the homogeneous *O*-demethylation of guaiacol, containing 170 water molecules. *c.* H-Beta unit cell with one Al substitution, one guaiacol and 22 water molecules. The framework atoms are in white for the sake of clarity.

($q_1 < -0.2$, $q_2 \sim 0.2$ in Figure 4.10). In the homogeneous case (Figure 4.11*a*), guaiacol is preferentially in interaction with a neutral water molecule. Only when forcing the reaction towards the S$_{\text{N}}$2 transition state H_3O^+ is forced to interact with guaiacol, gradually becoming a more prevalent state. At the onset of the S$_{\text{N}}$2 reaction (top profile, $q_1 < -0.2$, $q_2 \sim 0.8$ in Figure 4.10), it can be seen how the protonation state lies in between a neutral H_2O and a H_3O^+ ion, in line with the formation of the ionic contact pair described before. For the heterogeneous case, on the other hand, the interaction between guaiacol and H_3O^+ is already predominant in proximity of the reactant basin, indicating that there are no entropic factors disfavoring the association between the two species. When biasing the system towards the S$_{\text{N}}$2 transition state, H_3O^+ becomes fundamentally the only observed species in interaction with guaiacol. Finally, at the onset of the S$_{\text{N}}$2 step, the systems ends up in the ionic contact pair as in the homogeneous case.

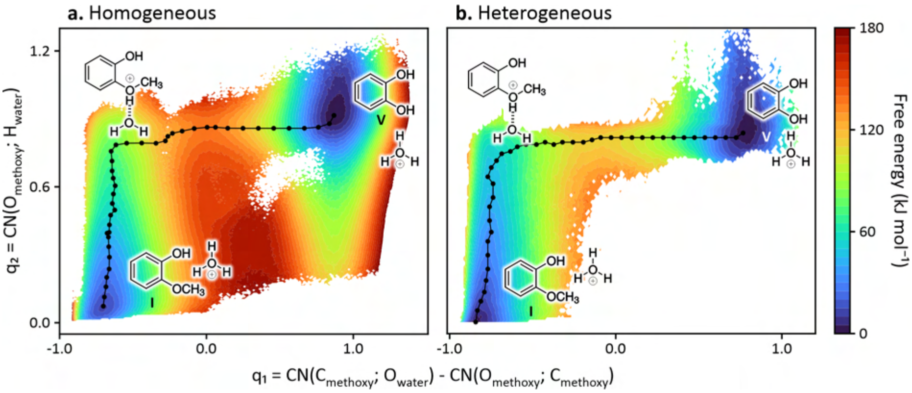


Figure 4.10: 2-dimensional free energy surfaces for the S_N2 reaction of guaiacol with water in homogeneous (*a*) and heterogeneous (*b*) phase. q_1 encodes the progression of the S_N2 reaction while q_2 the protonation state of the guaiacol methoxy oxygen.

We propose therefore that this enhanced association between guaiacol and the hydronium ion contributes to the increased rate observed for the heterogeneously-catalyzed reaction. This insights build atop previous seminal works in the field of confined hydronium ions (as explained at the beginning of the paragraph) and, hopefully, will be a step forward towards a better understanding of the fascinating—yet very complex—chemistry behind them.

4.4 Outlook

In this chapter, we introduced the concept of MD and enhanced sampling simulations for the study of rare events. We then showcased how this type of simulations can be used to derive chemical insights for zeolite-catalyzed reactions in complex reactive environments, in particular in the presence of co-adsorbed water. Elucidating the role of water in zeolite-catalyzed reactions has been one of the main goals in my thesis, going from its impact on diffusion of oxygenated compounds[183], to its role as proton-transferring agent[227] and finishing with reactions conducted in bulk HPW.

I believe the behavior of water is a key case study to highlight the power of enhanced sampling methods. Indeed, the presence of so many small molecules all in interaction with each other creates a very rough PES, with countless minima separated by negligible barriers. A single point on this landscape is clearly insufficient to capture a good statistical picture of the system and only with enhanced sampling one can hope to explore a sufficient portion of it. Even worse is the case of proton transfer reactions, were the hydronium

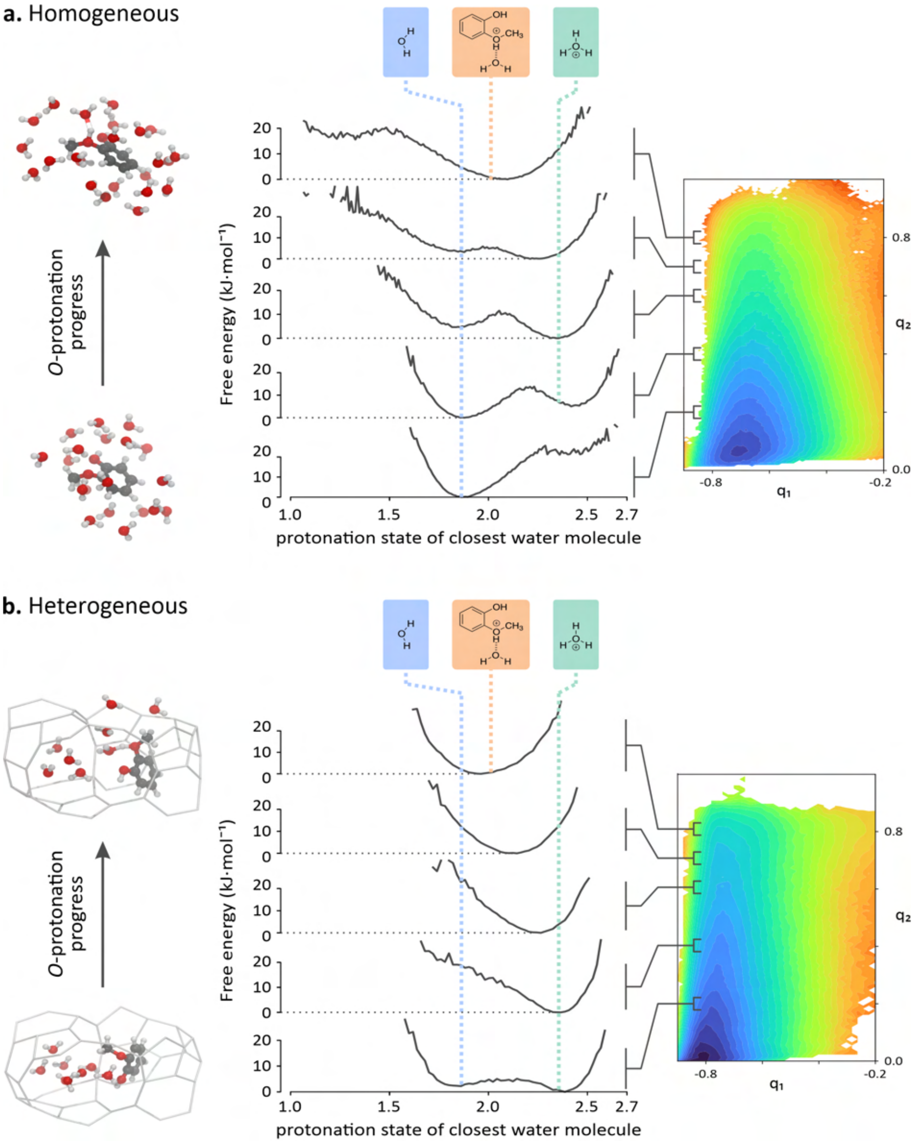


Figure 4.11: Free energy profiles showing the preferred protonation state of the water molecule closest to the methoxy O atom, for the homogeneous (*a*) and heterogeneous (*b*) systems. On the right, an extract of the 2-D FESs from Figure 4.10 is shown to highlight the regions from which the states are taken.

ions can quickly diffuse through Grottuß mechanism and dynamically jump from water to water. This behavior would also be completely impossible to capture with only few points on the PES.

In spite of this, enhanced sampling simulations based on first-principles methodologies still suffer from a central problem, i.e. their massive computational cost. While obtaining a free energy profile from static calculations can require at most few thousands of single-point energy and forces evaluations (including a full optimization followed by a frequency calculation), a typical umbrella sampling simulation requires in general hundreds of thousands of steps per umbrella to obtain well-converged results. With a typical time of about 10 seconds per step on modern supercomputing infrastructures (this number will obviously vary quite a bit based on the level of theory and on the system size) it can take months to finalize an US simulation for a single reactive step. Moreover, while enhanced sampling allows to describe activated events, a proper description of the adsorption step would also be necessary to build microkinetic models and compare the theoretical rates with the experiment. In theory, accurate results could be obtained with hybrid procedures based on MD and grand canonical Monte Carlo. Unfortunately, also in this case we are limited by the cost of DFT simulations, that makes this kind of simulation completely unfeasible. While there exists procedures to estimate accurate adsorption energies (on this topic a lot of work has been done by Sauer's group, see for instance Ref. 231 and references therein) we are still far from obtaining accurate adsorption isotherms in zeolites with first-principles accuracy.

This brings us to the final chapter of the thesis, where we introduce the concept of machine learning potentials (MLPs). MLPs can be trained on a limited set of first-principles data points to learn the underlying PES. In this way, they can be used to perform any kind of simulation with a speedup of many orders of magnitude, while retaining the accuracy of the calculations used for training. As we will see, MLPs are very promising but still need a lot of fundamental research to be applied on the complex systems tackled thus far in this thesis.

5

Beyond DFT: machine learning potentials for zeolite catalysis

While MD provides a straightforward way to explore the DFT PES, these simulations are also extremely inefficient. Indeed, the system often oscillates around a minimum, exploring similar conformations times and times again. It would therefore be ideal to have a fast algorithm capable of learning the DFT PES from a handful of single-point calculations and then use it to perform the MD simulations. This is the idea behind the recently exploding field of machine learning potentials (MLPs), where the energy of an atom is learned based on its environment. Very efficient architectures can be trained on a limited set of DFT single points and perform MD simulations with a speedup of many orders of magnitude. After introducing the clever way in which the DFT PES can be learned, we will see how the MLP can be combined with enhanced sampling MD simulations to compute the rate constant of the proton hopping reaction in zeolites with inclusion of nuclear quantum effects (NQEs), as extensively reported in ***Paper V****.*

5.1 Machine learning in a nutshell

5.1.1 A perfectly reasonable introduction to this chapter

"In this chapter, we will investigate the application of neural network potentials in modeling zeolite catalysis. We will evaluate their effectiveness in representing the intricate interactions between zeolites and catalytic species, and assess their potential to enhance the predictive abilities of computational models in the field of zeolite catalysis. By utilizing these potentials, we aim to gain a more comprehensive understanding of the mechanisms and factors influencing zeolite catalytic activity."

In spite of its apparent human-level quality, the paragraph above has been fully generated by the `ChatGPT` machine learning model from OpenAI[232] as an answer to the question "*write a short introduction (max 4 sentences) for a chapter of a PhD thesis concerning neural network potentials applied to zeolite catalysis. Assume that the audience is familiar with the concept of zeolites*". This, in my opinion, showcases the astonishing degree of complexity that machine learning algorithms have reached nowadays. Aside from producing human-grade sentences and dialogues, `ChatGPT` can also produce ready-to-use code from a set of human-readable instructions or propose a plausible plot for the next inevitable Star Wars movie. In the following section, we will explore the basic principles underlying machine learning methods and, subsequently, how these can be adapted to be applied in the field of molecular modeling.

5.1.2 How can a machine learn?

Suppose a farmer would have the following issue: wine producing companies will buy his grapes only if the sugar content is higher than a certain threshold. Unfortunately, sending a grape from each plant to the lab for testing is very expensive, and the farmer would therefore like a cheaper way to decide if the grapes of a certain plant can be sold or not. According to his experience, the sugar content depends mostly on the hours of sun a certain plant gets during the summer and on the amount of water it receives. Luckily, he diligently kept track of the grapes that he could and could not sell in the previous years as function of the two aforementioned variables. You then try to plot them with different colors and obtain the results shown in Figure 5.1*a*, where the x axis are the hours of sun a plant is exposed to during a summer day and the y axis the liters of water it has received from the farmer.

Dividing the data is in this case very straightforward, as a simple line $w_1x + w_2y + b = 0$ is an effective decision boundary. We will refer to the coefficients that multiply the variables as **weights** (w_i) and to the additive constant as **bias** (b). If these three parameters are properly tweaked, a line as the one shown in Figure 5.1*a* can be obtained. Thus, you can report back to the farmer and tell him that next year he can plug his two variables in the equation $w_1x + w_2y + b$; if the result is greater than 0 he can sell the grapes of the plant, otherwise he better keep them to be eaten (ideally, you could also provide him with some uncertainty metrics for results that are very close to zero).

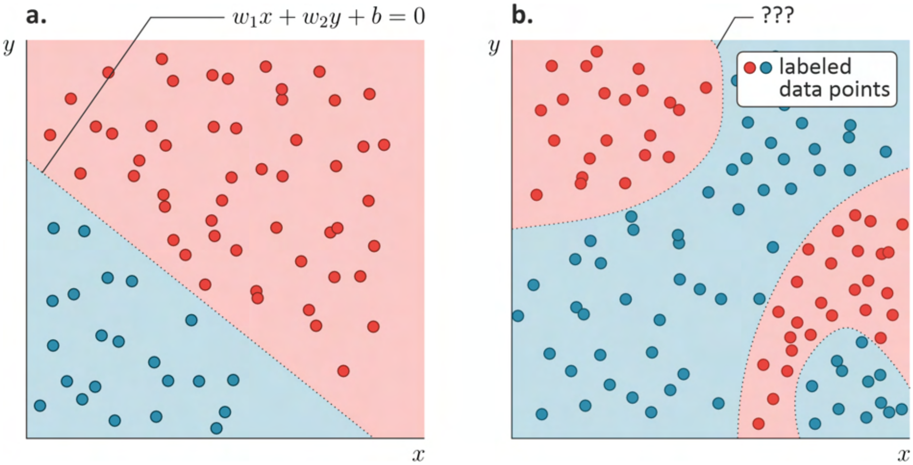


Figure 5.1: *a.* 2-dimensional data points can be classified in two groups usign a simple linear function. *b.* If the distribution of the points becomes more complex, linear models are insufficient to capture the system complexity.

The equation parameters can of course be tweaked manually, but ideally a **loss function** should be defined to quantify how bad the model is performing. In this case, labeled points that are not in the right region could contribute positively to this function while a well-classified point with 0. In this way, finding the best parameters is seen as minimizing the loss function—which is an important step given the plethora of computational methods to perform functions minimization. In the broad sense, this could be seen as an example of **machine learning**, where a set of labeled data is used to train some function capable of labeling new data points that have not been seen yet. Thus, in a sense, we can make a machine learn the classification process. Of course, this was a trivial example and most data distributions in the real world might look more like Figure 5.1*b*, where it is clear that many more parameters and non-linear functions should be used to achieve an accurate classification of the data.

It might be surprising, but the working principles of **machine learning potentials** for molecular simulations (also known as machine learning force fields) are actually analogous to the farmer example[233]. As the farmer was looking for a cheaper way to determine the quality of grapes without needing expensive lab analysis, we are looking for a cheaper way to compute the energy and the forces of a molecular system without needing all the times an expensive DFT calculation. While in the example the input parameters were the sun hours and the liters of water, for a molecular system the input is going to depend on the atomic coordinates. Hereafter, the working principles of the `SchNet` neural network potential[234, 235] will be resumed, being the one we extensively employed in **Paper V** and, at the end, we will quickly mention the most recent improvements to the architecture that have been proposed to increase its efficiency.

5.2 Neural networks for molecular simulations

`SchNet` is an invariant deep learning architecture that makes use of continuous-filter convolutional layers. The idea that lies at the core of most modern MLPs architectures, traceable back to Behler and Parrinello[236], is that the energy of the system is expressed as a sum of atom-wise contributions which only depend on the atomic environment up to a certain cutoff radius.

$$E_{\text{tot}} = \sum_{\alpha=0}^{P} E_{\alpha} \,. \tag{5.1}$$

This idea stems from the fact that the energy of an atom is strongly dependent on its local environment (one can think about the difference between an sp^2 and an sp^3 carbon) and all interactions tend to decay relatively rapidly with distance. Such principle offers immense advantages. The key one is that energy becomes a size-intensive quantity, i.e. atoms with similar environments should return similar energies independently on the system size. A consequence of this is that the MLP can be highly transferable. For instance, silicon atoms in two zeolite frameworks with similar SBUs will experience relatively similar environments and, therefore, training on one could be sufficient to describe the second as well. Moreover, the MLP can be used with any kind of periodicity.

Through the layers l of a typical neural network, the atoms are described by a tuple of features $X^l = \left(\mathbf{x}_1^l, ..., \mathbf{x}_n^l\right)$. The features of one layer $\mathbf{x}_i^l$ are propagated to the following layer ($\mathbf{x}_i^{l+1}$) by multiplying them by a weights matrix W^l and then adding the biases $\mathbf{b}^l$. Similarly to our previous example, these are the parameters that will be optimized.

$$\mathbf{x}_i^{l+1} = W^l \mathbf{x}_i^l + \mathbf{b}^l \,. \tag{5.2}$$

The number of features and the number of layers are two hyperparameters that can be adjusted by the user based on the complexity of the problem on hand. In general, the output of a layer is still passed through an activation function that rescales all values between 0 and 1. This function can be for instance a sigmoid and it is useful to introduce non-linear behaviors in the network output (which is necessary for an example like the one presented in Figure 5.1*b*). Using an activation function also strengthen the connection between neural networks and human brains. Indeed, we will now have mostly 'active' neurons (i.e. features that are close to 1) or 'inactive' ones (features close to 0), which is in crude approximation a similar behavior as human neurons.

Clearly, if one would only have layers as the one shown in eq. 5.2, the atoms would have no idea about their surroundings. Therefore, `SchNet` is built based on the principles of message passing neural networks (MPNNs), which are ideally suited to treat graph-like objects. A graph is a mathematical entity made of nodes with a specific connectivity pattern. In principle, every molecule can be seen as a graph in which the atoms are the nodes and the bonds between them the connections. Remark that, in the field of MPNN, the connections between atoms are not related to bonds in the chemical sense of the word. Two atoms are considered connected simply if they are within a certain cutoff distance from each other, which is defined by the user and that will affect the final accuracy and cost of the MLP (the concept is visually portrayed in Figure 5.2).

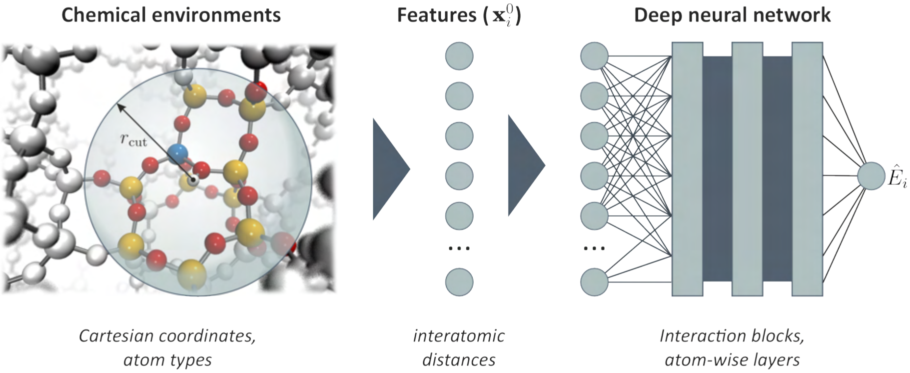


Figure 5.2: In MLPs based on deep neural networks, like `SchNet`, the atoms are described by a set of features that depend on their chemical environment up to a certain cutoff radius. In the example on the left, the chemical environment up to r_{cut} of the hydrogen atom in the H–CHA framework would include all the colored Si (yellow), Al (blue) and O (red) atoms.

Without diving into the mathematical details, in some of the network layers (how many is an hyperparameter that can be tweaked by the user) the atomic features are convoluted (i.e. they are forced to interact) with the features of all the surrounding atoms included in the cutoff sphere (Figure 5.3*a*). This interaction is modulated by a filter matrix, which is generated by a neural network based on the distance between the atoms (in this way the network is invariant with respect to rotations and translations of the molecular system). The neural network that generates the filter matrix is of course also optimized during training. Interestingly, every time we apply a convolution the interaction sphere of our atom is expanded. Indeed, after the first convolution, the atomic features will depend on the surrounding atoms up to the cutoff radius. When a second convolution is applied, we are making the features of our atom (which now contain information about its surrounding) interact again with the features of its surrounding atoms (which also contain information about their surroundings). Therefore, after this second convolution, the features of our atom will no longer depend only on its surrounding, but also on the surroundings of the surrounding atoms (I apologize for the number of repetitions, the concept is hopefully more clearly displayed in a schematic manner in Figure 5.3). In practice, this means that every convolution expands our actual interaction sphere beyond the initial cutoff radius.

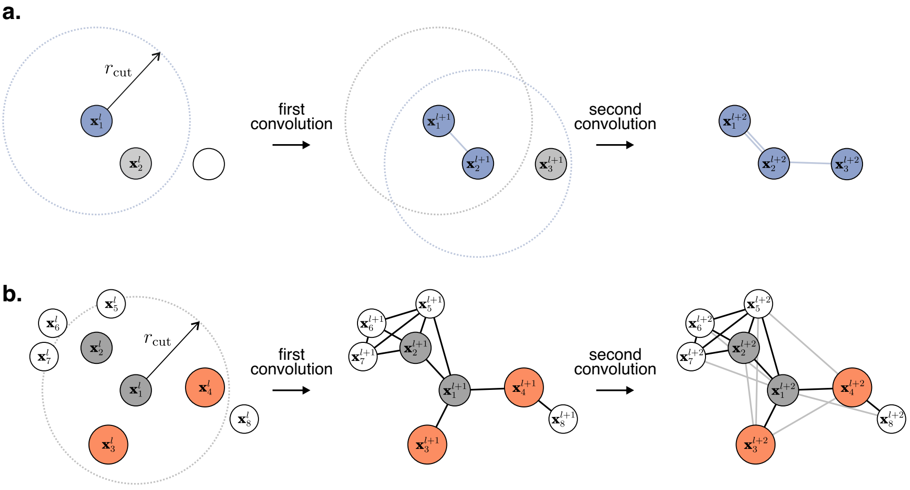


Figure 5.3: *a*. In a message passing neural network, the convolutional layers allow the features describing an atom to interact with each other. Here, atom 1 is described by the features $\mathbf{x}_1^l$ at layer l and, after the first convolution, it will have interacted with the features of atom 2 ($\mathbf{x}_2^l$), being now described by a new set of features $\mathbf{x}_1^{l+1}$. If a second convolutional layer is applied, the two features will interact again. This time, however, the features of atom 2 will have interacted with the ones of atom 3 as well during the first convolution. This has the practical outcome of atom 1 'seeing' atom 3 via the exchange of information with atom 2. **b**. The same concept outlined in *a*, applied to a whole molecule. After each layer, the interactions between atoms are propagated and each one feels atoms well beyond its initial cutoff radius.

To train the network, a set of DFT simulations has to be performed for the system of interest in order to obtain reference energy and forces. The loss function that has to be minimized is expressed as follows:

$$\ell\left((\hat{E}, \hat{\mathbf{F}}_1, ..., \hat{\mathbf{F}}_n), (E, \mathbf{F}_1, ..., \mathbf{F}_n)\right) = \\ = \rho \| E - \hat{E} \|^2 + \frac{1}{n} \sum_{i=0}^{n} \left\| \mathbf{F}_i - \left(-\frac{\partial \hat{E}}{\partial \mathbf{R}_i} \right) \right\|^2 \tag{5.3}$$

Where the hat denotes quantities predicted by the MLP and ρ is a weighting factor that can be used to give more importance to either the energy or the forces. For most MD applications we found that increasing the weight given to the forces is beneficial, given that the latter are used in the velocity-Verlet algorithm (Equations 4.1 and 4.2). The available data points have to be divided in a training and validation set (most often a 8:2 proportion is a good starting point). The former are used to minimize the loss function (each minimization step is usually referred to as an **epoch**) while the latter are evaluated to ensure that the error is decreasing for structures the MLP is not being trained on. Indeed, if the error on the validation set would increase, it would mean that the large amount of parameters in the network is allowing it to over fit on the available training data. The issue can be mitigated by increasing the chemical diversity of the training data and/or by reducing the size of the network.

5.3 Computing the heck out of an MLP PES

With a basic understanding of the working MLP principles, we will now summarize the main results of **Paper V**. There, a `SchNet` MLP was trained based on high-temperature US simulations of the proton hopping reaction in the H–CHA zeolite (Figure 5.4*a*). We will see how having access to a computationally cheap estimator of the DFT PES allows to perform a whole set of simulations that are normally unfeasible for the high cost of first-principles simulations.

In Brønsted-acidic silicoaluminate zeolites it is well established that the BAS is not restricted to a specific oxygen in the first coordination sphere of the Al defect, but it rather hops from one oxygen to another with a moderate free energy barrier[55]. This very simple activated event is more interesting than it appears. Indeed, it can be regarded as the archetypal example of proton transfer reaction and, as such, it can provide general indications about the zeolite propensity to transfer its proton to an adsorbed organic substrate. For our proof-of-concept study, we focused on the Brønsted acidic CHA topology (H–CHA, the material is also often denoted H–SSZ–13). This framework offers a series of advantages as (i) the unit cell is quite small (36 T atoms), allowing for relatively fast DFT evaluations and (ii) all T sites are equivalent by symmetry, eliminating the problem of multiple possible Al locations. On the other hand, the 4 oxygen atoms surrounding the Al are not equivalent and, therefore, 6 hopping paths are possible (Figure 5.4*b*).

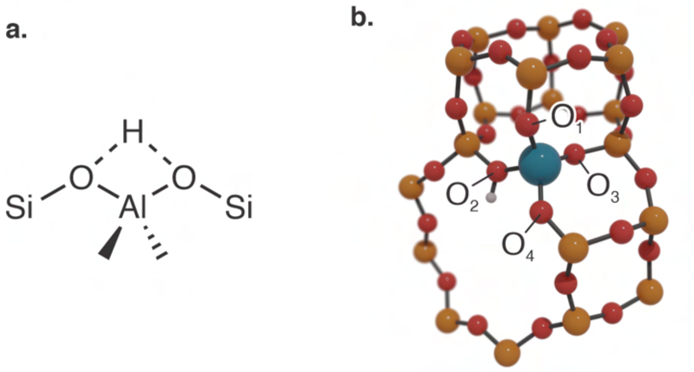


Figure 5.4: *a.* Schematic depiction of the proton hopping reaction between two oxygen atoms in the first coordination sphere of the Al defect. *b.* A portion of the H–CHA unit cell, showing the nomenclature of the four inequivalent O atoms in the first coordination sphere of the Al defect adopted hereafter. Adapted from Ref. 237.

To train the MLP, we performed a set of high-temperature (873 K) DFT US simulations to sample all the 6 possible hopping paths. In principle, it is advantageous to generate the training set at high temperature, to extend as much as possible the explored phase space. Since we wanted to validate the MLP against DFT FESs anyway, we ran sufficiently long simulations to obtain converged enough profiles. This also meant that a massive amount of training data was available ($\sim 1,200,000$ structures, randomly divided in a training and validation set with a 80:20 ratio). Efficient data generation was not a concern here, but in the conclusions of this chapter we will outline some strategies to dramatically reduce the burden of the training data generation step.

The speedup offered by the MLP is really remarkable. Indeed, each DFT evaluation took more than 8 s on 56 Xeon E5-2680v4 CPUs@2.4GHz cores, while the MLP evaluation, run on a single NVIDIA Volta V100 GPU, requires a bare hundredth of a second. With a DFT-quality PES that can be evaluated at such speed, a lot of additional information can be extracted from the system under study. Hereafter, we resume the main benefits achieved by using the MLP.

5.3.1 Benefit 1: improving convergence of DFT results

Even when using enhanced sampling to explore activated events, properly converging a free energy surface is often an hassle. Ideally, one would have to use a very large number of umbrellas with tight κs (eq. 4.6) and simulate for very long times, especially if the diffusive motion in the umbrella is slow and produces highly-correlated samples. Thanks to the MLP, one can fundamentally increase the simulation length and the number of umbrella to taste, obtaining extremely well-converged FEPs.

As an example, Figure 5.5*a* shows the FEP of the 2–3 hopping computed with DFT using 19 umbrellas and 50 ps simulation time per umbrella. These values are quite standard for typical DFT US simulations. In the same figure, the MLP FEP is also shown, which is however obtained from 39 umbrellas and 100 ps simulation time per umbrella. Remark that the MLP was trained on simulations performed at 873 K, while the sampling is here performed at 573 K. The DFT and MLP results are then completely uncorrelated. As it can be seen, a strange spike is present atop the transition state in the DFT profile, which is completely absent in the MLP case. To test whether this is an issue in the MLP accuracy or rather caused by limited sampling in the DFT case, we introduced 2 additional umbrellas atop the transition state region and increased the simulation time to 90 ps per umbrella. The new profile (Figure 5.5*b*) now perfectly matches the MLP one. This clearly shows that not only does the MLP provides DFT-quality results at lower temperatures than the training one, but it also allows to obtain extremely well-converged profiles for a fraction of the computational cost.

5.3.2 Benefit 2: going beyond the transition state theory approximation

In Section 4.2.4 we discussed how, starting from the Bennett–Chandler approach to transition state theory, it is possible to approximate the reaction rate by setting the chances of barrier recrossing to zero. On the other hand, there are ways to explicitly include barrier recrossing in the kinetic constant calculation. In CV-based enhanced sampling simulations, barrier recrossing can be accounted for with the **reactive flux** formalism[218]. There, a very large number of short unbiased MD trajectories is initialized atop the transition state region and monitored through time to assess whether they end up in the product basin or not. This allows to explicitly compute the ensemble average $\langle \dot{q}(0)\theta\left(q(t)-q^{*}\right)\rangle_{q(0)=q^{*}}$ (see eq. 4.7) and obtain the true kinetic constant of the reaction. We found that 5,000 trajectories of at least 50 fs each are needed to obtain converged results for the proton hopping reaction.

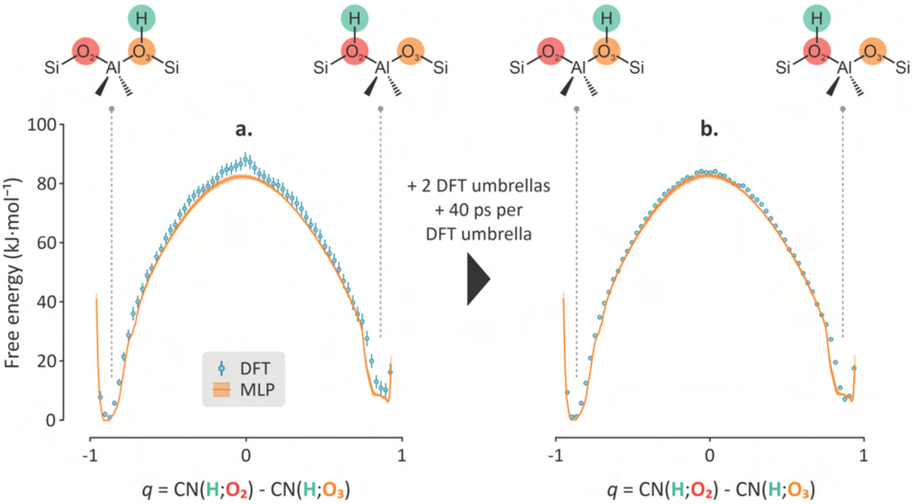


Figure 5.5: *a.* Comparison between the DFT and MLP FEP for the proton hopping reaction at 573 K between O_2 and O_3 (see Figure 5.4*b*). The DFT FEP was obtained using 19 umbrellas with 50 ps simulation time per umbrella while the MLP one using 39 umbrellas and 100 ps per umbrella. *b.* Same as in *a* but with two additional DFT umbrellas and a total of 90 ps simulation time per umbrella. The agreement between the two FEPs clearly improves. Adapted from Ref. 237.

It is important to consider that the rate was computed for all 6 hopping paths at 7 different temperatures uniformly spaced between 273 and 873 K. Clearly, this step is already pretty much unfeasible with traditional DFT simulations. The effect of barrier recrossing is, for this specific reaction, quite small. This can be seen in Figure 5.6*a*, where the overall rate of the proton hopping reaction is obtained using TST ($k^{\mathrm{c}}_{\mathrm{TST}}$) or reactive flux ($k^{\mathrm{c}}_{\mathrm{BC}}$). The explicit inclusion of barrier recrossing produces a negligible decrease in rate across the considered temperature range for this specific reaction, but it could become more important for processes with a less sharp transition state region.

5.3.3 Benefit 3: explicitly including nuclear quantum effects

Modelling the proton hopping reaction with high accuracy, aside from the chosen level of theory, poses an interesting challenge. Indeed, since it involves an hydrogen atom, quantum effects at the nuclear level could be of importance in affecting the computed reaction rate. As discussed in Section 4.1, Born–Oppenheimer MD considers nuclei as classical particles, whose motion is propagated with Newton's equations. To explicitly recover the quantum behavior of the system's nuclei, ring polymer MD (RPMD) can be

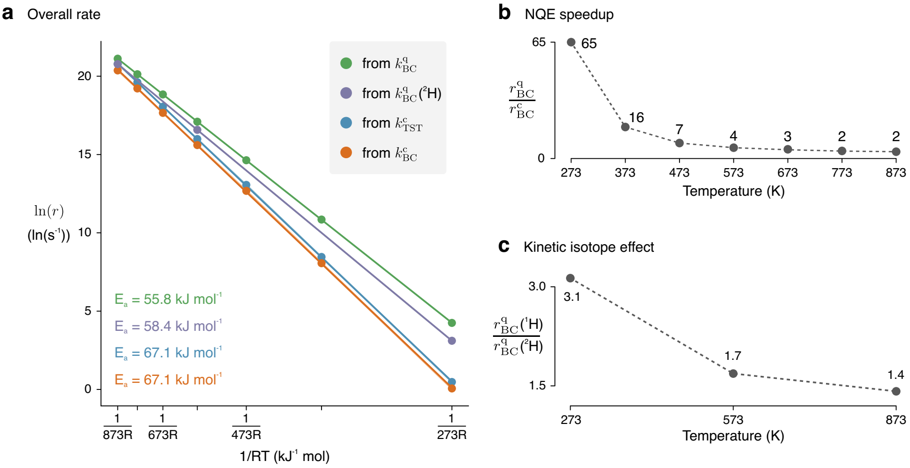


Figure 5.6: MLP simulations allow to compute rates for the proton hopping reaction over a broad temperature range. *a.* Arrhenius plot showing the overall hopping rate as function of 1/RT. The results are obtained from classical simulations and transition state theory (k^{c}_{TST}) or reactive flux (k^{c}_{BC}), from reactive flux and RPMD simulations with protium (k^{q}_{BC}) and with deuterium ($k^{q}_{BC}(^2H)$). *b.* Speedup in the rate due to the inclusion of the NQEs. *c.* KIE as function of reaction temperature. Adapted from Ref. 237.

used, which is based on Feynman's path integral formulation of quantum mechanics. Rather surprisingly for a non-physicist like me, the wave-like behavior of the nuclei can be approximated by considering P replicas of the system, called beads[238] (Figure 5.7). Equivalent atoms in each replica are connected by harmonic springs. At each MD step, this harmonic potential is taken into account when propagating the equation of motion and it can be shown that, at the limit for the number of beads $P \to \infty$, the right quantum statistics for the system is recovered.

The obvious problem is that we now have P MD simulations running synchronously. Considering that converged results require, in general, between 10 and 100 beads, it is clear that RPMD is out of reach even for GGA-DFT when medium-large systems are considered. On the other hand, the simulations run on the classical PES that, thanks to the MLP, we can estimate with extreme speed. We were therefore able to compute the quantum rate for the proton hopping reaction based on reactive flux simulations which, to the best of our knowledge, was never attempted in heterogeneous catalysis. Simply performing the RPMD US simulations already required more than 0.6 μs of simulation time, which is well-beyond the possibilities of DFT. We found that including nuclear quantum effects results in a remarkable increase in the proton hopping rate, as it can be seen by comparing the classical

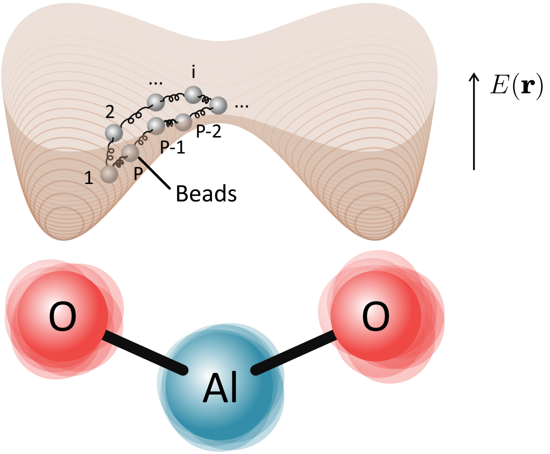


Figure 5.7: A schematic depiction of the PES experienced by the BAS when hopping between two zeolite oxygens. Atop the PES, a ring polymer representing the quantum hydrogen nucleus with P beads connected by harmonic potentials is shown. Adapted from Ref. 237.

reactive flux kinetic constant $k^{\mathrm{c}}_{\mathrm{BC}}$ with the quantum one obtained from RPMD simulations $k^{\mathrm{q}}_{\mathrm{BC}}$ in Figure 5.6*a*. A significant increase in rate due to the inclusion of nuclear quantum effects persists up to $\sim$473 K (Figure 5.6*b*), a temperature which is already in the range of zeolite-catalyzed reactions performed at mild conditions.

5.3.4 Benefit 4: easily computing kinetic isotope effects

As outlined in Section 4.3.2 when discussing guaiacol demethylation, KIE can be a powerful tool to deduce and validate reaction mechanisms. In that study, unfortunately, we could not validate the experimentally-measured KIE (^{1}H vs. ^{2}H) because of the prohibitive cost of DFT simulations, but only confirm that our mechanistic proposal was compatible with the experiment. To obtain a good KIE estimate, one would have to use RPMD to account for the ‘reduced’ quantum nature of the deuterium nucleus with respect to protium and, obviously, to perform a second simulation from scratch doubling the mass of the hydrogen nuclei. Also in this case, the MLP comes to the rescue. Indeed, the PES only depends on the atomic number of the atoms and not on their mass number. Therefore, the MLP trained on ^{1}H US can be used as such to immediately perform simulations with ^{2}H and compute the reaction KIE.

This approach is not new in the literature but, up to now, it has only been successfully applied to very small molecules for gas-phase reactions (see e.g. ref. 239). Thanks to the MLP, we were able to explicitly compute the deuterium hopping rate in D–CHA over the temperature range 273–873 K ($k^{\mathrm{q}}_{\mathrm{BC}}(^{2}\mathrm{H})$ in Figure 5.6*a*). The force error during the simulations remained extremely low, confirming that the system with deuterium does indeed explore a similar phase space as the protium one. We showed that, at 273 K, a KIE of about 3 has to be expected (although currently no experimental evidence is available to corroborate our results). The KIE value then decreases to 1.7 at 573 K and to 1.4 at 873 K, as expected based on standard Bigelesen–Mayer theory[240, 241] (Figure 5.6*c*). It has to be pointed out that the deuterium rates remain substantially higher than the classical protium ones, indicating that quantum effects persist even if the mass of the nucleus is doubled.

5.4 Outlook

MLPs are currently entering every aspect of theoretical chemistry and I'm sure they will play a key role in the next-generation studies of zeolite-catalyzed reactions[242]. Seminal work has already been performed on all-silica frameworks, highlighting how it is possible to generate highly transferable MLPs to compute thermal and mechanical properties of the materials[243, 244]. Still, a bottleneck was present, namely the large number of DFT datapoints required to train the MLP with high accuracy. Luckily, recent advances are making this problem less and less relevant.

Starting from the MLP architecture itself, the newer MLP generations have been making use of **equivariant** features rather than invariant. An example is given by the `NequIP`[245] equivariant deep neural network potential. 'Equivariant' means that the atomic features in the network are not only scalars, but may also be tensors that satisfy the atomic environment equivariance with respect to rotations. Using higher order tensors remarkably improves the data efficiency of the network, as within a tensor representation more information of the atomic environment can be captured. The MLP can then be trained on a much smaller set of DFT datapoints while achieving similar accuracy as invariant MLPs[237]. Secondly, data generation can also be drastically improved. Recently within our group, Sander Vandenhaute developed an innovative active learning procedure to quickly obtain an accurate MLP for activated events[246]. The simulation starts with a model trained on few structures (that can be generated, for instance, by random displacements of the atoms). This model, in spite of certainly not being very accurate, is used to perform many parallel metadynamics simulations. Every few hundreds of steps, the simulations are stopped and the final structures are added to the training set. A new (better) model is then retrained on such extended dataset and used to continue the metadynamics simulations, until a sufficient coverage of the phase space (and a sufficiently low error) is obtained. A summary of the procedure is graphically depicted in Figure 5.8.

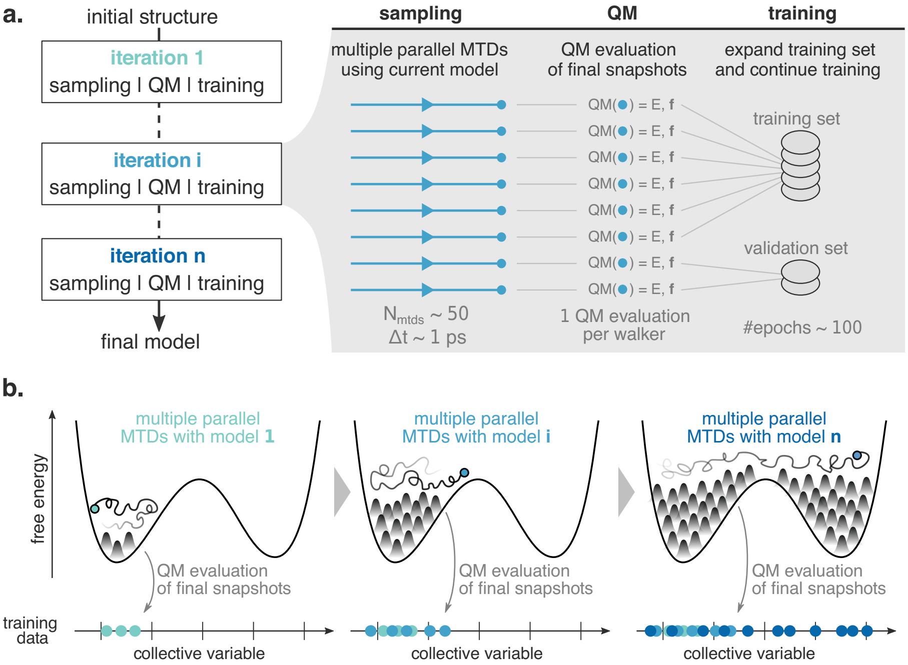


Figure 5.8: To reduce to a minimum the cost of an MLP training, it is possible to start with a bad model trained on few structures and use it to perform short parallel metadynamics simulations to sample the reaction of interest (*a*). At the end of the run, the final trajectory snapshot is added to the training set, the model is retrained and used to continue the metadynamics simulations (*b*). Adapted with permission from Ref. 246.

This procedure, tested on the proton hopping reaction as implemented in the software package `PsiFlow`[247], gave remarkable results. As it can be seen in Figure 5.9, we tested the 2–3 hopping reaction using 40 parallel walkers randomly divided in a training and validation sets with a 90:10 ratio. The training was performed with `NequIP`, using a relatively small network with a maximal rotational order for the features of 1. Over the course of 10 iterations, it can be seen how the CV space gets gradually sampled and the mean absolute error (MAE) on the forces tends to decrease. At the last iteration, the MAE is ∼60 meV $\cdot$ Å^{-1}, which is not much larger than the one we reported in **Paper V**. It must be noticed that (i) the amount of training points is massively reduced (440 vs. ∼1,200,000) and (ii) the actual user time to reach this preliminary results was barely one day, while almost a month was required to perform the 50 ps US simulations to generate the training set in **Paper V**. With a bit more training points, training time or a better network it should be easy to drastically improve the final accuracy of the network.

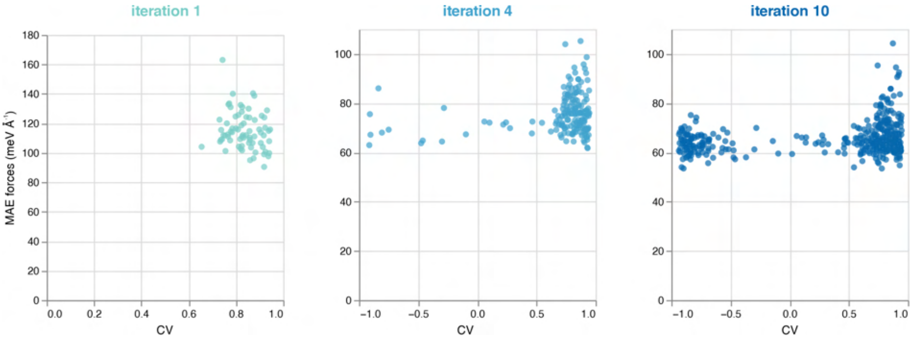


Figure 5.9: Mean absolute error (MAE) on the training set forces plotted as function of the reaction CV for various iteration of the active learning procedure depicted in Figure 5.8. The reaction under investigation is the hopping between O_2 and O_3 in H–CHA (see Figure 5.4*b*).

We hope that with this final example we were able to highlight how fast the field of MLPs for molecular simulations is currently evolving. It is not far-fetched, in my opinion, to state that MLPs are going to quickly relieve us from the computational burden of performing full ab initio MD simulations, allowing to reach length- and timescales so far unattainable. Indeed, the procedure illustrated here is readily applicable to any type of reaction and should allow to obtain in short amounts of time accurate kinetic constants and free energies. While obtaining rate estimates for complex reaction networks purely based on enhanced sampling simulations was considered almost impossible, I believe that with MLPs this goal has become not only achievable, but it might soon even be routine.

6

Conclusions and Perspectives

In this thesis, we have shown how different computational methodologies can be used to derive fundamental insights in the mechanism and reactivity of zeolite-catalyzed reactions. We began by investigating the mechanism of reactions involving relatively bulky and poorly mobile intermediates. This cases can be approached with static modeling, where a limited set of important points is identified on the PES and their local environment is explored with a normal mode analysis. **Paper I** is situated in the context of the BioFact EOS project, granted by the Flanders found for research and innovation (FWO) to consortia involving multiple universities and research centers[†]. We used static simulations to investigate and explain the findings of our experimental partners from the group of B. Sels in the KU Leuven concerning the zeolite-catalyzed *C*-dealkylation of lignin-derived alkylphenols. In that way, we could explain the reasons underlying the large reactivity differences between different substrates observed experimentally. Furthermore, we provided mechanistic insights on the most favorable mechanism followed by the reaction.

In **Paper II**, we worked together with our experimental partners from the group of D. De Vos in the KU Leuven. They recently discovered that Pd-loaded Beta zeolite could catalyze with high selectivity the direct cross-

---

[†] Center for Surface Chemistry and Catalysis, KU Leuven (Sels, De Vos); Center for Molecular Modeling, Universiteit Gent (Van Speybroeck); Organic Synthesis Division, Universiteit Antwerpen (Maes); Laboratory of Organic Chemistry, Université Libre de Bruxelles (Evano); Center for Education and Research on Macromolecules, Université de Liège (Detrembleur, Debuigne); Applied Sustainable Catalytic Processes, Leibniz Institut für Katalyse (Beller).

coupling of toluene to produce mainly *para,para'* bitolyl, a remarkable advantage with respect to the non-selective homogeneous catalysts investigated thus far. Again, static calculations were a useful resource to understand the unusual reaction mechanism, which goes through a concerted metalation-deprotonation followed by a migratory insertion step, none of which is part of the traditional Suzuki reaction. Moreover, we highlighted how the restrained zeolite environment strongly favors *para* selectivity in the first step, while for the second step the selectivity is at most moderate, in line with the experimental evidences.

It has to be pointed out that both **Paper I** and **Paper II** do not exclusively rely on static methodologies, but rather on a complementary approach with MD simulations. While we did not stress this extensively in Chapter 3 to maintain the discussion focused on the methodology, it is important to notice that the mobility and behavior of reactants and intermediates was extensively characterized by means of classical MD simulations prior to the actual mechanistic investigation. This provided us with a clearer picture of the molecular behavior at operando conditions, laying a better foundation for the subsequent static simulations of the reaction mechanism.

While proceeding in my PhD, it became apparent that a static methodology is very limited when one wants to study highly mobile species or a large number of co-adsorbed molecules. For this reason, we began to focus on enhanced sampling MD. **Paper III** is part of an FWO research project in collaboration with F. De Proft (Vrije Universiteit Brussel) and B. Weckhuysen (Utrecht University). The purpose of the project was to shed some light on the mechanism of aromatic electrophilic substitution, with particular focus on the formation of charged Wheland-type intermediates. Using enhanced sampling MD, we showed that this type of intermediate is indeed formed at operating conditions when benzene is ethylated with either ethene or ethanol. Moreover, we highlighted the important role of water in modulating the kinetic of protonation reactions in zeolites, which is highly susceptible to the loading. Water can easily form solvated hydronium ions starting at coverages of $\sim$3 molecules per BAS, which are highly mobile and drastically change the catalytic properties of the original active site. Therefore, they are an ideal case study for enhanced sampling techniques.

In this respect, **Paper IV** is a further step forward. It is again part of the BioFact project and, differently from **Paper I**, it focuses on the *O*-dealkylation of lignin-derived compounds. We used enhanced sampling techniques to simulate the dealkylation reaction conducted in hot liquid water with either an homogeneous Brønsted acid catalyst or a zeolite. We obtained some unexpected mechanistic features, which were fully supported by the experiments conducted in the group of B. Maes (University of Antwerp) for the homogeneous part and B. Sels (KU Leuven) for the heterogeneous one. This combined computational/experimental effort allowed to shed some new light on the fundamental reasons underlying the activity enhancement of hydronium ions in the restrained zeolite pores. This topic is of great interest in modern research, as many reactions in the field of biomass conversion are conducted at mild, solvated conditions.

In the context of enhanced sampling techniques, I believe that the value of the results lies not only in the mechanistic insights we obtained, but also on the methodology. Thanks to the efforts of Prof. L. Vanduyfhuys in our group, we were among the first to extensively apply transition state theory to derive CV-independent rates from our simulations. While gaining popularity, this practice is still not extensively applied in modern zeolite catalysis literature.

While going from static simulations to enhanced sampling MD improves the PES exploration, in **Paper V** we tried to boost the efficiency with which such exploration can be performed. We relied on recently proposed machine learning architectures to learn with high accuracy the PES of the proton hopping reaction in H–CHA starting from a set of high-temperature umbrella sampling simulations. The MLP fundamentally gave us the possibility of performing MD with an increase in speed of multiple orders of magnitude compared with standard DFT. With this power unlocked, we were able to perform a whole new set of simulations typically precluded to ab initio techniques, such as path integral molecular dynamics to include nuclear quantum effects and reactive flux to compute accurate kinetic constants. While many groups working on the topic have showcased similar methodologies for simpler systems, we were—to the best of our knowledge—among the first to perform such an extended investigation for a reaction in an heterogeneous catalyst.

To conclude, I would like to discuss what I believe is going to be achievable in the field of computational zeolite catalysis in the near future and where it could be headed in the upcoming years. In my opinion, the key word is machine learning potentials. I hope that the final Section of Chapter 5 was able to give an idea of the pace at which the field is currently evolving. Data generation practices and neural network architectures that were state-of-the-art two years ago (when I had the initial thoughts that led to **Paper V**) can already be considered obsolete. The possibility of training MLPs with few hundreds of data points efficiently generated with an active learning procedure will solve, I believe, three key issues currently encountered in zeolite catalysis modeling:

1. **Accuracy enhancement:** as extensively explained in Section 2.3, GGA DFT can have severe accuracy limitations when used to derive reaction

barriers or the stability of charged intermediates. Unfortunately, performing enhanced sampling simulations with higher ranks on Jacob's ladder (Figure 2.1) was thus far prohibitively expensive. As we saw in Section 5.4, on the other hand, a few hundreds of single point calculations are sufficient to train an accurate MLP. Such a small number of calculations is certainly feasible with hybrid functionals and, for small zeolites (due to memory issues), even with double hybrids or MP2. I am therefore quite confident that training an MLP to investigate a reaction with post-HF accuracy in a small framework should certainly be possible with the currently available hardware and software.

2. **Exploring full reaction networks:** even with GGA DFT, enhanced sampling remains expensive. Computing the whole reaction mechanism of benzene ethylation in **Paper IV**, for instance, involved a remarkable consumption of computational resources spanned over many months, despite consisting of 'only' 6 reaction steps. Training an MLP with thermodynamic transferability should make it possible to investigate in a reasonable time networks with hundreds of reaction steps, possibly across many different condition such as temperature, concentration of co-adsorbed molecules, etc. This will be an essential step to enhance the predictive power of simulations and their ability to guide the experiment towards the ideal conditions to perform a reaction.

3. **Achieving unprecedented length scales:** most investigations of zeolite-catalyzed reactions rely on one or (at most) few unit cells to represent the periodic material. However, there are important processes that occur on the crystal surface or within mesopores, such as the building up of large aromatic hydrocarbons leading to coking (Section 1.4). Studying the behavior of these molecules with full DFT MD is currently unfeasible. On the other hand, we saw that the MLP energies are defined in terms of single atoms and their local environment. It is then easy to imagine a divide-an-conquer approach, where the MLP is trained on portions of the system as large as allowed by the DFT software and then used to model systems with tens of thousands of atoms.

My hope is that the results presented in this thesis can be seen as an initial step to reach these goals and that, in the upcoming years, the quality of our models and simulations will get closer to the actual material used in both industry and research laboratories. Only then the predictive power of simulations would be sufficient to reach the dream of any computational chemist working in catalysis, i.e. the in silico design of reactions and materials with unprecedented performances.

Part II

A Selection of Published Papers

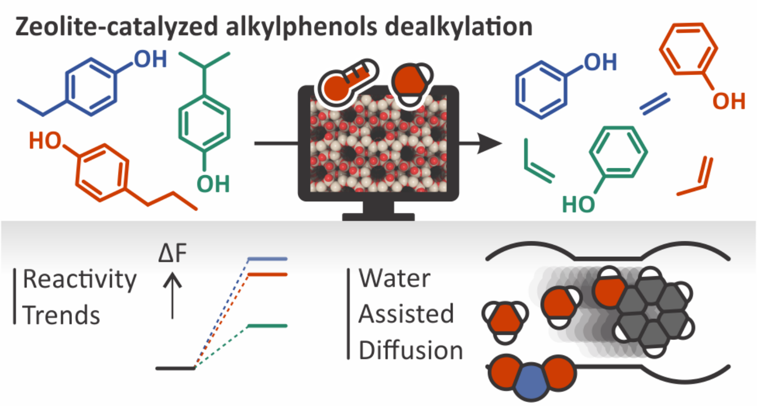
Not all the papers in which I had the pleasure to be involved are appended in this section, but only the ones in which I was the leading author or I had a major contribution. A full list of publications and conferences contributions is given in Appendix A.

Paper I

Insights into the mechanism and reactivity of zeolite catalyzed alkylphenol dealkylation

Massimo Bocus and Veronique Van Speybroeck*

ACS Catalysis, **2022**, *12*, 14227–14242



*Corresponding author
M. Bocus performed the simulations and prepared the initial draft of the manuscript.

ACS Catalysis

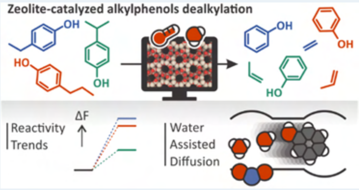
pubs.acs.org/acscatalysis Research Article

Insights into the Mechanism and Reactivity of Zeolite-Catalyzed Alkylphenol Dealkylation

Massimo Bocus and Veronique Van Speybroeck*

Cite This: *ACS Catal.* 2022, 12, 14227−14242 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: In the stride toward the production of low-carbon-footprint commodity chemicals, the development of a complete wood biorefinery plays a pivotal role. The lignin fraction of wood can be depolymerized and demethoxylated mainly into 4-alkylphenols. These phenolic compounds can further catalytically be C-dealkylated within the H-ZSM-5 zeolite at relatively high temperatures and in the presence of steam, producing phenol and olefins. Experimentally, the dealkylation reaction was found to have two striking features: first, different reactants possess very different reactivity. 4-Ethylphenol (4-EP) is somehow less reactive than 4-*n*-propylphenol (4-*n*-PP), which is in turn much less reactive than 4-isopropylphenol (4-*iso*-PP). Second, cofeeding of steam in the reaction mixture was necessary to prevent rapid and reversible catalyst deactivation. Herein, a combination of static and dynamic density functional theory (DFT) simulations is used to unravel the molecular and mechanistic origin of these observations. Free-energy profiles obtained from static calculations confirm the experimentally observed reactivity sequence, where our computations show that the secondary nature of the alkyl carbon involved in 4-*iso*-PP dealkylation strongly stabilizes the respective transition states. To investigate the effect of water on the mobility of the reactive species and their interaction with the active site, we investigated the diffusion of phenol along the H-ZSM-5 straight channel in the presence of water loadings from 0 to 3 molecules per zeolite unit cell. We show that water has a strongly beneficial effect in promoting desorption and diffusion of phenol away from the Brønsted acid site through competitive adsorption and by the formation of hydrogen bond chains with the diffusing phenol. This effect could lead to a shorter residence time inside the zeolite, preventing active site poisoning and condensation to bulkier biphenylether moieties.

KEYWORDS: *zeolite, DFT, enhanced sampling, dealkylation, water, molecular dynamics, lignin valorization*

1. INTRODUCTION

The depletion of easily accessible crude oil reservoirs around the world, joined with the increasing levels of anthropogenic CO_2 in the atmosphere, is exacerbating the need for new and renewable sources of carbon-based commodity chemicals.[1,2] Being the core of many polymeric materials, aromatics are indispensable for the chemical industries but are nowadays almost exclusively obtained from fossil resources.[3] The identification of sustainable and aromatic-rich raw materials is therefore crucial.

Among all natural polymers, lignin is likely to become such a source of aromatics. Together with (hemi)cellulose, it is one of the main polymeric components of the plant cell wall, and it is generally discarded or used as a low-value fuel in the paper-production process.[4,5] Nevertheless, it possesses the highest percentage by weight of aromatic units with respect to all other natural polymers. Lignin exploitation has been hampered by its heterogeneity and chemical stability that made it difficult to design a straightforward, selective, and efficient depolymerization process.[6] Despite this, the entirely new concept of lignin-first biorefinery has been developed in the past few years.[7,8] Thanks to advanced catalytic systems, it has become possible to depolymerize lignin directly from the raw biomass without significantly degrading the (hemi)cellulosic fraction in a process called reductive catalytic fractionation (RCF).[9,10] If—for instance—birch wood is used, the product oil contains mainly *para*-*n*-propyl derivatives of guaiacol and syringol (Figure 1, left)[11,12] that can then be demethoxylated using Ni under hydrogen atmosphere, producing *para*-substituted alkylphenols.[8]

The following step in the path toward the production of commodity chemicals consists of further defunctionalization of alkylphenols to obtain basic building blocks useful for the chemical industry. Recently, highly efficient gas-phase dealkylation of alkylphenols was proposed, catalyzed by acidic zeolites (Figure 1, right).[13−16] Among the many tested aluminosilicates, higher selectivity was achieved with the H-

Received: August 4, 2022
Revised: October 11, 2022
Published: November 4, 2022

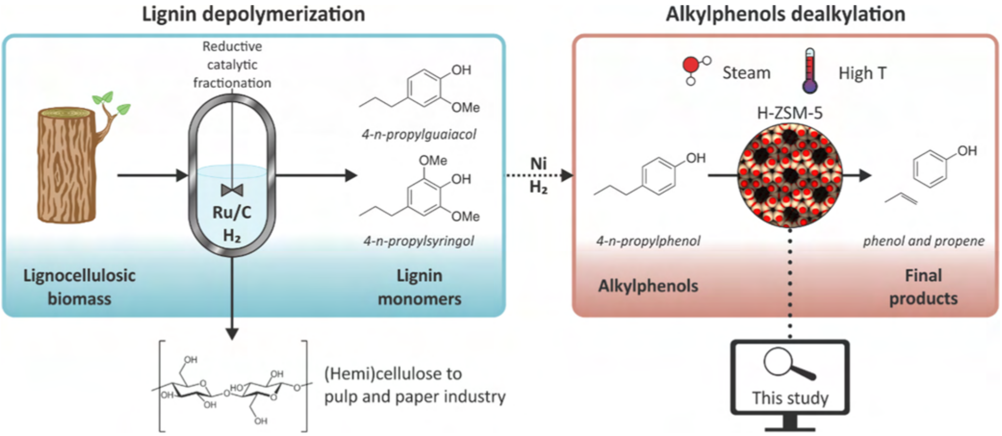


Figure 1. Schematic representation of the lignin-to-chemicals process. The lignocellulosic biomass undergoes reductive catalytic fractionation (RCF, blue box). In the process, lignin is depolymerized in a mixture of lignin monomers, while the (hemi)cellulosic fraction is recovered for further usage. The lignin monomers can be demethoxylated with a Ni catalyst under hydrogen to produce alkylphenols, which are then further dealkylated using the H-ZSM-5 zeolite at high temperatures and in the presence of steam (red box).

ZSM-5 zeolite with a moderately high Si/Al ratio (*vide infra*). It was also found that the presence of steam is important to prevent rapid catalyst deactivation. The reaction requires fairly high temperatures (500−600 K depending on the substrate) for the entropic gain to overcome the positive enthalpic variation.[15] The products, phenol and olefins, are both very valuable and requested by many industrial chemical processes.[8]

When 4-ethylphenol (4-EP), 4-*n*-propylphenol (4-*n*-PP), and 4-isopropylphenol (4-*iso*-PP) are used as model reactants for the dealkylation reaction (Figure 2a), it was experimentally observed that they present very different reactivities.[14,15] Specifically, 4-EP is the less reactive reagent, reaching ∼50% conversion at about 650 K (over H-ZSM-5 with Si/Al = 40). 4-*n*-PP is slightly more reactive, ∼50% of conversion being achieved around 575 K. 4-*iso*-PP, finally, is extremely active and almost full conversion can already be observed below 500 K. The reasons underlying such a large reactivity difference, however, were not fully understood.

A further option in the path toward commodity chemicals that was recently explored consists in avoiding the demethoxylation step and simply demethylate the lignin-derived 4-alkylguaicols.[17] This produces 4-alkylcatechols, that can be dealkylated in a similar fashion as the 4-alkylphenols to produce catechol and olefins. While the presence of the second hydroxyl group leads to a more complex chemistry and lower yields are observed, it does not seem to extensively influence the reactivity toward dealkylation, which is similar between 4-*n*-PP and 4-*n*-PC (4-*n*-propylcatechol).[17]

Another peculiar aspect of the alkylphenol dealkylation reaction, which remarkably differentiates it from dealkylations on purely hydrocarbon substrates, is that steam is required to prevent quick and reversible catalyst deactivation. This was attributed to water preventing the condensation of phenol to diphenylether, by both disfavoring the reaction equilibrium as well as promoting phenol desorption from the zeolite active sites.[15] This effect is very prominent, as Arrhenius activation energies measured for the dealkylation of 4-*n*-PP[15] resulted to be higher in the presence of water (from 67−68 to 99−103 kJ·mol^{-1}) but, on the other hand, the measured dealkylation rate was faster as in the absence of water an almost instantaneous poisoning of the dry catalyst was observed. While the impact of water on intrinsic reactivity is an active field of research,[18−20] its effect on diffusivity of reactive compounds is still relatively unexplored.

Within the field of zeolite catalysis, *ab initio* periodic calculations using density functional theory (DFT) have been extensively used to provide valuable mechanistic insights for a wide range of industrially relevant reactions.[21,22] To the best of our knowledge, the chemistry of alkylphenols in zeolites has not received—from a computational perspective—a large attention so far. The investigations have mostly focused on the alkylation of phenol with alcohols, such as methylation over H-FAU[23] and tert-butylation over H-BEA.[24] On the other hand, a well-characterized reaction is the alkylation of benzene with olefins,[20,25−29] used industrially to produce ethylbenzene. This reaction represents a relevant means of comparison for our case study as it consists in the opposite process with respect to dealkylation. Based on its known mechanistic features, it can be assumed that the dealkylation of alkylphenols can proceed through three possible pathways (Figure 2b): initially, the molecule is protonated at the *para* carbon by the BAS, forming an arenium ion commonly known as Wheland complex (**II**). This intermediate is a key representative of the rich carbocation chemistry within zeolite pores[30,31] and its existence has recently been proven experimentally during benzene ethylation in H-ZSM-5 through UV−vis spectroscopy[32] and by our group through enhanced sampling simulations at operating conditions.[20] Given the higher nucleophilicity of phenol with respect to benzene, the Wheland complex formation should be even more prominent in this case. After protonation, the substrate can then undergo a concerted dealkylation step (TS2 in Figure 2b) to immediately give the final products (**III**). Alternatively, the alkyl group can be transferred to the zeolite framework, forming a surface alkoxide species (SAS, **IV**) that can subsequently deprotonate to give the final alkene. Finally,

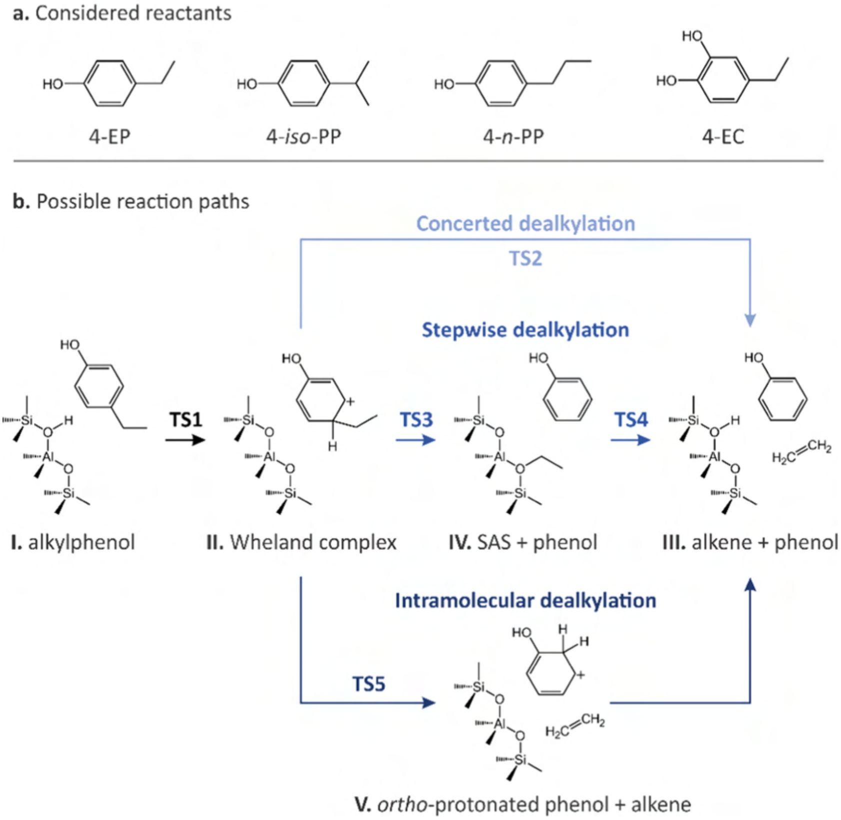


Figure 2. (a) The four reactants considered in this work, namely, 4-ethylphenol (4-EP), 4-*n*-propylphenol (4-*n*-PP), 4-isopropylphenol (4-*iso*-PP), and 4-ethylcatechol (4-EC). (b) Three possible mechanistic pathways for the conversion of an alkylphenol (**I**, 4-EP is used as the model reagent) into phenol and an alkene (**III**) are investigated: they all start with the formation of a Wheland complex (**II**) that can undergo concerted dealkylation (TS2), a stepwise path (TS3, TS4) with the formation of an intermediate surface alkoxide species (SAS, **IV**) or an intramolecular dealkylation (TS5), which leads to *ortho*-protonated phenol (**V**).

inspired by the work of Kolboe et al.,[33] we also investigated an intramolecular dealkylation where the alkyl tail transfers its excess proton to the *ortho* position of the aromatic ring (TS5 and intermediate **V** in Figure 2b). The latter can then transfer a proton back to the zeolite framework to give the final products.

To gain a better understanding of the preferred mechanism and the reactivity differences between various substrates during the H-ZSM-5-catalyzed alkylphenol dealkylation, we thoroughly investigated the behavior and reactivity of 4-EP, 4-*n*-PP, 4-*iso*-PP, and 4-EC (4-ethylcatechol) employing static and dynamic periodic DFT calculations. While the former represents a computationally efficient tool to unravel activation and adsorption energies of reacting species, the latter allows to better explore the flat potential energy surface (PES) in the catalyst pores for highly mobile species, thus better reproducing their behavior at realistic *operando* conditions.[34] We show that the static results allow to rationalize the reactivity differences between the various alkylphenols, with reaction barriers that are in line with the experimental kinetic measurements.

To obtain insight into the effect of water on the adsorption and diffusion of phenolic compounds, umbrella sampling (US) molecular dynamics simulations were performed to study the diffusion of phenol through the zeolite pore system in the anhydrous and hydrated case. Molecular dynamics simulations at operating conditions are quintessential to account for the high mobility of water in the pores of the zeolite. Water is shown to greatly improve the desorption and diffusion of phenol away from the BAS. This is caused by competitive adsorption between water and phenol on the BAS and, at the same time, to the formation of hydrogen bond chains between water and the phenol hydroxyl group when the latter moves away from the active site.

2. COMPUTATIONAL DETAILS

Three main types of periodic DFT calculations have been performed to investigate the reactivity differences between the considered alkylphenolic reactants (see Figure 2a) and to assess the impact of water on phenol diffusion. First, regular molecular dynamics (MD) simulations have been used to investigate the behavior of the main possible dealkylation intermediates (Figure 2b) and obtain insights in their stability and behavior at operating conditions. Then, static calculations allowed to compute the free-energy barriers for the whole dealkylation mechanism, providing some interesting reactivity trends, while being significantly less computationally demanding than MD simulations. Finally, to assess the role of water on phenol diffusion, the static approach is insufficient due to the high mobility of the water molecules and their strong interaction with the BAS, which can also result in its complete abstraction from the framework.[20,35,36] On the other hand, the diffusion of aromatics along the H-ZSM-5 straight channels is also an activated process;[37] therefore, it is impossible to sample the diffusion of phenolic compounds during a regular MD simulation. To simulate diffusion of phenol in the presence of water at operating conditions, we have used umbrella sampling molecular dynamics simulations according to a procedure introduced earlier by some of the presenting authors.[38,39]

2.1. Catalyst Model. All simulations are performed in a periodic H-ZSM-5 model, that allows to account for the confinement effects induced by the zeolite three-dimensional pore structure. The H-ZSM-5 zeolite possesses an MFI topology, consisting of straight 10-rings channels intersected by sinusoidal ones. The unit cell contains 96 tetrahedral silicon atoms connected by oxygen bridges. As previously done,[40,41] we substituted the Si at the T12 position with an aluminum atom obtaining a Si/Al ratio of 95. This is comparable with the experimental Si/Al ratio of 40,[15] which should mainly correspond to isolated active sites. The T12 site is located at the channel intersection and it ensures maximal accessibility for the substrate (Figure S1).[42] The negative charge created by the Al substitution is compensated by the addition of a proton on the O_{Z1}, adjacent to the defective site. The unit cell parameters used in the static simulations are analogous to the optimized ones reported in ref 43 (a = 20.02 Å, b = 20.25 Å, c = 13.49 Å, α = 89.87°, β = 89.69°, and γ = 90.10°).

2.2. Static Calculations. All of the computed reaction barriers were obtained by means of static periodic DFT simulations. We used the Vienna *Ab initio* Simulation Package (VASP 5.4.4)[44−46] with the Projected Augmented Wave (PAW) method.[47,48] The Perdew−Burke−Ernzerhof (PBE)[49] exchange-correlation functional coupled with Grimme's D3 dispersion scheme[50] was employed in all calculations. The plane wave energy cutoff was set to 600 eV, while the self-consistent field (SCF) convergence threshold was set to 10^{-5} eV. To ensure that the same truncated basis set is used in all calculations, we fixed the cell volume and parameters at their optimized values (*vide supra*). Because of the large unit cell dimensions, the sampling of the Brillouin zone was restricted to the Γ-point.

The transition states search and optimization was performed using the improved dimer method.[51] Initial guess structures were prepared with the ZEOBUILDER software.[52] The optimized geometries were then further refined using a quasi-Newton algorithm.[53] The reactant and product states for each optimized transition state were generated by slightly displacing the structure along the normal mode with imaginary frequency corresponding to the transition-state crossing and then fully optimized with the conjugate gradient method. In all cases, the geometry optimization was stopped when the energy difference between two subsequent steps was lower than 10^{-4} eV.

Each stationary point was characterized with a normal mode analysis (NMA) in the harmonic approximation. A partial Hessian vibrational analysis (PHVA)[54] was used, including the adsorbed molecules and a T8 cluster around the BAS position. This choice was shown to be sufficient to obtain good results for the calculated thermodynamic properties.[55] A minimum in the potential energy surface (PES) should be characterized exclusively by positive vibrational frequencies, while a transition state by a single imaginary frequency along the reaction coordinate. In some cases, however, the flat PES in the zeolite pores makes it extremely difficult to remove all imaginary modes with very low frequency, even after multiple reoptimizations of the system. Those modes correspond with (almost free) rotations and translations of the adsorbed species in the zeolite pores. We then considered an optimization converged if, after two subsequent reoptimization attempts following the residual undesired imaginary mode(s), the computed energy variation was lower than 1 kJ·mol^{-1}. The remaining imaginary frequencies were replaced with an arbitrary value of 60 cm^{-1}, as proposed by De Moor et al.[55]

The same procedure was performed for other low-frequency modes (below 60 cm^{-1}) similarly to what has already been proposed in the literature, to reduce the numerical uncertainty and their impact on the calculation of the entropy.[56−58] Finally, the normal modes spectrum was used to calculate enthalpy, entropy, and free energy of adsorption for each species at the reaction temperature,[59] as implemented in our in-house-developed software TAMkin.[60]

2.3. Molecular Dynamics Simulations. To study the behavior and stability of adsorbed species in the zeolite fully taking into account their mobility and the flexibility of the catalyst, ab initio molecular dynamics (AIMD) simulations were performed. The CP2K software package (version 5.1)[61,62] was used, combining plane waves with a Gaussian atom-centered basis set (GPW).[63,64] In line with the static calculations, energies and forces were evaluated at a PBE-D3[49,50] level of theory, where we however resort to its revPBE parametrization because of its improved results.[65] A DZVP quality basis set with GTH pseudopotentials[66] was adopted and the plane wave energy cutoff set to 350 Ry. All simulations were performed in the NPT ensemble with a time step for integration of the equation of motion set to 0.5 fs. The temperature was controlled by a chain of five Nosé−Hoover thermostats,[67,68] while the pressure was controlled by an MTK barostat.[69]

2.4. Umbrella Sampling. The effect of water coadsorbed on the active site with phenol in potentially assisting and favoring desorption and diffusion of the latter was probed using umbrella sampling (US) simulations.[70,71] As a model system, a single phenol molecule was placed in proximity of the active site and, for the hydrated cases, one, two, or three additional water molecules were also introduced in its proximity.

In US, a set of quadratic bias potentials (V_i (CV)) defined as

$$V_i(\mathrm{CV}) = \frac{\kappa_i}{2}(\mathrm{CV} - \mathrm{CV}_{0,i})^2 \tag{1}$$

is applied along an *a priori* defined collective variable (CV). Each bias is characterized by a harmonic spring constant κ_i and it is centered at the CV value $\mathrm{CV}_{0,i}$. In this case, the CV was chosen to effectively describe the diffusion of phenol along the straight channel of H-ZSM-5, moving from the channel intersection with the BAS to the adjacent one, as schematically illustrated in Figure 3. While *ad hoc* CVs have been reported to study diffusion in other zeolite frameworks,[38,39] we here found that using the *x* component of the distance between the geometric center of the C and O phenol atoms and the geometric center of the skewed double ring that separates two channel intersections in H-ZSM-5 is effective in describing the diffusion process along the straight channel (Figure 3).

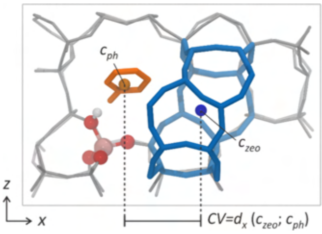


Figure 3. *x* component of the distance between the geometric center (c) of the phenol C and O atoms (in orange) and the geometric center of the double ring separating the two channel intersections in the H-ZSM-5 unit cell (in blue) is chosen as CV to describe the diffusion of phenol along the zeolite straight channel. The location of the Al tetrahedra and the BAS is highlighted in ball-and-stick representation, while the rest of the framework atoms (only half of them is shown) are in gray for the sake of clarity.

A set of 12 umbrellas were used to sample the diffusion from the channel intersection with the BAS (CV ∼ −6 Å) up to CV ∼ 3 Å (being solely interested in the diffusion barrier, the actual diffusion in the second intersection was not considered), whose $CV_{0,i}$ and κ_i parameters are reported in Table S4. These parameters were kept the same independently of the water loading in the unit cell. An AIMD was then performed in each one of the umbrellas. To speed up the simulations, a switch was made to the NVT ensemble, where the average cell lengths of the 653 K MD simulation of 4-EP were used (20.22, 20.40, 13.60 Å). Cell angles were fixed to 90°, as the average values were negligibly different from 90° (<0.1°) and an orthorhombic unit cell symmetry remarkably reduces the computational time in CP2K.

A 100 ps long MD simulation was performed for each umbrella and the first 2.5 ps was excluded as equilibration time. To obtain reliable statistics out of the simulations, it is important that the CV time series are subsampled to decorrelate the datapoints. This was done using the Pymbar library.[72,73] The molecular motion in the zeolite channel is a slow process and, therefore, very high statistical inefficiencies were obtained (∼0.3 ps). Hence the long simulation time per umbrella. The decorrelated time series were then used to obtain the final free-energy profiles (FEPs) through the weighted histogram analysis method (WHAM), as implemented in our in-house-developed ThermoLIB library.[74] Additional benchmark calculations were performed to validate that the resulting FEPs negligibly depend on the plane-waves cutoff (Section S5.2). The error estimate on the FEP was also computed with the ThermoLIB using the Fisher information matrix of the maximum likelihood estimator (MLE).[74] An additional set of US simulations were performed to investigate the impact of water on TS2 for 4-EP, 4-*n*-PP, and 4-*iso*-PP, as in all cases, it was found to be the most favorable dealkylation path (*vide infra*). The related computational details are extensively reported in Section S4.1.

3. RESULTS AND DISCUSSION

3.1. Mobility and Stability of Reaction Intermediates at *Operando* Conditions.

As explained in the introduction (Figure 2), various intermediates can be expected to occur in the H-ZSM-5 catalyst during alkylphenols dealkylation. To gain preliminary insights into their behavior and stability at *operando* conditions, AIMD simulations were performed on the reactive alkylphenols, on the *para*-protonated Wheland complexes and on the SAS that can be formed during the dealkylation of 4-EP, 4-*n*-PP, and 4-*iso*-PP. Moreover, alkyl cations corresponding to the protonated product alkenes were also simulated, to understand whether their existence might be possible at relatively high reaction temperatures. Each simulation was carried out for 40 ps at a temperature of 653 K and/or 900 K. The former corresponds to the one at which the least reactive 4-EP reaches full conversion experimentally,[14]

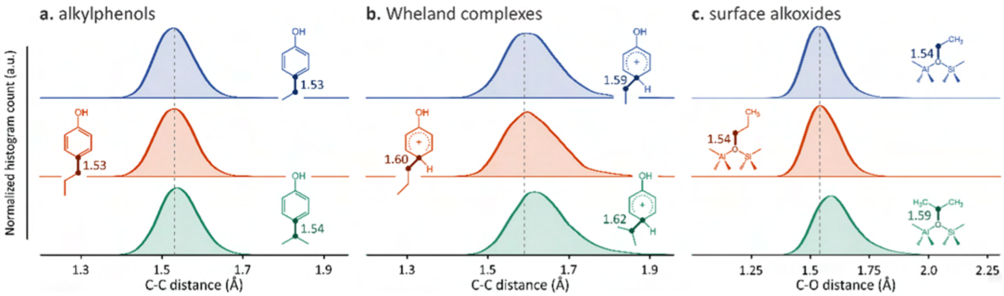


Figure 4. Wheland complex and surface alkoxide intermediates of 4-*iso*-PP present, on average, an elongated reactive bond. This can be seen in the normalized histogram distributions of relevant bond lengths (highlighted with a dark color in the structures drawings) observed in the 653 K MD simulations of (a) neutral alkylphenols, (b) Wheland complexes, and (c) surface alkoxides of 4-EP (blue), 4-*n*-PP (red), and 4-*iso*-PP (green). The dotted gray line is centered at the maximum of the 4-EP distribution and it is only meant to guide the reader's eye. The values reported near the structures drawings are the bond lengths at which the various distributions reach their maximum.

while the latter was selected to increase the probability of sampling rare events during the simulation time.

3.1.1. 4-Alkylphenols. A single molecule of 4-EP, 4-*n*-PP, or 4-*iso*-PP was introduced in the zeolite channel intersection, close to the BAS location. All of them resulted to be stable during the simulation at 653 K. The observed mobility in the zeolite pore system is low in all cases, as none was able to diffuse away from the channel intersection (see Figure S2). 4-EP and 4-*n*-PP positioned themselves with the alkyl tail aligned along the straight channel. This configuration allows them to strongly interact with the BAS, mainly through a hydrogen bond with their hydroxyl group. Despite such hydrogen bond being the prevalent interaction, the molecules have enough freedom to slightly move back and forth along the straight channel, occasionally exposing the aromatic π cloud to the BAS (Figure S3a,b). The possibility of the molecule to interact with the BAS through the aromatic π cloud is important as the transfer of the BAS proton to the ring represents the initial step of the dealkylation reaction.

4-*iso*-PP behaves quite differently. Its isopropyl tail, indeed, is too hindered to be easily accommodated into the straight channel and pushes the aromatic ring forward (Figure S2c). This prevents the formation of a strong hydrogen bond between the BAS (mainly because of its location in our model) and the OH group of 4-*iso*-PP, with an O···H distance constricted between 2 and 3 Å (Figure S3c). On the other hand, the hindrance of the isopropyl tail has a low impact on the C−C bond involved in the dealkylation reaction, whose length distribution is fundamentally identical to 4-EP and 4-*n*-PP (Figure 4a).

In the 900 K simulations, interestingly, we observed protonations and deprotonations of the aromatic ring for 4-*n*-PP and 4-*iso*-PP (Figure S4). This indicates that the protonation of the aromatic ring is a low-activated process and an equilibrium between protonated and neutral alkylphenols may exist in the confined pores. In both cases, the *ortho* position was protonated. This can be expected as the positive charge in the ring is stabilized by the presence of the hydroxyl group. Moreover, there are two equivalent *ortho* carbon and only one *para* carbon, making the protonation on the former likelier to occur.

3.1.2. Wheland Complexes. For each of the three reactants, the relative *para*-protonated Wheland complexes were simulated, both at high and low temperatures. They are all stable during the whole simulation at a low temperature and, as the neutral reactants, their mobility is limited to the channel intersection. In the high-temperature simulations, 4-EP and 4-*n*-PP remained stable during the whole simulation. For 4-*iso*-PP, we observed proton hopping from the *para* to the *meta* carbon after 35 ps of simulation. The *meta* protonated arenium ion remained stable for 4 ps, after which the proton hopped again back on the *para* carbon. Differently from the simulation of the reactants, no deprotonation/protonation events were observed in this case, which is most likely due to the relatively short simulation times.

By analyzing the C−C bond distance of the 653 K simulations, it is observed that in all of the cases, the average bond length is longer than the neutral reactants, increasing from 1.53−1.54 to 1.59−1.62 Å (Figure 4b). Such an increase is to be expected because of the change in the *para* carbon hybridization. The 4-protonated 4-*iso*-PP has, in this case, a significantly elongated C−C bond than the other two reactants (Figure 4b), pointing toward a higher reactivity of the intermediate.

3.1.3. Surface Alkoxides. Surface alkoxide species were simulated in the presence of a phenol molecule coadsorbed in their proximity, to mimic the dealkylation intermediate **III** (Figure 2b). The differences between the reactive bond lengths become even more accentuated than for the Wheland complexes, with the surface isopropoxide having an average C−O bond length of 1.59 *vs* 1.54 Å for the ethoxide and *n*-propoxide (Figure 4c). The general increase in the analyzed bond lengths for the 4-*iso*-PP intermediates can be attributed to two main effects caused by the double substitution on the central alkyl carbon. First, the two methyl groups produce a larger steric hindrance and, second, the incipient positive charge on the central carbon is stabilized by the inductive effect of two substituents instead of one, as in the case of 4-EP and 4-*n*-PP. Interestingly, during the 900 K simulation of the surface isopropoxide, a complete breaking of the C−O bond was observed after ∼3 ps (Figure S5). The newly formed isopropyl cation remained stable for ∼4 ps, after which a proton was transferred back to the framework, leading to the formation of propene and phenol (Figure S5). This large decrease in the stability of the surface alkoxides based on the number of alkyl substituents on the framework-bounded carbon is in good agreement with the observations made by Cnudde et al. for the intermediates involved in the alkenes cracking reaction[75] and points, again, toward a higher reactivity of the 4-*iso*-PP intermediates.

A similar analysis was performed for the alkyl cations (see Section S2.3), where it was found that they are, in general, not stable. Only the isopropyl cation can survive for a limited amount of time during the molecular dynamics simulations.

In this section, we have analyzed the stability and behavior of the reaction intermediates that might be formed in the H-ZSM-5 zeolite during the dealkylation reaction of 4-EP, 4-*n*-PP, and 4-*iso*-PP. The reactants, 4-protonated Wheland complexes, and surface alkoxides represent minima in the free-energy surface at reaction conditions, thus constituting relevant reaction intermediates. Except for the likely metastability of the isopropyl cation, alkyl cations are not stable and will then not be explicitly considered in our mechanistic investigation. By looking at the bond length distribution of the reactive Wheland complexes and surface alkoxides, it can be seen how the 4-*iso*-PP intermediates present a slightly elongated reactive bond, which is attributed to steric hindrance and inductive stabilization of the incipient positive charge formed on the reactive carbon when the bond is stretched.

3.2. Free-Energy Profile of Alkylphenols Dealkylation from Static Calculations. As shown in the introduction (Figure 2b), three possible mechanisms for the dealkylation of alkylphenols are investigated based on previous studies performed on the alkylation of benzene. To summarize, the aromatic ring of the alkylphenol (**I**) is initially protonated on the *para* position (TS1) to form an intermediate Wheland complex (**II**). This intermediate can undergo a concerted dealkylation (TS2), immediately giving the final phenol and alkene products (**III**). Alternatively, the alkyl group can be transferred to the zeolite framework (TS3) forming a surface alkoxide species (**IV**) which, by concerted scission of the C−O bond and deprotonation (TS4), leads again to the final alkene product. Finally, an intramolecular mechanism is also considered, where a proton of the alkyl tail is transferred to

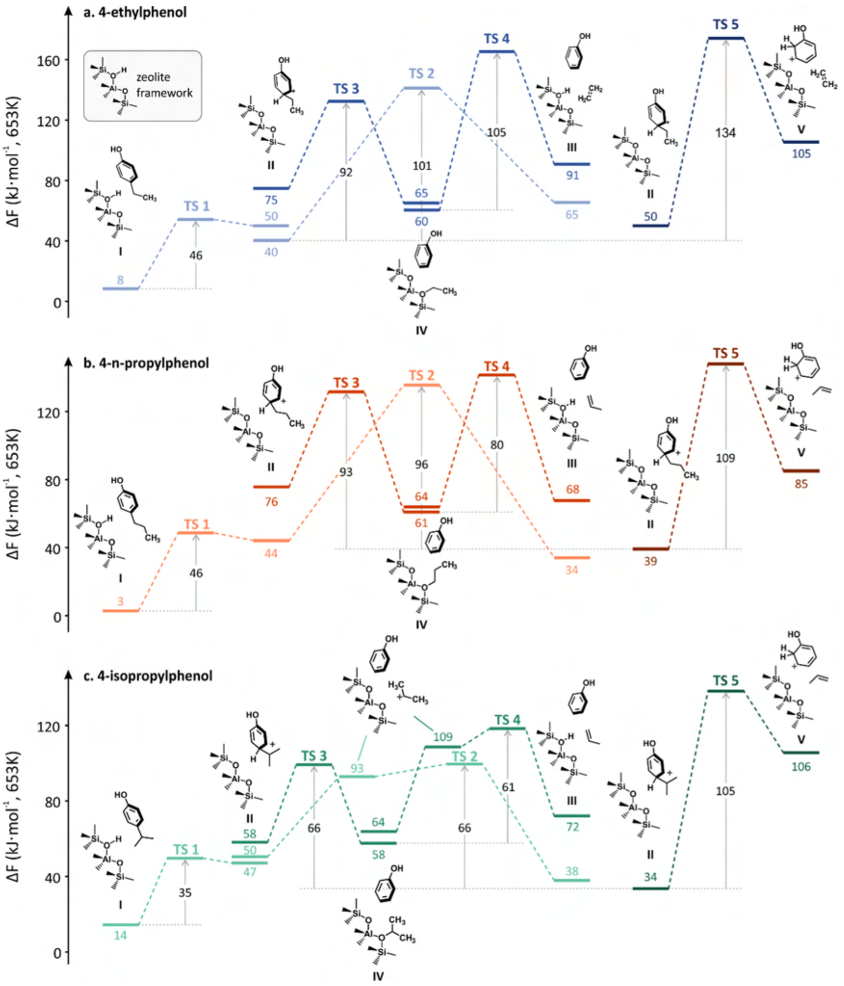


Figure 5. Free-energy profiles (653 K) for all possible dealkylation mechanisms of (a) 4-EP, (b) 4-*n*-PP, and (c) 4-*iso*-PP. The reference value for the free energies corresponds to the empty zeolites and the reacting alkylphenol in gas phase. Activation energies and free energies of the intermediates are reported in kJ·mol^{-1}. For the sake of clarity, TS5 is reported on the right of the profiles, despite its reactant state corresponding to the reactant states of TS2 and TS3.

the *ortho* position of the aromatic ring instead of the zeolite (TS5).

The energetics of all of these transition states were computed through static calculations, as well as the reactants and products connecting them for 4-EP, 4-*n*-PP, and 4-*iso*-PP, as described in Section 2. The free-energy profiles for the three species are shown in Figure 5 and the explicit activation free energies are listed in Table 1. All free-energy values, together with the respective enthalpic and entropic contributions, are listed in Table S2. A graphical visualization of the optimized transition-state geometries is provided in Figure 6.

For all reactants, the protonation of the aromatic ring (TS1) is relatively low activated, with barriers of 46 kJ·mol^{-1} for 4-EP and 4-*n*-PP and 35 kJ·mol^{-1} for 4-*iso*-PP, in good agreement with the observations made from the MD simulations. The Wheland complex was always localized as a stable intermediate; however, its energy can vary substantially, *e.g.*, variations up to 35 kJ·mol^{-1} in the case of 4-EP, depending on its specific orientation in the zeolite framework. Such large dependencies of the computed energies on the orientation of a molecule in the zeolite framework is well-known for static calculations.[58] Here, this difference in orientation originates from the fact that the reactant and product geometries are relaxed directly from the respective transition state, thus ending up in different local minima on the flat potential energy surface of the channel intersection. A clear example is the

Table 1. Forward ($\Delta F_f^\ddagger$) and Backward ($\Delta F_b^\ddagger$) Activation Free Energies ($kJ \cdot mol^{-1}$) for the Various Transition States of the Dealkylation Process, as Obtained from Static Calculations[a]

	TS1		TS2		TS3		TS4		TS5		
reactant	$\Delta F_f^\ddagger$	$\Delta F_b^\ddagger$	$\Delta F_f^\ddagger$	$\Delta F_b^\ddagger$	$\Delta F_f^\ddagger$	$\Delta F_b^\ddagger$	$\Delta F_f^\ddagger$	$\Delta F_b^\ddagger$	$\Delta F_f^\ddagger$	$\Delta F_b^\ddagger$	$\Delta F_{tot}^\ddagger$
4-EP	46	4	101	76	58	67	105	74	124	69	141
4-*n*-PP	46	4	91	101	56	71	77	74	109	63	135
4-*iso*-PP	35	3	49[b]	62[b]	41	42	54[b]	46[b]	105	33	100
4-EC	56	−2[c]	95	89	47	61			129	70	137

[a] $\Delta F_{tot}^\ddagger$ is the overall free-energy barrier to overcome from the reference gas-phase molecules to TS2, which in all cases is the lowest activated. Note that these values refer to the reactant and product states obtained from the TS optimization and, therefore, do not necessarily correspond with the barriers reported in Figures 5 and 7. [b] Referred to the actual reactants, not the metastable isopropyl cation found as intermediate minimum. [c] Artifact due to the frequency calculation (see main text).

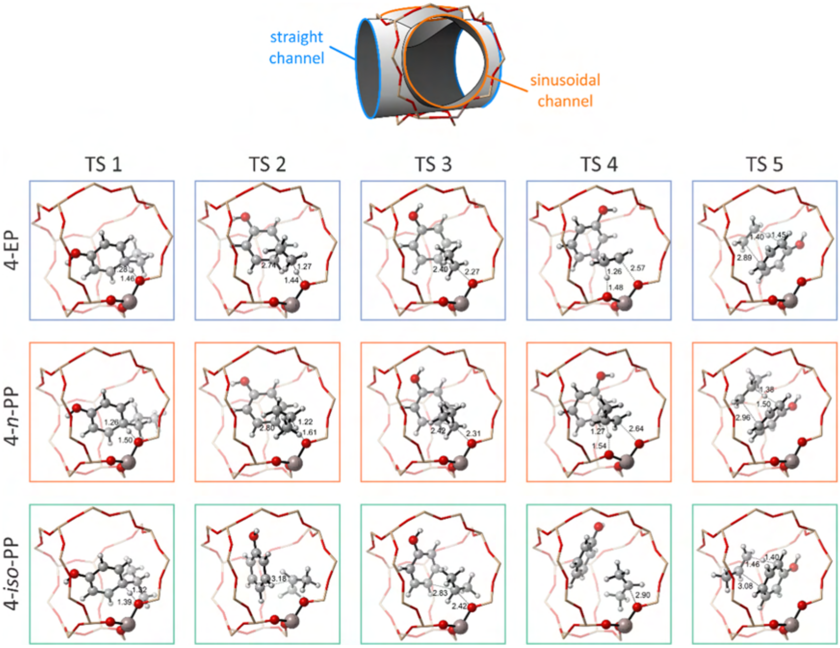


Figure 6. Relaxed transition-state geometries for the dealkylation of 4-EP, 4-*n*-PP, and 4-*iso*-PP. Relevant distances are reported in Å. For the sake of clarity, only the portion of framework surrounding the channel intersection is shown; a schematic representation of the straight and sinusoidal channel orientations is provided in the top figure.

reactant state of TS3 that, in all cases, is higher in energy than the reactant state of TS2 and the product state of TS1. The reason for this is that, to expose the alkyl tail toward the framework, the Wheland complex must rotate producing a hindered configuration where the acidic proton on the *para* carbon is no longer interacting with the oxygens in the first coordination sphere of the Al defect (compare with Figure 6). For the discussion of the reaction barriers hereafter, we systematically used the lowest energetic structure as reference (Figure 5). Furthermore, we opted to consider similar orientations for analogous transition states across the various reactants to allow a systematic comparison between the various alkylphenolic compounds. It can indeed be reasonably assumed that the interconversion between the local minima obtained when optimizing from different transition states is much faster than overcoming the actual reaction barriers.

For the dealkylation of 4-EP (Figure 5a), the concerted dealkylation (TS2) is somehow favorite with respect to the other two investigated mechanisms. With a computed free-energy barrier of 101 $kJ \cdot mol^{-1}$ (referred to as the Wheland complex **II**), the rate is slower than the formation of a surface ethoxide species (SES) in the stepwise mechanism (TS3, 92 $kJ \cdot mol^{-1}$). However, forming the SES is also an endergonic process (+20 $kJ \cdot mol^{-1}$) and its subsequent deprotonation (TS4) presents a barrier of 105 $kJ \cdot mol^{-1}$ (referred to as the surface ethoxide intermediate **IV**). Therefore, an overall barrier of 125 $kJ \cdot mol^{-1}$ must be overcome to move from the Wheland complex (**II**) to the products (**III**) through the stepwise path. Finally, the intramolecular mechanism (TS5) has the highest activation energy (134 kJ·mol^{-1}) and the deprotonation of the *ortho*-protonated phenol formed in the reaction was then no longer investigated. Overall, the main mechanistic features of the reaction are preserved with respect to the alkylation of benzene with ethene, for which Wang et al.[29] obtained a qualitatively similar profile using static simulations at a PBE-D3 level of theory.

Similar considerations can be made for the reaction profile of 4-*n*-PP (Figure 5b). Interestingly, the reaction barriers related to the C−C bond scission (TS2, TS4, and TS5) are all lower than in the 4-EP case (−5 kJ·mol^{-1} for TS2, −25 kJ·mol^{-1} for TS4 and TS5). As expected, forming a more substituted alkene (isopropene *vs* ethene) is energetically advantageous. Despite its comparably lower stabilization, TS2 remains the most favorable dealkylation path, with an activation energy of 96 kJ·mol^{-1}. 102 kJ·mol^{-1} is needed to reach TS4 from the most stable Wheland complex state, while the intramolecular TS5 has again the highest activation free energy, *i.e.*, 109 kJ·mol^{-1}. The transition-state geometries are quite similar to 4-EP, where we however observed a small but systematic increase in the C−C or C−O bond that must be broken to form isopropene (0.06−0.07 Å, Figure 6)

When comparing the profiles for 4-EP and 4-*n*-PP, our results show that when starting from the neutral adsorbed reactants, 133 kJ·mol^{-1} is needed to reach TS2 for 4-EP and 132 kJ·mol^{-1} for 4-*n*-PP (141 and 135 kJ·mol^{-1} starting from the gas-phase reactants, Table 1). This difference is very small and largely within the typical DFT uncertainty. However, when comparing the reaction barriers, there is a clear decrease in the barriers when moving from 4-EP to 4-*n*-PP, which is in line with the moderate reduction in temperature required experimentally to reach 50% conversion (from ∼650 K for 4-EP to ∼575 K for 4-*n*-PP). On the other hand, experimental kinetic analysis of the reaction showed that 4-EP has an apparent activation energy of 60±3 kJ·mol^{-1}, while 4-*n*-PP has an apparent activation energy of 98 ± 6 kJ·mol^{-1}, with the observed increase in rate mainly due to entropic reasons.[15] This large difference was tentatively attributed to the different heat of adsorption between the two molecules, as when considering their dehydroxylated counterparts, ethylbenzene has a higher heat of adsorption than *n*-propylbenzene.[76,77] This large difference in adsorption energy does not clearly emerge from our computations, where 4-*n*-PP has an adsorption enthalpy of −135 kJ·mol^{-1}, while 4-EP has an adsorption enthalpy of −121 kJ·mol^{-1} (Table S1), a difference mainly originating from the difference in dispersive interactions. It seems then that other factors must be into play when comparing the theory results with the experimental kinetic data. The likelier source of error could lie in our estimate of the adsorption enthalpy, which might be heavily biased when working in the harmonic approximation and accounting for a single conformation of a relatively mobile adsorbate.[78] In principle, one could go beyond this approach using MD simulations and a higher level of theory to more accurately account for long-range interactions.[79,80] This is however well beyond the scope of the present study.

Finally, 4-*iso*-PP presents, in general, very reduced barriers (green scheme in Figure 5) with respect to the other two reactants. Only 66 kJ·mol^{-1} is needed to activate TS2 and TS3, while TS4 proceeds with an activation energy of 61 kJ·mol^{-1}. As in the previous cases, the formation of the surface alkoxide is endergonic, leading to an overall barrier of 85 kJ·mol^{-1} to undergo stepwise dealkylation. The intramolecular dealkylation still has a very high activation energy compared to the other mechanisms (105 kJ·mol^{-1}). Therefore, also in this case, the concerted path is computed to be the lowest activated. Furthermore, the large decrease in activation energy (only 100 kJ·mol^{-1} separate the gas-phase reactants from TS2, see Table 1) agrees very well with the experiment, where almost full conversion was reached already at less than 500 K. The transition-state geometries of TS2, TS4, and TS5 present very elongated bond lengths (∼3 Å, Figure 6) indicating—not surprisingly—a larger charge accumulation on the more substituted reactive carbon. The increase in stability moving from primary to secondary carbenium ions is sufficient to create, for both TS2 and TS4, a metastable isopropyl cation moiety that was formed in the optimization toward the reactant basin (Figure 5). Frequency analysis revealed that this is indeed a minimum. However, a minimal perturbation of the C−C (for TS2) or C−O (for TS4) bond length followed by a geometry optimization brings the system in the expected reactant state, confirming that the isopropyl cation minimum is extremely shallow.

Our calculations show overall a good qualitative agreement with the reactivity differences observed experimentally. For 4-*iso*-PP, the pronounced reactivity is easily explained by the stabilization of the transition states caused by the double hyperconjugation with the methyl groups. The stabilization is strong enough to make the isopropyl cation a possible metastable intermediate for the reaction. In the 4-*n*-PP case, the methyl group cannot directly conjugate with the carbon on which the partial positive charge is located. Nevertheless, it can still interact with the forming π orbital of the incipient double bond, stabilizing both the transition state and the produced substituted olefin. While this is possibly not the main reason for the increase in reactivity observed experimentally, it indicates that 4-*n*-PP should be intrinsically slightly more reactive that 4-EP toward dealkylation. A similar trend was also observed by Arstad et al.[26] for the alkylation of benzene with ethene and propene over a cluster model of the zeolite.

From a mechanistic perspective, it is very likely for all of the investigated reactants to undergo concerted dealkylation (TS2), as the energy requirement is in all cases lower than TS4. Differently from the ethylation of benzene, the concerted path does not require the substrate and ethylating agent to come together in proximity of the active site, which was found to potentially hamper the reaction progress and favor the stepwise path.[29] On the other hand, SAS formation (in equilibrium with the Wheland complex and neutral alkylphenol) seems plausible, TS3 being similar in energy with respect to TS2, but the endergonic character of the reaction makes its concentration in the pores likely small. This is potentially different than in benzene ethylation, where the stepwise path becomes more favorable at high ethylating agent concentrations and where the presence of surface ethoxide species was detected through NMR spectroscopy.[32]

While demethoxylation of the lignin compounds can lead to the alkylphenols investigated so far (Figure 1), it has also recently been shown that the alkylcatechol compounds obtained after O-demethylation of the lignin-derived alkylguaiacols can also be C-dealkylated in a similar fashion.[17] To understand the effect of a second hydroxyl substituent on the aromatic ring, we also recomputed the dealkylation mechanism of 4-EC. All transition states were reoptimized except for TS4, as it does not directly involve the aromatic moiety and thus we did not expect any significant change in its energy. The free-

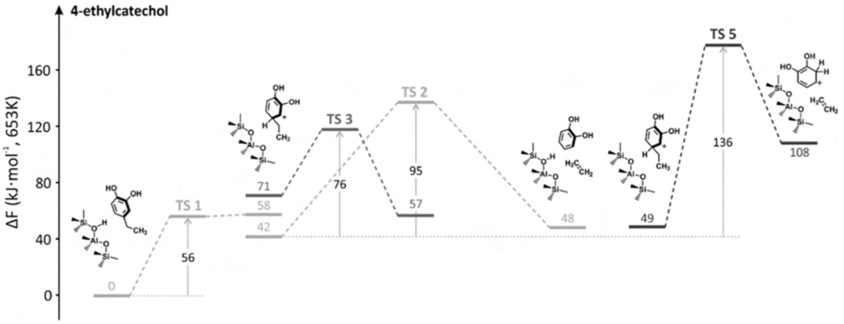


Figure 7. Free-energy profiles (653 K) for all possible dealkylation mechanisms of 4-EC. The reference value for the free energies corresponds to the empty zeolites and the reacting 4-EC in gas phase. Activation energies and free energies of the intermediates are reported in kJ·mol^{-1}. For the sake of clarity, TS5 is reported on the right of the profiles, despite its reactant state corresponding to the reactant states of TS2 and TS3.

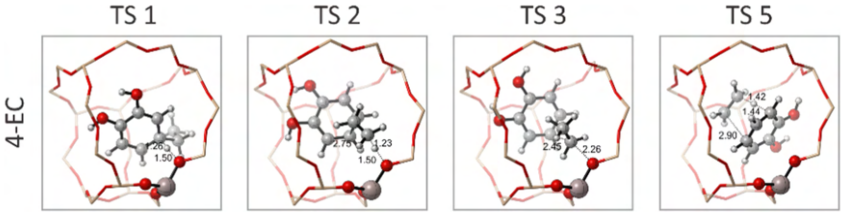


Figure 8. Relaxed transition-state geometries for the dealkylation of 4-EC. Relevant distances are reported in Å. The shown framework portion is the same as in Figure 6.

energy profile is shown in Figure 7, while the geometry of the transition states is shown in Figure 8.

The protonation step (TS1) has a moderately higher activation energy with respect to 4-EP (56 *vs* 46 kJ·mol^{-1}). The free energies of the products of TS1 are slightly higher in energy than the transition state (Figure 7), which is an artifact caused by the approximations in the frequency calculations on one particular configuration of the products resulting from TS1. Indeed, when comparing the 0 K electronic energy of the protonated arenium ion, the product is indeed lower in energy than the transition state.

TS3 is surprisingly low in energy, with a barrier of only 76 kJ·mol^{-1}. The C−C distance is somehow more stretched than in the 4-EP case (2.45 *vs* 2.40 Å, compare Figures 6 and 8), while the C−O distance remains almost unaffected. However, the product SAS remains higher in energy (+15 kJ·mol^{-1}) compared to the most stable Wheland intermediate. This value, combined with the 4-EP activation energy of TS4, leads to an overall barrier of 120 kJ·mol^{-1} to overcome in the stepwise path. Therefore, also here, the concerted pathway would be preferred.

When comparing the overall activation energies for 4-EC and 4-EP, the differences are relatively small (137 *vs* 132 kJ·mol^{-1} with respect to the neutral reactants are found for 4-EC and 4-EP, respectively) and certainly well beyond the methodological accuracy. Therefore, we can conclude that the effect of a second hydroxyl group on the aromatic ring—if any—is very small. This is in line with the experimental observations, where the dealkylation of 4-*n*-propylcatechol occurs at analogous temperature as the dealkylation of 4-*n*-PP.

Before moving on with the final section, some interesting observations about the connection between our calculations and the realistic reaction environment can be made, in particular, concerning the water coverage. Because of their strong interaction with the BAS, adsorbed water molecules can interfere with the kinetics of proton transfer reactions, both by acting as proton shuttling medium as well as by fully solvating the BAS as hydronium ion (compare with the following section). Recently, we indeed showed that low water coverages (1−3 molecules per BAS) can significantly speed up the protonation reaction of aromatic hydrocarbons in H-ZSM-5, helping the proton shuttling from the framework to the molecule.[20] At higher coverages (6 molecules per BAS), the rate goes back to similar values as in the anhydrous case. Since there are no reasons to suspect very different behaviors for the four considered reactants, it seems plausible that the kinetics of TS1 will be affected in a similar fashion by the presence of coadsorbed water. Therefore, even if the absolute values of the free-energy barriers could change, it seems unlikely that this will affect the computed reactivity trends.

The possible water effect is less clear on the actual dealkylation transition states. For this reason, TS2 (for 4-EP, 4-*n*-PP, and 4-*iso*-PP) was further investigated by means of dynamic US simulations in the presence of three water molecules coadsorbed in the zeolite unit cell. For a complete discussion of the results, the interested reader is referred to Section S4. While the reactivity trends across the reactants do not seem to be particularly affected, it was found that the dealkylation reaction competes with the deprotonation of the Wheland complex hydroxyl group to form a 4-alkyl-2,5-cyclohexadienone moiety. The deprotonation occurs preferen-

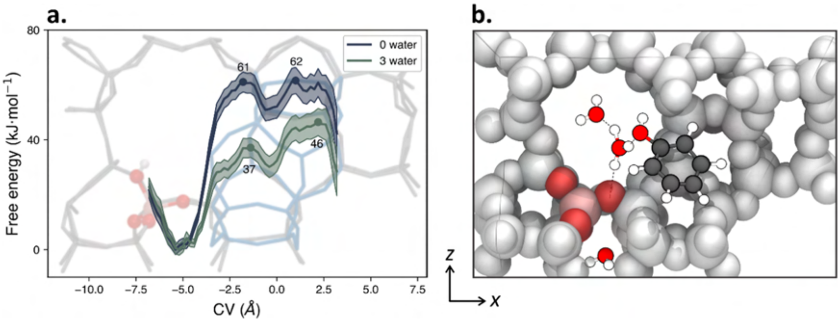


Figure 9. (a) Free-energy profile for the diffusion of phenol along the straight channel of H-ZSM-5 with zero and three water molecules in the zeolite unit cell. The image in the background shows the rough location of the various CV values within the framework, where the Al tetrahedra and BAS location are highlighted in ball-and-stick representation, while the double ring separating the two channel intersections in light blue. (b) Snapshot arbitrarily extracted from the umbrella centered at CV = −1.5, where the H-ZSM-5 unit cell is seen along the straight channel. It is possible to see how the water maintains an interaction with the phenol hydroxyl group, thereby assisting the diffusion.

tially when multiple water molecules interact with the hydroxyl group, in line with the higher proton affinity of water clusters with high molecularity.[20,35,36] The deprotonated form is not particularly stable ($\Delta F \sim +70$ kJ·mol^{-1}) and it is indeed not observed in the 4-*iso*-PP case where the dealkylation barrier is lower in magnitude. On the other hand, it is possible that at higher water loadings, the O-deprotonation of the reactants could become more favorable. Since the neutral form cannot undergo dealkylation, an equilibrium between the Wheland complex and 4-alkyl-2,5-cyclohexadienone could be tentatively indicated as a cause for the increase in Arrhenius activation energy for the 4-*n*-PP dealkylation observed experimentally in the presence of water.[15]

3.3. Effect of Water on Phenol Diffusion. Differently from traditional hydrocarbon chemistry, alkylphenols dealkylation requires to cofeed steam in the reactants stream to prevent a quick and reversible deactivation of the catalyst. This was attributed to water preventing the condensation of phenol to diphenylether, but it was also speculated that it could facilitate desorption of phenol from the catalyst. The reaction barriers computed for the diphenylether formation mechanism in anhydrous conditions are of similar magnitude as the dealkylation ones, in line with a fast build-up of the molecule in the catalyst pores (for a complete overview of the results, see Section S6). Since water is formed as the reaction product, the presence of steam will help to shift the reaction equilibrium toward the reacting phenols. On the other hand, it is less clear as to how far water could enhance the diffusion and desorption of the latter from the zeolite framework. Therefore, the role of water in modulating phenol diffusion along the straight channels of H-ZSM-5 was investigated by means of four separate umbrella sampling simulations, with water loadings varying from 0 to 3 water molecules per unit cell. Larger water clusters (up to 8 molecules/BAS) have been observed in the zeolite at high water partial pressures and low temperatures[81] and water is also known to adsorb quite nonhomogeneously on the BAS.[82] Nonetheless, the free-energy profiles with two and three water molecules are fundamentally indistinguishable, making it unlikely that a higher coverage would significantly affect the results (*vide infra*). Since the results related to the simulations with one and two water molecules were found to be intermediate between the zero and three water cases, only the latter will be discussed in detail. For a complete overview of the results across all loadings, the interested reader is referred to Section S5.4. Diffusion through the sinusoidal channel was not considered as it is known to be significantly slower, based on both theoretical and experimental evidence.[37,83] For instance, based on a set of static simulations, DeLuca and Hibbits found that the diffusion barrier for toluene in all-silica MFI increases from 15 to 56 kJ·mol^{-1} when moving from the straight to the sinusoidal channel.[37]

The obtained free-energy profiles for phenol diffusion along the H-ZSM-5 straight channel with zero and three water molecules are shown in Figure 9a.

The main feature that emerges from the simulated FEPs is the significant reduction in the diffusion barrier computed in the presence of water. The barrier to diffuse through the straight channel in the anhydrous case amounts to 61 kJ·mol^{-1}, whereas in the presence of three water molecules in the zeolite unit cell, the barrier is reduced to only 37 kJ·mol^{-1}. Importantly, phenol also has to desorb from the BAS to travel through the straight channel, an effect contributing to the overall free-energy profile. When traveling through the straight channel, the molecule encounters two skewed 10-membered rings, which separate the channel intersections. The maxima in the profile originate from the actual crossing of the two 10-membered rings. In both the anhydrous and hydrated cases, the FEP has a shallow minimum located in between the two skewed 10-membered rings. A similar behavior has been recently reported for aromatic diffusion in all-silica MFI by DeLuca and Hibbitts,[37] where two minima around the channel center were observed instead for benzene diffusion. The presence of a single minimum in our FEP is likely due to the dynamic nature of the simulation and the asymmetric nature of phenol. Indeed, the phenol oxygen also being included in the computation of the molecule center (Figure 3), its orientation can produce different CV values that are associated with a similar position of the aromatic ring in the channel. This effect can produce a "blurring" when moving from well-defined states on the PES to an averaged FEP.

When comparing the two FEPs, it is observed that the profile is quite symmetric in the anhydrous case and that the second transition state to travel through the second 10-membered ring is nearly isoenergetic with the first (61 *vs* 62 kJ·

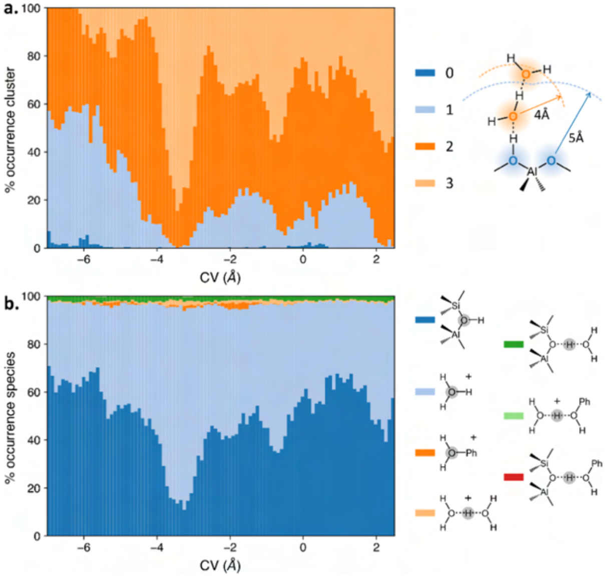


Figure 10. (a) Percentage of occurrence of a certain water cluster size in proximity of the active site as a function of the US CV. The scheme on the right sketches how the cluster size is computed, by first searching for water molecules within a 5 Å cutoff from the zeolite oxygen atoms in the first coordination sphere of the Al site and then, starting from the closer water, looking for other water molecules within 4 Å and so on until no further water molecules can be found. (b) Percentage of occurrence of the various possible protonated species during the US simulations as a function of the US CV. The atom highlighted in gray in the legend shows the location of the fictitious proton atom in various cases (see main text). In both cases, the trajectory is subsampled every 2.5 fs.

mol^{-1}). This seems to indicate that, once the phenol reaches the first barrier, its interaction with the BAS has already completely vanished and diffusing even further from it does not require additional energy. On the other hand, when water is present in the unit cell, the FEP becomes significantly asymmetric and the barrier to cross the second ring increases to 46 kJ·mol^{-1}. When looking at the simulations, it was noticed that water tends to form a chain of hydrogen bonds between the active site and the diffusing phenol molecule (Figure 9b). Only when moving even further along the straight channel (CV > 0), this chain of H-bonds begins to get stretched, causing the moderate energy increase to cross the second ring. We remark that previous considerations are mostly qualitative as the actual energy barriers might vary due to error bars of the simulations. Despite not being directly comparable, the activation free energies for phenol diffusion computed in the presence of water are largely lower than the static (but also dynamic, see Section S4) barrier of the dealkylation process for 4-EP and 4-*n*-PP, in line with the experimental observation that the reaction kinetics does not seem to be diffusion-limited for 4-*n*-PP.[15] This might not be the case for 4-*iso*-PP, as the hindrance of the isopropyl tail reduces its mobility in the framework (Figure S3) while, at the same time, leading to reduced dealkylation barriers.

To better understand the reasons underlying the large difference in diffusion barriers with different water loadings, we assessed whether the water molecules (whose position is not restrained during the simulations) do remain in proximity of the BAS and the diffusing phenol. A cutoff radius of 5 Å was defined around the O atoms in the first coordination sphere of the Al defect. If no water molecules are found within such a radius, the cluster size is set to 0. Otherwise, the closest water molecule is taken as reference and a second cutoff radius of 4 Å is defined around it. All water molecules included in such a cutoff are then added to the cluster size and are then used as new centers to look for other water molecules until no others are found. This procedure was done for every umbrella trajectory (subsampling it every 2.5 fs to make the amount of data more manageable) and the percentage of occurrence of each cluster size reported as a function of the US CV (Figure 10a). A cluster size of 0 is almost never observed during the simulations, indicating that water strongly interacts with the BAS—as expected. In the simulations with one water molecule, it is more likely for the water to diffuse away from the active site. On the other hand, even for low CV values, the phenol molecule is not capable of completely displacing the water molecule, which is far from the Al substitution at most 60–70% of the times for CV < −6 Å (Figure S13a). This percentage is largely reduced moving to two waters and basically disappears with three waters, where at least one water molecule remains in the BAS proximity for more than 90% of the simulation time at any considered CV value. When phenol moves away from the active site (increasing CV values), water does not have any competition to access the BAS and cluster sizes of larger molecularity become predominant. Quite some fluctuation is observed in the cluster sizes while moving toward higher CV values. We attribute this to the fact that the data comes from separate umbrellas, and in some of them, larger cluster sizes are preserved most of the simulation time, while for the others, more diffusive events are observed. This is very likely a random process and, water diffusion being relatively slow with respect to our simulation times, it is not possible to achieve a fully equilibrated distribution of cluster sizes in all umbrellas. The observation of various cluster sizes also indicates that considering more molecules per unit cell is not necessary as, if much larger cluster sizes would be favorable, we would almost never observe the three molecules diffusing away from the active site.

Having assessed that water remains in proximity of the active site during the simulations, we investigated how the BAS interacts with the protic molecules present in the unit cell. We first defined the location of a fictitious proton atom using a combination of the collective variables proposed by Pérez de Alba Ortiź et al.[84] and Grifoni et al.[85] (for more information, see Section S5.3). In this approach, such a fictitious atom is located on the oxygen carrying an excess of hydrogen atoms with respect to a selected default value (for instance, 2 for a water oxygen or 0 for a zeolite oxygen). To practically visualize how the location of the proton atom evolves during a simulation, we invite the interested reader to refer to the movie in the Supporting Information, which shows the 40−50 ps interval of the umbrella centered at $CV_0 = 0.0$ with three water molecules. We then computed, in each simulation, where this proton atom is located with respect to the proton-acceptor species and used this data to obtain the percentage of occurrence of a determined protonated species (Figure 10b).

What immediately emerges from the figure is that two states dominate the simulation in all of the umbrellas, with the extra proton being either located on the zeolite (thus a "regular" BAS) or on a water molecule, forming an eigen-like hydronium ion (H_3O^+). The balance between the two is dictated by the amount of water molecules in proximity of the active site (compare the similar up-down trends in the top and bottom graphs of Figure 10 and with the results obtained at different water loadings in Figure S13b) as it is well-known that at least two water molecules are required to lead to a significant solvation of the BAS.[35,36] Phenol seems, on the other hand, to have a much more limited interaction with the BAS. As it can be seen, the instances in which its hydroxyl group is protonated are very infrequent (<10%) even when in close proximity of the Al site.

Interestingly, it can be seen how the calculations predict that the proton is preferentially located on a specific atom and the sampling of structures in which it is shared between two atoms (Zundel-like ions) is very small independently on the phenol location. According to these results, phenol and water are then in direct competition for the zeolite active site and this will certainly favor the desorption and diffusion of the former. Further insights in this behavior were obtained by computing the adsorption enthalpies for coadsorption of phenol and one water molecule with static simulations (an extensive discussion of the results is reported in Section S5.5). The computed adsorption enthalpy for phenol decreases from 132 to 113 $kJ\cdot mol^{-1}$ if water is coadsorbed on the BAS while, at the same time, the adsorption enthalpy of water also decreases from 83 to 64 $kJ\cdot mol^{-1}$ if phenol is coadsorbed on the BAS. This suggests that competitive coadsorption penalizes both molecules with respect to a single-molecule adsorption and thus not only does water enhance the kinetics of phenol diffusion but it also modulates its adsorption thermodynamics.

In Figure 10b, it can be seen that protonated phenol seldomly occurs up to relatively high CV values, where a direct transfer of the BAS from the framework would no longer be possible. This is caused by water that, as the phenol diffuses away from the active site, creates a chain of hydrogen bonds along which the extra proton is allowed to move (Figure 9b). Furthermore, spontaneous *ortho* protonation of phenol was observed in the anhydrous case when located close to the BAS (in umbrellas 1 and 2, see Table S1) after about 10 ps of simulation and the so-formed arenium ion proved to be stable for the rest of the simulation. Phenol protonation was also observed with one and two water molecules in the unit cell, but not with three (Figure S13b), indicating that larger loadings could also have an active effect on the formation of protonated phenolic species, as already highlighted in the case of ethylbenzene protonation.[20]

Calculations then suggest that not only does water compete with phenol to directly adsorb on the BAS, thereby facilitating its desorption, but it also assists its diffusion through the straight zeolite channel by forming a chain of H-bonds thanks to which the phenol hydroxyl group does not have to completely break its interaction with the active site when crossing between the two channel intersections.

4. CONCLUSIONS

In this work, we have explored the two main features of the H-ZSM-5-catalyzed alkylphenol dealkylation reaction. Through a combination of static and dynamic DFT simulations, we showed that 4-*iso*-PP is much more reactive than 4-*n*-PP and 4-EP, as the secondary nature of the central carbon atom of the isopropyl tail stabilizes the (partially) positively charged transition states. 4-*n*-PP presents somehow lower activation energies than 4-EP, in line with the experimentally observed reactivity trends, although the latter also suggests the decrease in reactivity to be mainly entropic in nature. This could be due to contributions deriving from the adsorption step of the reactant in the zeolite, which is however outside the scope of this contribution. Mechanistically, we found that the reaction proceeds mainly through a concerted mechanism, being characterized by the lowest computed activation energy; however, the formation of surface alkoxide species is also possible in low amounts. Umbrella sampling simulations of the dealkylation reaction in the presence of three water molecules showed that the intermediate Wheland complex can be O-deprotonated by the water cluster. This reaction competes with dealkylation for 4-EP and 4-*n*-PP, being possibly one of the reasons underlying the higher activation energies measured experimentally in the presence of steam.

Umbrella sampling was also used to investigate the mobility of the phenol product along the straight zeolite channel in anhydrous conditions and with one, two, and three water molecules in the unit cell. We found that water may play an important role in assisting phenol desorption and diffusion away from the zeolite BAS, with a computed decrease in the free-energy barrier from 62 to 46 $kJ\cdot mol^{-1}$ when moving from zero to three coadsorbed molecules. This is caused by the competitive adsorption of water and phenol on the BAS. Water is shown to preferentially interact with the BAS, thereby effectively shielding it from phenol. On the other hand, it also appears that a chain of water-mediated hydrogen bonds between the BAS and the phenol can be formed while the latter is diffusing, which could assist its passage between the two channel intersections. The assistance of water in desorption and diffusion away from the BAS could lead to shorter residence times of phenolic compounds in the zeolite, preventing their condensation to bulkier biphenylether species. It would be interesting to explore in future studies as to how far this behavior is extendable to other (moderately) bulky and polar molecules diffusing along the straight channels of the H-ZSM-5 zeolite and to provide further atomistic details in the role played by water in modulating the kinetics of relevant zeolite-catalyzed reactions.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscatal.2c03844.

> Catalyst model, extended details on the behavior and stability of reaction intermediates from the unbiased MD simulations, explicit values of the free energy (with enthalpic and entropic contributions) for all of the optimized stationary points, and extended details on the US simulations of TS2 in the presence of three water molecules on the diffusion of phenol with one and two water molecules and on the mechanism of phenol condensation to diphenylether (PDF)
>
> Movie showing the location of the fictitious proton atom (orange sphere) in the 40−50 ps interval taken from the simulation of the umbrella centered at CV_0 = 0.0 Å (MP4)
>
> POSCAR of all of the optimized stationary points reported in the manuscript; Cartesian coordinates of the initial structures for the US simulations; CP2K and PLUMED input file examples (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Veronique Van Speybroeck − *Center for Molecular Modeling, Ghent University, 9052 Zwijnaarde, Belgium;* orcid.org/0000-0003-2206-178X;
Email: Veronique.Vanspeybroeck@Ugent.be

Author

Massimo Bocus − *Center for Molecular Modeling, Ghent University, 9052 Zwijnaarde, Belgium;* orcid.org/0000-0001-9474-6644

Complete contact information is available at:
https://pubs.acs.org/10.1021/acscatal.2c03844

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge the Fund for Scientific Research—Flanders (FWO, BioFact Excellence of Science project G0H0918N, ID EOS: 30902231, and project G024019N) as well as the Research Board of Ghent University (BOF). The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Ghent University, FWO, and the Flemish Government—Department of EWI.

■ REFERENCES

(1) Chu, S.; Majumdar, A. Opportunities and Challenges for a Sustainable Energy Future. *Nature* **2012**, *488*, 294−303.

(2) Martens, J. A.; Bogaerts, A.; De Kimpe, N.; Jacobs, P. A.; Marin, G. B.; Rabaey, K.; Saeys, M.; Verhelst, S. The Chemical Route to a Carbon Dioxide Neutral World. *ChemSusChem* **2017**, *10*, 1039−1055.

(3) Weissermel, K.; Arpe, H.-J.*Industrial Organic Chemistry*, 4th ed.; Wiley-VCH Verlag GmbH: Weinheim, Germany, 2003.

(4) Zakzeski, J.; Bruijnincx, P. C. A.; Jongerius, A. L.; Weckhuysen, B. M. The Catalytic Valorization of Lignin for the Production of Renewable Chemicals. *Chem. Rev.* **2010**, *110*, 3552−3599.

(5) Calvo-Flores, F. G.; Dobado, J. A. Lignin as Renewable Raw Material. *ChemSusChem* **2010**, *3*, 1227−1235.

(6) Li, C.; Zhao, X.; Wang, A.; Huber, G. W.; Zhang, T. Catalytic Transformation of Lignin for the Production of Chemicals and Fuels. *Chem. Rev.* **2015**, *115*, 11559−11624.

(7) Renders, T.; Van den Bossche, G.; Vangeel, T.; Van Aelst, K.; Sels, B. Reductive Catalytic Fractionation: State of the Art of the Lignin-First Biorefinery. *Curr. Opin. Biotechnol.* **2019**, *56*, 193−201.

(8) Liao, Y.; Koelewijn, S. F.; van den Bossche, G.; van Aelst, J.; van den Bosch, S.; Renders, T.; Navare, K.; Nicolaï, T.; van Aelst, K.; Maesen, M.; et al. A Sustainable Wood Biorefinery for Low-Carbon Footprint Chemicals Production. *Science* **2020**, *367*, 1385−1390.

(9) Van Den Bosch, S.; Schutyser, W.; Koelewijn, S. F.; Renders, T.; Courtin, C. M.; Sels, B. F. Tuning the Lignin Oil OH-Content with Ru and Pd Catalysts during Lignin Hydrogenolysis on Birch Wood. *Chem. Commun.* **2015**, *51*, 13158−13161.

(10) Van Den Bosch, S.; Schutyser, W.; Vanholme, R.; Driessen, T.; Koelewijn, S. F.; Renders, T.; De Meester, B.; Huijgen, W. J. J.; Dehaen, W.; Courtin, C. M.; et al. Reductive Lignocellulose Fractionation into Soluble Lignin-Derived Phenolic Monomers and Dimers and Processable Carbohydrate Pulps. *Energy Environ. Sci.* **2015**, *8*, 1748−1763.

(11) Schutyser, W.; Renders, T.; Van Den Bosch, S.; Koelewijn, S. F.; Beckham, G. T.; Sels, B. F. Chemicals from Lignin: An Interplay of Lignocellulose Fractionation, Depolymerisation, and Upgrading. *Chem. Soc. Rev.* **2018**, *47*, 852−908.

(12) Sun, Z.; Fridrich, B.; De Santi, A.; Elangovan, S.; Barta, K. Bright Side of Lignin Depolymerization: Toward New Platform Chemicals. *Chem. Rev.* **2018**, *118*, 614−678.

(13) Verboekend, D.; Liao, Y.; Schutyser, W.; Sels, B. F. Alkylphenols to Phenol and Olefins by Zeolite Catalysis: A Pathway to Valorize Raw and Fossilized Lignocellulose. *Green Chem.* **2016**, *18*, 297−306.

(14) Liao, Y.; D'Halluin, M.; Makshina, E.; Verboekend, D.; Sels, B. F. Shape Selectivity Vapor-Phase Conversion of Lignin-Derived 4-Ethylphenol to Phenol and Ethylene over Acidic Aluminosilicates: Impact of Acid Properties and Pore Constraint. *Appl. Catal., B* **2018**, *234*, 117−129.

(15) Liao, Y.; Zhong, R.; Makshina, E.; d'Halluin, M.; van Limbergen, Y.; Verboekend, D.; Sels, B. F. Propylphenol to Phenol and Propylene over Acidic Zeolites: Role of Shape Selectivity and Presence of Steam. *ACS Catal.* **2018**, *8*, 7861−7878.

(16) Liao, Y.; Zhong, R.; D'Halluin, M.; Verboekend, D.; Sels, B. F. Aromatics Production from Lignocellulosic Biomass: Shape Selective Dealkylation of Lignin-Derived Phenolics over Hierarchical ZSM-5. *ACS Sustainable Chem. Eng.* **2020**, *8*, 8713−8722.

(17) Wu, X.; Liao, Y.; Bomon, J.; Tian, G.; Bai, S.-T.; Van Aelst, K.; Zhang, Q.; Vermandel, W.; Wambacq, B.; Maes, B. U. W.; et al. Lignin-First Monomers to Catechol: Rational Cleavage of C-O and C-C Bonds over Zeolites. *ChemSusChem* **2021**, No. e202102248.

(18) Van Der Mynsbrugge, J.; Moors, S. L. C.; De Wispelaere, K.; Van Speybroeck, V. Insight into the Formation and Reactivity of Framework-Bound Methoxide Species in h-Zsm-5 from Static and Dynamic Molecular Simulations. *ChemCatChem* **2014**, *6*, 1906−1918.

(19) Nastase, S. A. F.; Cnudde, P.; Vanduyfhuys, L.; De Wispelaere, K.; Van Speybroeck, V.; Catlow, C. R. A.; Logsdail, A. J. Mechanistic Insight into the Framework Methylation of H-ZSM-5 for Varying

Methanol Loadings and Si/Al Ratios Using First-Principles Molecular Dynamics Simulations. *ACS Catal.* **2020**, *10*, 8904−8915.

(20) Bocus, M.; Vanduyfhuys, L.; De Proft, F.; Weckhuysen, B. M.; Van Speybroeck, V. Mechanistic Characterization of Zeolite-Catalyzed Aromatic Electrophilic Substitution at Realistic Operating Conditions. *JACS Au* **2022**, *2*, 502−514.

(21) Van Speybroeck, V.; De Wispelaere, K.; Van Der Mynsbrugge, J.; Vandichel, M.; Hemelsoet, K.; Waroquier, M. First Principle Chemical Kinetics in Zeolites: The Methanol-to-Olefin Process as a Case Study. *Chem. Soc. Rev.* **2014**, *43*, 7326−7357.

(22) Van Speybroeck, V.; Hemelsoet, K.; Joos, L.; Waroquier, M.; Bell, R. G.; Catlow, C. R. A. Advances in Theory and Their Application within the Field of Zeolite Chemistry. *Chem. Soc. Rev.* **2015**, *44*, 7044−7111.

(23) Jansang, B.; Nanok, T.; Limtrakul, J. Structure and Reaction Mechanism of Alkylation of Phenol with Methanol over H-FAU Zeolite: An ONIOM Study. *J. Phys. Chem. C* **2008**, *112*, 540−547.

(24) Nie, X.; Janik, M. J.; Guo, X.; Liu, X.; Song, C. Reaction Mechanism of Tert-Butylation of Phenol with Tert-Butyl Alcohol over H-β Zeolite: An ONIOM Study. *Catal. Today* **2011**, *165*, 120−128.

(25) Vos, A. M.; Schoonheydt, R. A.; De Proft, F.; Geerlings, P. Reactivity Descriptors and Rate Constants for Acid Zeolite Catalyzed Ethylation and Isopropylation of Benzene. *J. Phys. Chem. B* **2003**, *107*, 2001−2008.

(26) Arstad, B.; Kolboe, S.; Swang, O. Theoretical Investigation of Arene Alkylation by Ethene and Propene over Acidic Zeolites. *J. Phys. Chem. B* **2004**, *108*, 2300−2308.

(27) Hansen, N.; Brüggemann, T.; Keil, F. J.; Bell, A. T. Theoretical Investigation of Benzene Alkylation with Ethene over H-ZSM-5. *J. Phys. Chem. C* **2008**, *112*, 15402−15411.

(28) Hansen, N.; Kerber, T.; Sauer, J.; Bell, A. T.; Keil, F. J. Quantum Chemical Modeling of Benzene Ethylation over H-ZSM-5 Approaching Chemical Accuracy: A Hybrid MP2:DFT Study. *J. Am. Chem. Soc.* **2010**, *132*, 11525−11538.

(29) Wang, D.; Wang, C. M.; Yang, G.; Du, Y. J.; Yang, W. M. First-Principles Kinetic Study on Benzene Alkylation with Ethanol vs. Ethylene in H-ZSM-5. *J. Catal.* **2019**, *374*, 1−11.

(30) Chen, W.; Yi, X.; Liu, Z.; Tang, X.; Zheng, A. Carbocation Chemistry Confined in Zeolites: Spectroscopic and Theoretical Characterizations. *Chem. Soc. Rev.* **2022**, *51*, 4337−4385.

(31) Gong, X.; Çağlayan, M.; Ye, Y.; Liu, K.; Gascon, J.; Dutta Chowdhury, A. First-Generation Organic Reaction Intermediates in Zeolite Chemistry and Catalysis. *Chem. Rev.* **2022**, *122*, 14275−14345.

(32) Chowdhury, A. D.; Houben, K.; Whiting, G. T.; Chung, S. H.; Baldus, M.; Weckhuysen, B. M. Electrophilic Aromatic Substitution over Zeolites Generates Wheland-Type Reaction Intermediates. *Nat. Catal.* **2018**, *1*, 23−31.

(33) Kolboe, S.; Svelle, S.; Arstad, B. Theoretical Study of Ethylbenzenium Ions: The Mechanism for Splitting off Ethene, and the Formation of a π Complex of Ethene and the Benzenium Ion. *J. Phys. Chem. A* **2009**, *113*, 917−923.

(34) Collinge, G.; Yuk, S. F.; Nguyen, M. T.; Lee, M. S.; Glezakou, V. A.; Rousseau, R. Effect of Collective Dynamics and Anharmonicity on Entropy in Heterogenous Catalysis: Building the Case for Advanced Molecular Simulations. *ACS Catal.* **2020**, *10*, 9236−9260.

(35) Liu, P.; Mei, D. Identifying Free Energy Landscapes of Proton-Transfer Processes between Brønsted Acid Sites and Water Clusters Inside the Zeolite Pores. *J. Phys. Chem. C* **2020**, *124*, 22568−22576.

(36) Grifoni, E.; Piccini, G. M.; Lercher, J. A.; Glezakou, V. A.; Rousseau, R.; Parrinello, M. Confinement Effects and Acid Strength in Zeolites. *Nat. Commun.* **2021**, *12*, No. 2630.

(37) DeLuca, M.; Hibbitts, D. Predicting Diffusion Barriers and Diffusivities of C6−C12 Methylbenzenes in MFI Zeolites. *Microporous Mesoporous Mater.* **2022**, *333*, No. 111705.

(38) Cnudde, P.; Demuynck, R.; Vandenbrande, S.; Waroquier, M.; Sastre, G.; Van Speybroeck, V. Light Olefin Diffusion during the MTO Process on H-SAPO-34: A Complex Interplay of Molecular Factors. *J. Am. Chem. Soc.* **2020**, *142*, 6007−6017.

(39) Cnudde, P.; Redekop, E. A.; Dai, W.; Porcaro, N. G.; Waroquier, M.; Bordiga, S.; Hunger, M.; Li, L.; Olsbye, U.; Van Speybroeck, V. Experimental and Theoretical Evidence for the Promotional Effect of Acid Sites on the Diffusion of Alkenes through Small-Pore Zeolites. *Angew. Chem., Int. Ed.* **2021**, *60*, 10016−10022.

(40) Bhan, A.; Joshi, Y. V.; Delgass, W. N.; Thomson, K. T. DFT Investigation of Alkoxide Formation from Olefms in H-ZSM-5. *J. Phys. Chem. B* **2003**, *107*, 10476−10487.

(41) Van Der Mynsbrugge, J.; Hemelsoet, K.; Vandichel, M.; Waroquier, M.; Van Speybroeck, V. Efficient Approach for the Computational Study of Alcohol and Nitrile Adsorption in H-ZSM-5. *J. Phys. Chem. C* **2012**, *116*, 5499−5508.

(42) Moors, S. L. C.; De Wispelaere, K.; Van Der Mynsbrugge, J.; Waroquier, M.; Van Speybroeck, V. Molecular Dynamics Kinetic Study on the Zeolite-Catalyzed Benzene Methylation in ZSM-5. *ACS Catal.* **2013**, *3*, 2556−2567.

(43) Yarulina, I.; De Wispelaere, K.; Bailleul, S.; Goetze, J.; Radersma, M.; Abou-Hamad, E.; Vollmer, I.; Goesten, M.; Mezari, B.; Hensen, E. J. M.; et al. Structure−Performance Descriptors and the Role of Lewis Acidity in the Methanol-to-Propylene Process. *Nat. Chem.* **2018**, *10*, 804−812.

(44) Kresse, G.; Hafner, J. *Ab initio* molecular-dynamics simulation of the liquid-metal−amorphous-semiconductor transition in germanium. *Phys. Rev. B* **1994**, *49*, 14251−14269.

(45) Kresse, G.; Furthmiiller, J. Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set. *Comput. Mater. Sci.* **1996**, *6*, 15−50.

(46) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for *Ab Initio* Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B* **1996**, *54*, 11169.

(47) Blöchl, P. E. Projector Augmented-Wave Method. *Phys. Rev. B* **1994**, *50*, 17953.

(48) Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method. *Phys. Rev. B* **1999**, *59*, 1758.

(49) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.

(50) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, No. 154104.

(51) Heyden, A.; Bell, A. T.; Keil, F. J. Efficient Methods for Finding Transition States in Chemical Reactions: Comparison of Improved Dimer Method and Partitioned Rational Function Optimization Method. *J. Chem. Phys.* **2005**, *123*, No. 224101.

(52) Verstraelen, T.; Van Speybroeck, V.; Waroquier, M. ZEOBUILDER: A GUI Toolkit for the Construction of Complex Molecular Structures on the Nanoscale with Building Blocks. *J. Chem. Inf. Model.* **2008**, *48*, 1530−1541.

(53) Pulay, P. Convergence Acceleration of Iterative Sequences. the Case of Scf Iteration. *Chem. Phys. Lett.* **1980**, *73*, 393−398.

(54) Ghysels, A.; Van Neck, D.; Waroquier, M. Cartesian Formulation of the Mobile Block Hessian Approach to Vibrational Analysis in Partially Optimized Systems. *J. Chem. Phys.* **2007**, *127*, No. 164108.

(55) De Moor, B. A.; Ghysels, A.; Reyniers, M. F.; Van Speybroeck, V.; Waroquier, M.; Marin, G. B. Normal Mode Analysis in Zeolites: Toward an Efficient Calculation of Adsorption Entropies. *J. Chem. Theory Comput.* **2011**, *7*, 1090−1101.

(56) Zhi, Y.; Shi, H.; Mu, L.; Liu, Y.; Mei, D.; Camaioni, D. M.; Lercher, J. A. Dehydration Pathways of 1-Propanol on HZSM-5 in the Presence and Absence of Water. *J. Am. Chem. Soc.* **2015**, *137*, 15781−15794.

(57) Herrmann, S.; Iglesia, E. Elementary Steps in Acetone Condensation Reactions Catalyzed by Aluminosilicates with Diverse Void Structures. *J. Catal.* **2017**, *346*, 134−153.

(58) DeLuca, M.; Kravchenko, P.; Hoffman, A.; Hibbitts, D. Mechanism and Kinetics of Methylating C 6 −C 12 Methylbenzenes with Methanol and DME in H-MFI Zeolites. *ACS Catal.* **2019**, *9*, 6444−6460.

(59) De Wispelaere, K.; Vanduyfhuys, L.; Van Speybroeck, V.Entropy Contributions to Transition State Modeling. In *Modelling and Simulation in the Science of Micro- and Meso-Porous Materials*; Elsevier Inc.: Amsterdam, 2018; pp 189−228.

(60) Ghysels, A.; Verstraelen, T.; Hemelsoet, K.; Waroquier, M.; Van Speybroeck, V. TAMkin: A Versatile Package for Vibrational Analysis and Chemical Kinetics. *J. Chem. Inf. Model.* **2010**, *50*, 1736−1750.

(61) Vandevondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. Quickstep: Fast and Accurate Density Functional Calculations Using a Mixed Gaussian and Plane Waves Approach. *Comput. Phys. Commun.* **2005**, *167*, 103−128.

(62) Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schütt, O.; Schiffmann, F.; et al. CP2K: An Electronic Structure and Molecular Dynamics Software Package -Quickstep: Efficient and Accurate Electronic Structure Calculations. *J. Chem. Phys.* **2020**, *152*, No. 194103.

(63) Lippert, G.; Hutter, J.; Parrinello, M. A Hybrid Gaussian and Plane Wave Density Functional Scheme. *Mol. Phys.* **1997**, *92*, 477−487.

(64) Lippert, G.; Hutter, J.; Parrinello, M. The Gaussian and Augmented-Plane-Wave Density Functional Method for Ab Initio Molecular Dynamics Simulations. *Theor. Chem. Acc.* **1999**, *103*, 124−140.

(65) Yang, K.; Zheng, J.; Zhao, Y.; Truhlar, D. G. Tests of the RPBE, RevPBE, τ -HCTHhyb, Ωb97X-D, and MOHLYP Density Functional Approximations and 29 Others against Representative Databases for Diverse Bond Energies and Barrier Heights in Catalysis. *J. Chem. Phys.* **2010**, *132*, No. 164117.

(66) Goedecker, S.; Teter, M.; Hutter, J. Separable Dual-Space Gaussian Pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703−1710.

(67) Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255−268.

(68) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé-Hoover Chains: The Canonical Ensemble via Continuous Dynamics. *J. Chem. Phys.* **1992**, *97*, 2635−2643.

(69) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101*, 4177−4189.

(70) Torrie, G. M.; Valleau, J. P. Monte Carlo Free Energy Estimates Using Non-Boltzmann Sampling: Application to the Sub-Critical Lennard-Jones Fluid. *Chem. Phys. Lett.* **1974**, *28*, 578−581.

(71) Torrie, G. M.; Valleau, J. P. Monte Carlo Study of a Phase-Separating Liquid Mixture by Umbrella Sampling. *J. Chem. Phys.* **1977**, *66*, 1402−1408.

(72) Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Seok, C.; Dill, K. A. Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations. *J. Chem. Theory Comput.* **2007**, *3*, 26−41.

(73) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129*, No. 124105.

(74) *ThermoLIB | Center for Molecular Modeling*. https://molmod.ugent.be/software/thermolib (accessed Nov 19, 2021).

(75) Cnudde, P.; De Wispelaere, K.; Van Der Mynsbrugge, J.; Waroquier, M.; Van Speybroeck, V. Effect of Temperature and Branching on the Nature and Stability of Alkene Cracking Intermediates in H-ZSM-5. *J. Catal.* **2017**, *345*, 53−69.

(76) Lee, C. K.; Chiang, A. S. T. Adsorption of Aromatic Compounds in Large MFI Zeolite Crystals. *J. Chem. Soc. Faraday Trans.* **1996**, *92*, 3445−3451.

(77) Schumacher, R.; Karge, H. G. Thermodynamics and Kinetics of Adsorption of Selected Monoalkylbenzenes in H-ZSM-5. *J. Phys. Chem. B* **1999**, *103*, 1477−1483.

(78) Piccini, G.; Alessio, M.; Sauer, J.; Zhi, Y.; Liu, Y.; Kolvenbach, R.; Jentys, A.; Lercher, J. A. Accurate Adsorption Thermodynamics of Small Alkanes in Zeolites. Ab Initio Theory and Experiment for H-Chabazite. *J. Phys. Chem. C* **2015**, *119*, 6128−6137.

(79) Amsler, J.; Plessow, P. N.; Studt, F.; Bucko, T. Anharmonic Correction to Adsorption Free Energy from DFT-Based MD Using Thermodynamic Integration. *J. Chem. Theory Comput.* **2021**, *17*, 1155−1169.

(80) Galimberti, D. R.; Sauer, J. Chemically Accurate Vibrational Free Energies of Adsorption from Density Functional Theory Molecular Dynamics: Alkanes in Zeolites. *J. Chem. Theory Comput.* **2021**, *17*, 5849−5862.

(81) Eckstein, S.; Hintermeier, P. H.; Zhao, R.; Baráth, E.; Shi, H.; Liu, Y.; Lercher, J. A. Influence of Hydronium Ions in Zeolites on Sorption. *Angew. Chem., Int. Ed.* **2019**, *58*, 3450−3455.

(82) Vjunov, A.; Wang, M.; Govind, N.; Huthwelker, T.; Shi, H.; Mei, D.; Fulton, J. L.; Lercher, J. A. Tracking the Chemical Transformations at the Brønsted Acid Site upon Water-Induced Deprotonation in a Zeolite Pore. *Chem. Mater.* **2017**, *29*, 9030−9042.

(83) Fu, D.; Erik Maris, J. J.; Stanciakova, K.; Nikolopoulos, N.; van der Heijden, O.; B Mandemaker, L. D.; Siemons, M. E.; Salas Pastene, D.; Kapitein, L. C.; Rabouw, F. T.; et al. Unravelling Channel Structure−Diffusivity Relationships in Zeolite ZSM-5 at the Single-Molecule Level. *Angew. Chem., Int. Ed.* **2022**, *61*, No. e202114388.

(84) Pérez de Alba Ortíz, A.; Tiwari, A.; Puthenkalathil, R. C.; Ensing, B. Advances in Enhanced Sampling along Adaptive Paths of Collective Variables. *J. Chem. Phys.* **2018**, *149*, No. 072320.

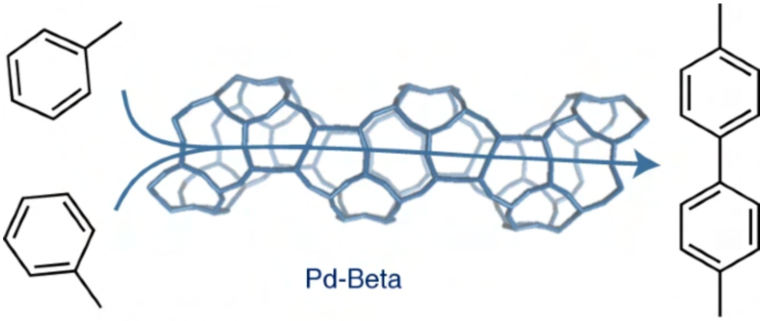
(85) Grifoni, E.; Piccini, G. M.; Parrinello, M. Microscopic Description of Acid−Base Equilibrium. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 4054−4057.

Paper II

Shape-selective C–H activation of aromatics to biarylic compounds using molecular palladium in zeolites

Jannick Vercammen, Massimo Bocus, Samuel E. Neale, Aram Bugaev, Patrick Tomkins, Julianna Hajek, Sam Van Minnebruggen, Alexander Soldatov, Andraž Krajnc, Gregor Mali, Veronique Van Speybroeck* and Dirk E. De Vos*

Nature Catalysis, **2020**, *3*, 1002–1009



- ✓ TON up to 450
- ✓ Selectivity up to 80% for 4,4′-bitolyl
- ✓ Heterogeneous, easily available catalyst

*Corresponding authors

M. Bocus performed the computational investigation of the reaction in collaboration with S. E. Neale and J. Hajek, and contributed to the writing of the manuscript.

nature catalysis

ARTICLES

https://doi.org/10.1038/s41929-020-00533-6

Shape-selective C–H activation of aromatics to biarylic compounds using molecular palladium in zeolites

Jannick Vercammen[1], Massimo Bocus[2], Sam Neale[2], Aram Bugaev[3], Patrick Tomkins[1], Julianna Hajek[2], Sam Van Minnebruggen[1], Alexander Soldatov[3], Andraž Krajnc[4], Gregor Mali[4], Véronique Van Speybroeck[2] ✉ and Dirk E. De Vos[1] ✉

[illegible]

Biarylic bonds are an important structural motif in numerous organic chemicals. Currently, the industrial production of biaryl compounds relies on traditional coupling reactions (Suzuki, Ullmann and so on) which use pre-activated arenes (for example, aryl halides and arylboronic acids)[1]. The high cost of these intermediates and the associated reaction waste generated usually direct the application of these coupling reactions towards making high-end, speciality chemicals. Especially in the polymer industry there has been a growing demand for biarylic monomers because of their superior chemical and physical stability, and their favourable health and safety profiles[2–4].

The direct functionalization of unreactive C–H bonds via transition-metal-mediated C–H activation has emerged as a promising alternative to traditional multi-step approaches[1]. C–H/C–H arene–arene couplings can in principle be used for the direct synthesis of biaryls from simple arenes with, for example, Pd carboxylates as the catalysts[5–10]. If O_2 is used as the oxidant, water is the only by-product[11]. Unfortunately, the multitude of C–H bonds present in organic reactants and the poor differentiation between them often result in poor regioselectivity. Consequently, reactions with simple arenes (for example toluene) lead to useless mixtures of many isomers[12]. Selectivity for the *ortho* position can be achieved if the existing substituent exerts a directing effect, as is typical for ketone groups, carbamates, amides and so on[13]. Complex schemes must be adopted to achieve some *meta* selectivity, for example by installing pendant groups on the ligand of the catalytic complex[14]. Finally, *para* selectivity realized recently *via* pre-functionalization with highly electrophilic radicals[15,16] and by installing extreme steric bulk on the aromatic reactant, for example by using pivaloyl-derivatized reactants[17]. Clearly, these complex reactants are quite far from the simple, monosubstituted aromatics that one would desire to functionalize functionalizing in a regioselective way.

A so-far unexplored option is to implement the molecular mechanisms of transition-metal-catalysed C–H activation in the shape-selective confinement of a porous material like a zeolite. Zeolites are microporous, crystalline aluminosilicates; they are applied on a large scale in industrial reactions because of their excellent activity and stability, along with their low cost. The molecular dimensions and shape of the pores may direct the selectivity to particular products via the imposed steric constraints[18]. This shape selectivity has been extensively studied for acid-catalysed reactions. However, zeolites have so far not been used to create a shape-selective environment around molecular catalysts that activate C–H bonds of arenes to form more complex organics[19].

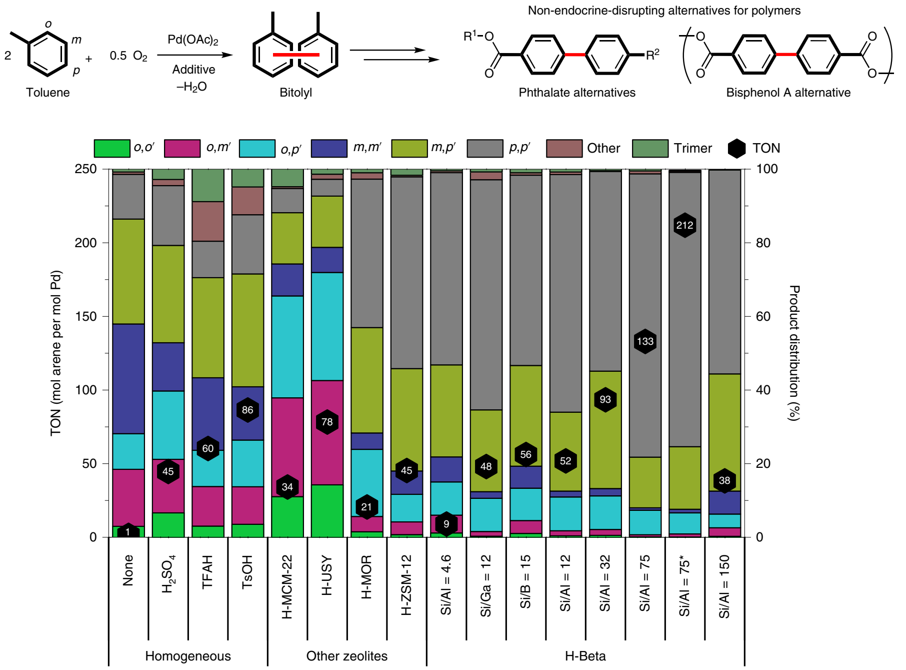
Here we demonstrate that Pd(ɪɪ) catalysts, site-isolated in the framework of zeolite Beta or some other 12-membered-ring (12-MR) zeolites, uniquely allow oxidative coupling of toluene and other monosubstituted aromatics to biaryls, with exceptional selectivity for the *p*,*p*′-dimer.

Results

Pd catalysts at work in zeolite hosts. The most active homogeneous catalysts for the oxidative coupling of arenes generally comprise a palladium salt, a strongly acidic additive, for example, *p*-toluenesulfonic acid (TsOH) or trifluoroacetic acid (TFAH) and acetic acid (AcOH)[5–10]. When applying such homogeneous catalysts to toluene (Fig. 1), an unbiased regioisomer distribution is obtained, with all C–H bonds being activated by Pd to a similar extent (Supplementary Figs. 2 and 3; Supplementary Table 1). Furthermore, side products were identified, resulting from oxygenation, from benzylic coupling and, especially at higher conversions, from consecutive coupling to trimers (4.8 mol% trimer at 8.0% toluene conversion). We first set out to replace the strongly acidic

---

[1]Centre for Membrane Separations, Adsorption, Catalysis, and Spectroscopy for Sustainable Solutions (cMACS), KU Leuven, Leuven, Belgium. [2]Center for Molecular Modeling (CMM), Ghent University, Zwijnaarde, Belgium. [3]The Smart Materials Research Institute, Southern Federal University, Rostov-on-Don, Russia. [4]Department of Inorganic Chemistry and Technology, National Institute of Chemistry, Ljubljana, Slovenia. ✉e-mail: Veronique.vanspeybroeck@ugent.be; dirk.devos@kuleuven.be

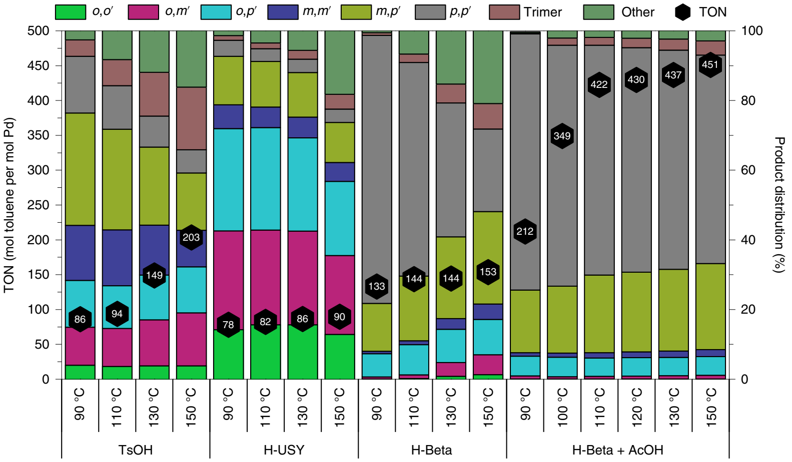


Oxidative coupling of toluene with different additives. Reaction conditions: 15 μmol $Pd(OAc)_2$, 2 ml toluene, 90 °C, 16 bar O_2, 16 h. Homogeneous reactions employed 0.3 mmol of the additive and 1 ml AcOH, while reaction with zeolites used 50 mg at Si/Al < 50 and 100 mg at Si/Al > 50. *AcOH (200 equiv. to Pd) was added.

additives with acid zeolites, initially without acetic acid, to simplify the reaction mixture. At equal loadings of $Pd(OAc)_2$, bitolyl formation with the addition of acid zeolites not only surpassed that with any other solid acid tested, but occasionally even the homogeneous catalyst tested under the same conditions (Fig. 1). Large-pore zeolites (that is with 12-MRs) were required to offer sufficient space for toluene and other substituted arenes to react; medium-pore zeolites such as H-ZSM-5 were only effective for the oxidative coupling of benzene. To our delight, zeolites not only with three-dimensional pores (H-Beta) but also with one-dimensional pores (H-MOR, H-ZSM-12 and so on), which are more susceptible to diffusion restrictions, could be used to boost the catalytic performance of Pd in this reaction.

The most remarkable feature of this disclosure is the regioselectivity in the biaryl product if zeolites are applied. Out of the six possible bitolyl regiomers, zeolites containing only 12-MR channels and no large cavities (H-MOR, H-ZSM-12, H-Beta) were highly selective for *para*-substituted products. In H-Beta, the combined selectivity for all products with at least one *para* substituent reaches 97% in H-Beta and with a selectivity of 77% for *p*,*p*′-bitolyl, compared to only 59 and 16% respectively for a homogeneous catalyst with TsOH. Zeolites with large cages (H-USY, H-MCM-22) showed a high combined selectivity for products with at least one *ortho* substituent (72% for H-USY compared to 28% for TsOH). Furthermore, side reactions, such as coupling at the benzylic position or oxygenation reactions, were suppressed in 12-MR zeolites like H-Beta. On MCM-22, known for its surface activity[20], some trimer formation was observed, but such consecutive coupling at higher conversions was inhibited in the microporous environment of the other tested 12-MR zeolites.

Towards high activity and selectivity. In view of the intriguing shape selectivity for a double C–H activation at the *para* position combined with the exceptional activity of Pd-loaded H-Beta, we decided to investigate this system further (Supplementary Figs. 4–11). Addition of small amounts of acetic acid (1–500 equiv. to Pd) further boosted the performance to a turnover number (TON) of 212 or a yield of 16.8% while maintaining an identical product distribution (Supplementary Fig. 12). High zeolite Si/Al ratios were required for obtaining higher TON values (Fig. 1). This is likely unrelated to acid strength since similar results were obtained with Ga- and B-substituted Beta. Rather, the beneficial effect of the high Si/Al ratio seems related to the more hydrophobic properties of such a zeolite (Fig. 1). A decline in regioselectivity during the reaction was observed at a Si/Al ratio of 32 (Supplementary Fig. 6) while the regioselectivity remained constant and even improved initially at the higher Si/Al ratio of 75 (Fig. 4b). This was attributed to a poisoning effect of the formed water[21], which was subsequently confirmed by spiking reactions with 100 μl of water (370 equiv. to Pd): here the TON was decreased drastically from 93 to 2 for a Si/Al of 32, but the TON decreased only slightly, from 133 to 96, for the more hydrophobic zeolite Beta with a Si/Al of 75. The excellent catalytic stability of H-Beta is further highlighted by a doubling of the TON from 212 to 422 (bitolyl yield 34%) upon increasing the reaction temperature from 90 to 110 °C (Fig. 2). High TONs were obtained at O_2 pressures of 16 bar or more; however, the addition



Variation of the reaction temperature in the oxidative coupling of toluene. Reaction conditions: 15 µmol $Pd(OAc)_2$, 2 ml toluene, 16 bar O_2, 16 h. Following amounts were employed: TsOH (0.3 mmol) and 1 ml AcOH, H-USY (Si/Al = 40, 50 mg), H-Beta (Si/Al = 75, 100 mg), H-Beta + AcOH (200 equiv. AcOH to Pd).

of $Fe(OTf)_3$ also allows the zeolite Beta system to reach significant TONs under only 1 bar of O_2 (see Supplementary Table 2).

Nature of the active site. The catalytic species Pd-Beta, formed upon loading the H-Beta zeolite with $Pd(OAc)_2$, was characterized in detail (Fig. 3). X-ray absorption near edge structure (XANES) spectra are consistent with Pd(II) being surrounded by four neighbouring oxygen atoms in a square planar coordination; Pd prevails in the +2 state, while the contribution of Pd(0) nanoparticles is only minor (Supplementary Figs. 14 and 15, Supplementary Table 3). When the loading is performed in toluene, the spectral changes suggest that one of these oxygen atoms may be replaced by a nearby toluene molecule (Supplementary Fig. 16). Extended X-ray absorption fine structure (EXAFS) spectra were successfully modelled with mononuclear Pd(II) with four oxygen neighbours; although there could be a small contribution from Pd(0) nanoparticles, the presence of di- or trinuclear Pd, as in the precursor $Pd(OAc)_2$, can be excluded. Again, upon loading in toluene, the best fit is obtained with three oxygen atoms from the zeolite and an acetate, together with a nearby toluene (Fig. 3a; Supplementary Fig. 17; Supplementary Tables 4–6). Solid-state ^{13}C magic-angle spinning NMR of the acid zeolite Beta loaded with ^{13}C-enriched $Pd(OAc)_2$ proved that one acetate out of the two reacts to form acetic acid, moving freely in the pores, while the second acetate remains immobilized at a Pd centre (Supplementary Fig. 21). The mutual proximity of acetate ligands was probed using 2D ^{13}C–^{13}C spin-diffusion experiments and 1D ^{13}C double-quantum-filtered experiments (Fig. 3; Supplementary Figs. 22 and 23), which proved that the individual acetates are spatially isolated from other acetate ligands. This combined information points to a molecular Pd(II) species that receives one charge compensation from the framework and one from a pendant acetate ligand. However, it does not rule out the possibility that a fraction of Pd is compensated by two anionic zeolite charges.

Such catalytic species can be formed in zeolite Beta, either in situ in toluene or prior to the reaction (Supplementary Fig. 13). The zeolite amount was varied at a fixed amount of Pd; the reaction rate increased linearly until an Al/Pd ratio of 1 was reached. This supports the single charge compensation of each Pd centre by the zeolite (Fig. 4a). The crucial presence of the acetate ligand was shown by the absence of any reaction if a Pd(II)-zeolite was prepared via classical ion-exchange with $Pd(NH_3)_4^{2+}$ and subsequent calcination (Supplementary Fig. 24), where the catalytic activity could only be restored by the addition of acetic acid. Palladium is strongly bound to the framework, as was shown by a successful filtration test which demonstrated a fully heterogeneous Pd-zeolite catalyst (Fig. 4b; Supplementary Figs. 25 and 26; Supplementary Table 7).

Producing useful biarylic monomers. The methodology could easily be expanded to the oxidative coupling of other simple arenes of interest in polymer chemistry. Oxidative coupling of anisole could lead to an alternative synthetic strategy for 4,4′-dihydroxybiphenyl, which is not only a starting material for heat-resistant materials but also a replacement for the disputed bisphenol A-derived polycarbonates[22,23]. A high activity (TON of 197) and a selectivity of 77% for *p,p*′-bianisyl were obtained in the oxidative homocoupling of anisole with H-Beta (Fig. 5a). While the homogeneous reaction with TsOH formed up to 56% of bisanisyl F (bis(anisyl)methane), which was due to the consecutive hydrolysis of anisole, oxidation of the formed methanol to formaldehyde and condensation reaction with anisole, only 4% of bisanisyl F was formed with H-Beta. Lastly, *p,p*′-bianisyl selectively precipitated from the reaction mixture upon cooling, not only simplifying the work-up but also yielding high-purity crystals (Supplementary Figs. 30 and 31). The regioselective coupling also applies to disubstituted arenes like *o*-xylene, from which the desired 3,3′,4,4′-tetramethylbiphenyl was obtained in 90% overall selectivity (Fig. 5b). This tetramethyl-substituted compound is a direct precursor of 3,3′,4,4′-biphenyldicarboxylic acid, from which high value polyimide can be produced.

Density functional theory calculations. To support the experimental observations and to obtain in-depth mechanistic information, a full catalytic cycle was proposed and calculated by means of a combination of first-principles static and molecular dynamics-based dispersion-corrected density functional theory (DFT) calculations. The first-principles molecular dynamics (MD) calculations were

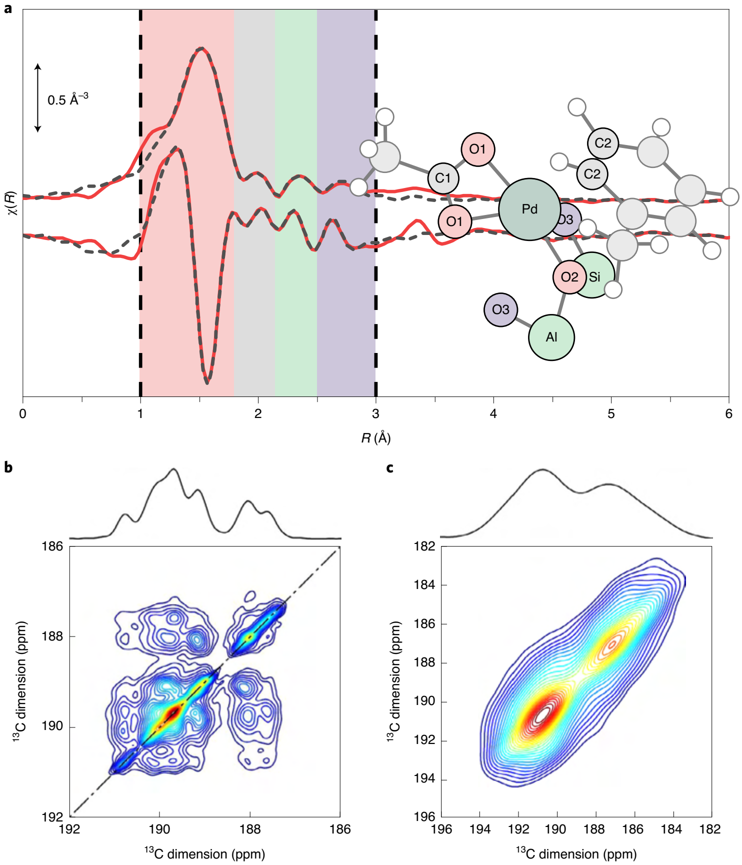
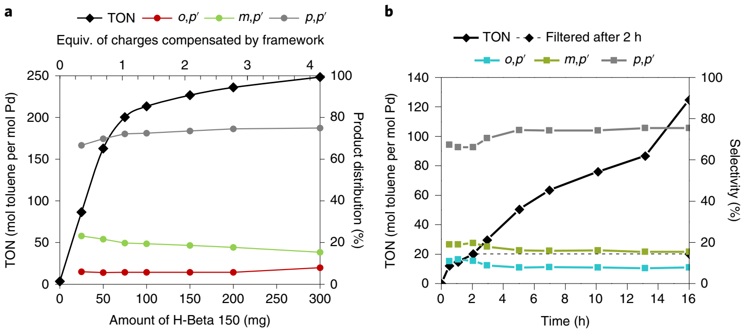


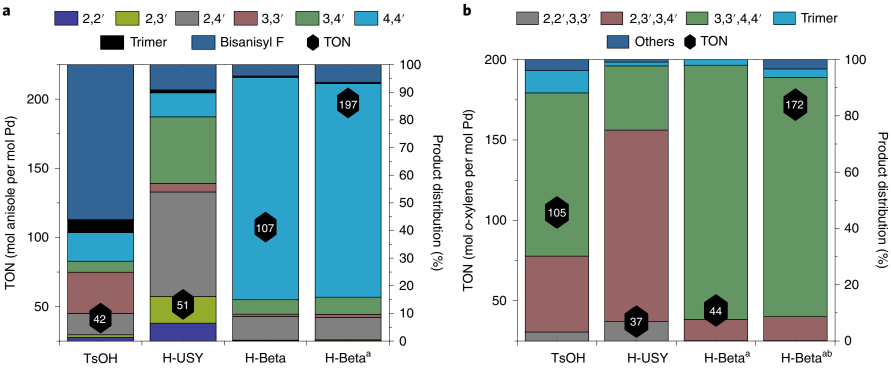
Fig. 3 | Spectroscopic characterization of Pd-loaded zeolite Beta. a, Experimental (red) and fitted (dashed grey) EXAFS spectra for Pd-Beta, pre-loaded in toluene. For the fit, the theoretically proposed intermediate **V** was used (Fig. 6b). The different background colours highlight the region with the major contribution from the atoms of the corresponding colour. b,c, 2D ^{13}C–^{13}C spin-diffusion magic-angle spinning NMR spectra of the ^{13}C-enriched $Pd(OAc)_2$ (b) and Pd-Beta (c). The absence of cross-peaks in the latter spectrum proves that the acetate ligands are spatially isolated.

employed to explore the flexible behaviour of the active catalytic site when docked to the zeolite in various stages of catalysis at realistic temperatures and toluene loadings, whereas the static DFT calculations afforded adsorption free energies and allowed for the construction of regiodivergent reaction profiles to explore the origin of shape selectivity in the reaction. The proposed catalytic cycle for the homocoupling reaction of toluene in H-Beta is reported in Fig. 6a, whereas the associated free energy profile is shown in Fig. 6b.

The catalytic reaction begins with the adsorption of $Pd(OAc)_2$ into the zeolite framework, where, relative to the empty H-Beta and the isolated $Pd(OAc)_2$, the adsorption free energy is favourable and amounts to −75 kJ mol^{-1} (species **II**) at 363 K (Fig. 6b). Subsequently, the actual docking can take place as described experimentally to release AcOH and afford Pd(OAc) species **III** (−80 kJ mol^{-1}). This is mildly exergonic with respect to the preceding intermediate **II**, with a decrease in free energy of 5 kJ mol^{-1}. This proposed mechanism of docking to form **III** is fully supported by ^{13}C NMR measurements (vide supra, Fig. 3). The docked Pd species **III** then acts as the initial species in the catalytic cycle and, based on our calculations, we propose that the reaction proceeds through four elementary reaction steps (Fig. 6). After the first toluene molecule has been adsorbed on the zeolite with a favourable free energy gain of 37 kJ mol^{-1}, C–H activation takes place through a concerted metallation–deprotonation mechanism (CMD, Step 1 in Fig. 6), a pathway that has been well-studied in the literature[24]. First, the toluene must enter into the coordination sphere of Pd, substituting one of the zeolite oxygens to afford the η^2-(tolyl)Pd species **V** (−122 kJ mol^{-1}). The presence of **V** as a key reaction intermediate is in good agreement with the X-ray absorption spectroscopy (XAS) measurements of the catalyst pre-loaded with toluene (Fig. 3). Then, CMD can take place, where the calculated activation energies indicate that the reaction involving the *para* carbon is the most accessible, since the barrier amounts to 91 kJ mol^{-1} versus 101 kJ mol^{-1} and 103 kJ mol^{-1} for activation at the *meta* and *ortho* positions respectively (Supplementary Fig. 40). This suggests that the large overall 97% *para* selectivity observed experimentally is mainly governed by the kinetics of this initial



▯▯▯▯ ▯**Variation of the amount of H-Beta, kinetics and filtration test.** ▯, Variation of the amount of H-Beta (Si/Al = 75) in the oxidative coupling of toluene (2 ml) with 3 mmol AcOH added. Reaction conditions: 15 μmol $Pd(OAc)_2$, 90 °C, 16 bar O_2, 16 h. ▯, Time profile and filtration test for the oxidative coupling of toluene with H-Beta (100 mg, Si/Al = 75) without additional AcOH. After 2 h of reaction, the solids were removed by filtration and the filtrate was reacted further with fresh H-Beta (Si/Al = 75, 100 mg) for 14 h (dashed line; see Supplementary Figs. 26 and 27 for detailed description). Reaction conditions: 15 μmol $Pd(OAc)_2$, 2 ml toluene, 90 °C, 16 bar O_2.



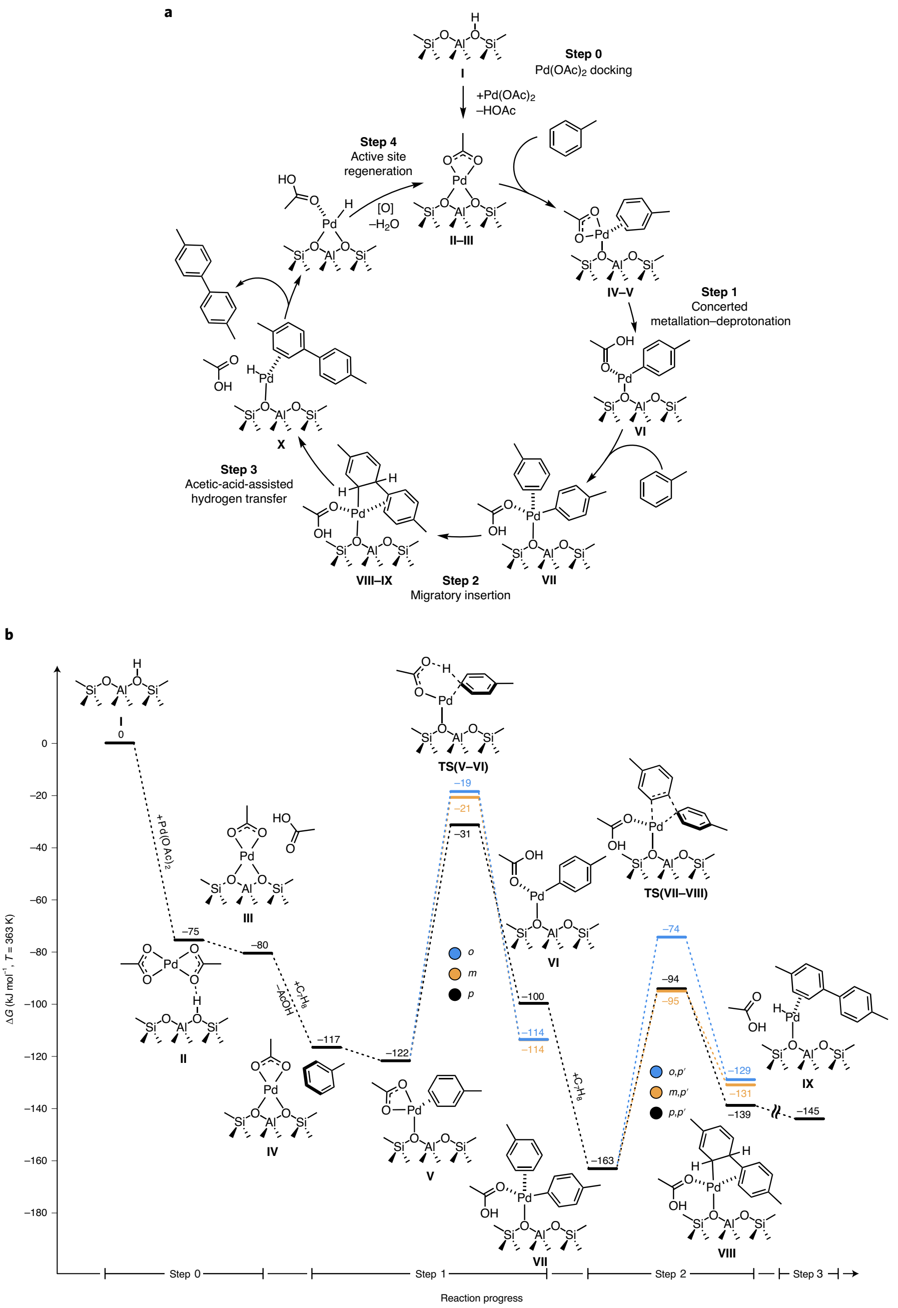
▯▯▯▯ ▯**Substrate scope.** ▯,▯, Oxidative coupling of anisole (▯) and *o*-xylene (▯). Reaction conditions: 15 μmol $Pd(OAc)_2$, 2 ml arene, 90 °C, 16 bar O_2, 16 h. Following amounts were employed: TsOH (0.3 mmol) and 1 ml AcOH, H-USY (Si/Al = 40, 50 mg), H-Beta (Si/Al = 75, 100 mg). [a]AcOH added, 3 mmol. [b]Reaction temperature, 110 °C.

CMD step. It might be noticed that the *para* product has a higher energy than the others; this is caused by the methyl group directly pointing toward the zeolite walls and creating some short contacts (Supplementary Fig. 41). Further support for *para* activation was gained through an MD simulation of the reactant state, in which a predominant interaction between Pd and the toluene *para* carbon was observed (Supplementary Figs. 42–44).

After the first CMD step, we propose further C–C coupling to occur through a migratory insertion reaction (MI, Step 2 in Fig. 6, and Supplementary Figs. 45–49)[25,26]. The calculated reaction barriers for this process are relatively low (68 kJ mol^{-1} for *m,p*′-selectivity, 69 kJ mol^{-1} for *p,p*′-selectivity and 89 kJ mol^{-1} for *o,p*′-selectivity). To test the validity of the proposed MI mechanism further, we computed an alternative second CMD mechanism. The resulting barriers for such processes were much higher (with $\Delta G^{\ddagger} = 187$ kJ mol^{-1} for *p,p*′-selectivity, Supplementary Fig. 51 and Supplementary Tables 14 and 15), thus giving further support to the MI mechanism.

Moreover, experimental kinetic studies revealed a first-order rate dependence in Pd (Supplementary Fig. 36), thus further supporting the proposed MI mechanism as the favoured means of C–C coupling over a bimetallic transmetallation process. While the MI barrier difference between *m,p*′ and *p,p*′ is small (with $\Delta\Delta G^{\ddagger} = 1$ kJ mol^{-1} for the regiodivergent analogues of **TS(VII–VIII)**), the *o,p*′ activation energy is much higher, which arises from the increased steric clashing between the methyl group and the zeolite pore (Supplementary Fig. 49). The calculations thus suggest that the predominant overall *para* selectivity is primarily dictated by the first CMD step, while more subtle effects are at play in the rest of the catalytic cycle to afford the overall 77% yield for *p,p*′-bitolyl (Supplementary Fig. 47).

Subsequently, from **VIII**, the transfer of a hydrogen atom from the C–C coupled bitolyl moiety to Pd needs to take place. However, the lack of an available base does not allow for a direct reductive deprotonation, while the *anti*-configuration of Pd and the hydrogen prevent a β-hydride elimination[26,27]. We thus propose that in Step 3 a hydrogen-transfer process from the resulting C–C coupled cyclohexadienyl adduct **VIII** to the Pd centre can take place, facilitated by AcOH itself. However, accurate characterization of the corresponding transition states proved challenging, and so explicit



Proposed catalytic cycle for the oxidative coupling of toluene using H-Beta. , Reaction scheme. , Free energy profile of the proposed mechanism for the oxidative coupling of toluene. The profile (kJ mol⁻¹) was computed at the B3LYP-D3//PBE-D3 level of theory at 363 K. The schematic representations of the intermediates and transition states are related to the mechanism leading to the *p,p′*-bitolyl product. Enthalpic and entropic contributions to the free energy are reported in Supplementary Tables 10–15.

transition-state energies are not reported in Fig. 6b (Supplementary Fig. 50). The eventually obtained Pd hydride **IX** (−145 kJ mol^{-1}) was found to be a stable energy minimum in the free energy reaction cycle. Finally, from **IX**, active site regeneration (Step 4) is proposed to take place via O_2 insertion. Characterization of this process was carried out using a 5T cluster model to efficiently compute the spin-forbidden nature of the mechanism, as previously identified by Stahl and collaborators[28,29]. These calculations revealed that hydrogen atom abstraction (HAA) is favoured to form the resulting Pd–OOH species and ultimately release H_2O_2 (Supplementary Figs. 52 and 53)

Conclusions

We report the development of a strategy for controlling the regioselectivity in the C–H functionalization of aromatics without a functional handle. The elusive *para* selectivity is obtained through a non-traditional shape selectivity in which exchanged cationic palladium in zeolites is utilized, rather than the acidic sites of zeolites. The technology grants access to a novel range of monomers for advanced polymers; the methodology also provides a paradigm for introducing shape selectivity in the transition-metal-catalysed C–H activation of arenes.

Methods

Reaction procedure. Generally, toluene (2.00 ml, 18.88 mmol, Acros), acetic acid (172 µl, 0.3 mmol, Acros) and *n*-decane (50.0 µl, internal standard) were added to $Pd(OAc)_2$ (3.37 mg, 0.015 mmol, Acros) and H-Beta (100.0 mg, Si/Al = 75, Süd-Chemie) in a 3 ml glass liner. Homogeneous reactions were performed with 0.3 mmol of the additive and 1 ml of acetic acid, while reactions with zeolites employed 50 mg of the solid without acetic acid. Zeolites with a Si/Al ratio above 50 employed 100 mg of the solid. The liner was sealed in a homemade, stainless steel autoclave (Supplementary Fig. 1), and purged five times with pure O_2 and subsequently pressurized with pure O_2 to 16 bar. The reactor was placed in a pre-heated four-well aluminium block. The internal temperature dependence was calibrated beforehand, and the set temperature of the heating block was adjusted to obtain 90 °C internally.

After 16 h, the reactor was cooled and the zeolite was separated from the reaction mixture via centrifugation. The reaction mixture was analysed by GC-FID. When anisole was used as the aromatic reagent, the 4,4′-bianisyl product precipitated from the reaction medium upon cooling. Therefore, this mixture was extracted with $CHCl_3$ (20 ml) before GC-FID analysis to fully solubilize all products. In accordance with literature reports, the TON is defined as TON = 2 × mol biaryl / mol Pd.

Each reactor had an intentional weak spot at the capillaries on top of the reactor which functioned as a burst disc. Note: a thorough safety assessment must be made before applying such hazardous conditions and these reactions should only be performed using proper equipment that ensures safe handling at all times.

Reaction analysis. Reaction mixtures were analysed quantitatively using a Shimadzu GC-2014 equipped with a CP-SIL 5 CB column (Agilent, 100% PDMS, 60 m, 0.25 µm film thickness, 0.32 mm i.d.). Samples of 1 µl were injected automatically using an AOC-20s autosampler and AOC-20i auto-injector aided by the GCsolution software bundle (version 2.44.00). Products were identified using an Agilent 6890 gas chromatograph equipped with an HP-1 MS column and coupled to a 5973 MSD mass spectrometer or by comparison with commercially obtained or synthesized reference samples. ^{1}H NMR spectra of liquid samples were recorded using a Bruker Avance III HD 400 console at 400 MHz, equipped with a 5 mm PABBO BB/^{19}F–^{1}H/D probe and the data were analysed using the MestReNova 12.0.2 software package.

Pre-loading. Zeolites can be pre-loaded with $Pd(OAc)_2$ prior to the reaction by contacting the solid with an appropriate amount of $Pd(OAc)_2$ in a suitable solvent (for example $CHCl_3$, toluene). After stirring for 24 h, the zeolite was separated by centrifugation, and washed twice with the loading solvent (10 ml per g zeolite). The solid was then dried in vacuo (30 °C, 30 mbar, overnight) and used as such.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The findings of this study are available in the main text or the supplementary materials. Atomic coordinates of optimized computational models and initial and final configurations of molecular dynamics trajectories are supplied in a Supplementary Data file. All data are available from the authors upon reasonable request.

Received: 10 January 2020; Accepted: 30 September 2020;

References

1. Hassan, J., Sévignon, M., Gozzi, C., Schulz, E. & Lemaire, M. Aryl–aryl bond formation one century after the discovery of the Ullmann reaction. *Chem. Rev.* **102**, 1359–1469 (2002).
2. Mondschein, R. J. et al. Synthesis and characterization of amorphous bibenzoate (co)polyesters: permeability and rheological performance. *Macromolecules* **50**, 7603–7610 (2017).
3. De Smit, E. et al., Hydroalkylation catalyst and process for use thereof. US patent 2014/0378697 (2014).
4. Dakka, J. M. et al. Biphenyl esters, their production and their use in the manufacture of plasticizers. US patent 9,556,103 (2017).
5. Chen, X., Engle, K. M., Wang, D. -H. & Yu, J. -Q. Palladium(II)-catalyzed C–H activation/C–C cross-coupling reactions: versatility and practicality. *Angew. Chem. Int. Ed.* **48**, 5094–5115 (2009).
6. Xu, B.-Q., Sood, D., Iretskii, A. V. & White, M. G. Direct synthesis of dimethylbiphenyls by toluene coupling in the presence of palladium triflate and triflic acid. *J. Catal.* **187**, 358–366 (1999).
7. Izawa, Y. & Stahl, S. S. Aerobic oxidative coupling of *o*-xylene: discovery of 2-fluoropyridine as a ligand to support selective Pd-catalyzed C–H functionalization. *Adv. Synth. Catal.* **352**, 3223–3229 (2010).
8. Wang, D., Izawa, Y. & Stahl, S. S. Pd-catalyzed aerobic oxidative coupling of arenes: evidence for transmetalation between two Pd(II)–aryl intermediates. *J. Am. Chem. Soc.* **136**, 9914–9917 (2014).
9. Wang, D. & Stahl, S. S. Pd-catalyzed aerobic oxidative biaryl coupling: non-redox cocatalysis by $Cu(OTf)_2$ and discovery of $Fe(OTf)_3$ as a highly effective cocatalyst. *J. Am. Chem. Soc.* **139**, 5704–5707 (2017).
10. Álvarez-Casao, Y. et al. Palladium-catalyzed cross-dehydrogenative coupling of *o*-xylene: evidence of a new rate-limiting step in the search for industrially relevant conditions. *ChemCatChem* **10**, 2620–2626 (2018).
11. Yang, Y., Lan, J. & You, J. Oxidative C–H/C–H coupling reactions between two (hetero)arenes. *Chem. Rev.* **117**, 8787–8863 (2017).
12. Kuhl, N., Hopkinson, M. N., Wencel-Delord, J. & Glorius, F. Beyond directing groups: transition-metal-catalyzed C–H activation of simple arenes. *Angew. Chem. Int. Ed.* **51**, 10236–10254 (2012).
13. Sambiagio, C. et al. A comprehensive overview of directing groups applied in metal-catalysed C–H functionalisation chemistry. *Chem. Soc. Rev.* **47**, 6603–6743 (2018).
14. Leow, D., Li, G., Mei, T.-S. & Yu, J.-Q. Activation of remote *meta*-C–H bonds assisted by an end-on template. *Nature* **486**, 518–522 (2012).
15. Boursalian, G. B., Ham, W. S., Mazzotti, A. R. & Ritter, T. Charge-transfer-directed radical substitution enables *para*-selective C–H functionalization. *Nat. Chem.* **8**, 810–815 (2016).
16. Berger, F. et al. Site-selective and versatile aromatic C–H functionalization by thianthrenation. *Nature* **567**, 223–228 (2019).
17. Dey, A., Maity, S. & Maiti, D. Reaching the south: metal-catalyzed transformation of the aromatic *para*-position. *Chem. Commun.* **52**, 12398–12414 (2016).
18. Van Speybroeck, V. et al. Advances in theory and their application within the field of zeolite chemistry. *Chem. Soc. Rev.* **44**, 7044–7111 (2015).
19. Kosinov, N., Liu, C., Hensen, E. J. M. & Pidko, E. A. Engineering of transition metal catalysts confined in zeolites. *Chem. Mater.* **30**, 3177–3198 (2018).
20. Lawton, S. L., Leonowicz, M. E., Partridge, R. D., Chu, P. & Rubin, M. K. Twelve-ring pockets on the external surface of MCM-22 crystals. *Micropor. Mesopor. Mat.* **23**, 109–117 (1998).
21. Van Velthoven, N. et al. Single-site metal–organic framework catalysts for the oxidative coupling of arenes via C–H/C–H activation. *Chem. Sci.* **10**, 3616–3622 (2019).
22. Kaplan, G. Preparation of biphenols by oxidative coupling of alkylphenols using a recyclable copper catalyst. US patent 2003/0050515 (2003)
23. Eckardt, M., Greb, A. & Simat, T. J. Polyphenylsulfone (PPSU) for baby bottles: a comprehensive assessment on polymer-related non-intentionally added substances (NIAS). *Food Addit. Contam. A* **35**, 1421–1437 (2018).
24. Davies, D. L., Macgregor, S. A. & McMullin, C. L. Computational studies of carboxylate-assisted C–H activation and functionalization at group 8–10 transition metal centers. *Chem. Rev.* **117**, 8649–8709 (2017).
25. Chen, B., Hou, X., Li, Y. & Wu, Y. Mechanistic understanding of the unexpected meta selectivity in copper-catalyzed anilide C–H bond arylation. *J. Am. Chem. Soc.* **133**, 7668–7671 (2011).
26. Deng, C., Zhang, J. & Lin, Z. Theoretical studies on Pd(II)-catalyzed *meta*-selective C–H bond arylation of arenes. *ACS Catal.* **8**, 2498–2507 (2018).
27. Lane, B. S., Brown, M. A. & Sames, D. Direct palladium-catalyzed C-2 and C-3 arylation of indoles: a mechanistic rationale for regioselectivity. *J. Am. Chem. Soc.* **127**, 8050–8057 (2005).

28. Popp, B. V. & Stahl, S. S. Insertion of molecular oxygen into a palladium–hydride bond: computational evidence for two nearly isoenergetic pathways. *J. Am. Chem. Soc.* **129**, 4410–4422 (2007).
29. Konnick, M. & Stahl, S. S. Reaction of molecular oxygen with a Pd^{II}-hydride to produce a Pd^{II}-hydroperoxide: experimental evidence for an HX-reductive-elimination pathway. *J. Am. Chem. Soc.* **130**, 5753–5762 (2008).

Acknowledgements

We thank N. Van Velthoven for discussion. The XAS experiments were performed on beamline BM26A at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. This work was funded by grants from FWO (1S17620N for J.V.; project G0D0518N, G0F2320N, G078118N; EoS BioFACT), the Flemish government (CASAS Methusalem programme for D.D.V.). V.V.S., J.H. and M.B. acknowledge the Research Board of Ghent University (BOF) and funding from the European Union's Horizon 2020 research and innovation programme (consolidator ERC grant agreement No. 647755 – DYNPOR (2015-2020)). The computational resources and services used were provided by Ghent University (Stevin Supercomputer Infrastructure) and the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO). A.S. and A.B. acknowledge the funding from Russian Science Foundation (joint RSF-FWO grant No. 20-43-01015). A.K. and G.M. acknowledge the financial support from the Slovenian Research Agency (research core funding No. P1-0021 and project No. N1-0079).

Author contributions

Under the supervision of D.D.V., J.V. was responsible for the conception, design and interpretation of the experiments. S.V.M. performed additional experiments. Under the supervision of V.V.S., J.H., S.N. and M.B. performed the DFT calculations. A.B. and A.S. conceived and performed the XAS experiments. A.K. and G.M. conceived and performed the NMR experiments. All authors discussed the results and commented on the manuscript.

Competing interests

J.V., P.T. and D.D.V. filed a patent application GB1804905.6 prior to an international patent application PCT/EP2019/057746.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41929-020-00533-6.

Correspondence and requests for materials should be addressed to V.V. or D.D.V.

Reprints and permissions information is available at www.nature.com/reprints.

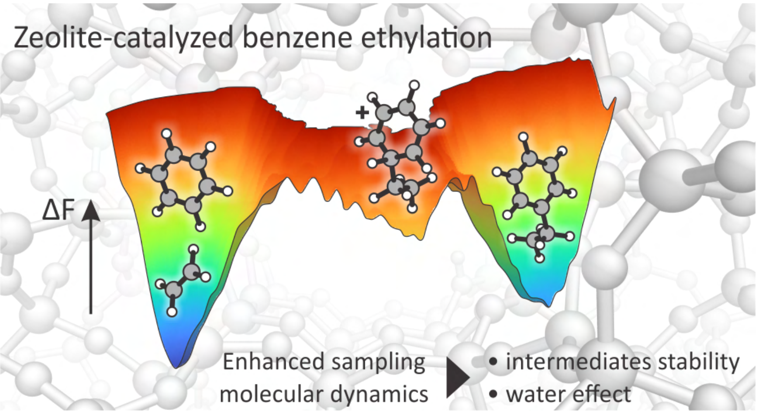
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Paper III

Mechanistic characterization of zeolite-catalyzed aromatic electrophilic substitution at realistic operating conditions

Massimo Bocus, Louis Vanduyfhuys, Frank De Proft, Bert M. Weckhuysen and Veronique Van Speybroeck*

JACS Au, **2022**, *2*, 502–514



*Corresponding author
M. Bocus performed the simulations and prepared the initial draft of the manuscript.

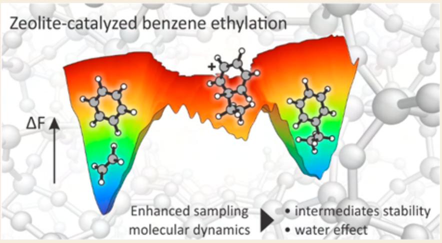
JACS Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/jacsau
Article

Mechanistic Characterization of Zeolite-Catalyzed Aromatic Electrophilic Substitution at Realistic Operating Conditions

Massimo Bocus, Louis Vanduyfhuys, Frank De Proft, Bert M. Weckhuysen, and Veronique Van Speybroeck*

Cite This: *JACS Au* 2022, 2, 502−514



ABSTRACT: Zeolite-catalyzed benzene ethylation is an important industrial reaction, as it is the first step in the production of styrene for polymer manufacturing. Furthermore, it is a prototypical example of aromatic electrophilic substitution, a key reaction in the synthesis of many bulk and fine chemicals. Despite extensive research, the reaction mechanism and the nature of elusive intermediates at realistic operating conditions is not properly understood. More in detail, the existence of the elusive arenium ion (better known as Wheland complex) formed upon electrophilic attack on the aromatic ring is still a matter of debate. Temperature effects and the presence of protic guest molecules such as water are expected to impact the reaction mechanism and lifetime of the reaction intermediates. Herein, we used enhanced sampling ab initio molecular dynamics simulations to investigate the complete mechanism of benzene ethylation with ethene and ethanol in the H-ZSM-5 zeolite. We show that both the stepwise and concerted mechanisms are active at reaction conditions and that the Wheland intermediate spontaneously appears as a shallow minimum in the free energy surface after the electrophilic attack on the benzene ring. Addition of water enhances the protonation kinetics by about 1 order of magnitude at coverages of one water molecule per Brønsted acidic site. In the fully solvated regime, an overstabilization of the BAS as hydronium ion occurs and the rate enhancement disappears. The obtained results give critical atomistic insights in the role of water to selectively tune the kinetics of protonation reactions in zeolites.

KEYWORDS: *zeolite, DFT, enhanced sampling, mechanism, water, molecular dynamics, benzene ethylation*

1. INTRODUCTION

Acid-catalyzed (de)alkylation reactions of aromatic substrates constitute a key class of organic chemistry processes. Presumably, one of the most remarkable examples is represented by the zeolite-catalyzed alkylation of benzene with ethene to produce ethylbenzene,[1,2] which is the first step in the synthesis of styrene monomers for polystyrene production.[3] Quite some interest was initially also directed to the use of ethanol as an alkylating agent; however, the low prices of fossil resources in the previous century made this alternative poorly appealing.[4] Recently, on the other hand, the relevance of the use of ethanol is growing together with the search for environmentally friendly processes, as ethanol may be obtained from renewable resources.[5] Moreover, zeolites have also been shown to effectively catalyze the dealkylation of the alkylphenolic monomers derived from lignin valorization processes to produce simpler and more useful molecules such as phenol and olefins.[6,7] Therefore, zeolite-catalyzed (de)alkylation reactions of aromatic substrates are regaining a great deal of attention as an effective tool in the conversion and valorization of biomasses to commodity chemicals.

Mechanistically, the prototypical alkylation of benzene with ethene or ethanol in the H-ZSM-5 zeolite has represented the main case study in the computational investigation of zeolite-catalyzed alkylation reactions.[8−15] There is consensus that two main reaction pathways are possible: (i) a stepwise mechanism (TS0, TS1, and TS3 in Figure 1) and (ii) a concerted mechanism (TS2 and TS3 in Figure 1). In the former, ethene or ethanol is first chemisorbed on the zeolite walls as a surface ethoxide species (SES, Figure 1b). The SES then acts as an electrophile in a typical electrophilic aromatic substitution reaction (S_EAr), which generates a Wheland complex (also known as σ complex, Figure 1c) as the intermediate, i.e., a protonated arenium ion. In the concerted mechanism, ethene or ethanol is directly activated by the zeolite Brønsted acid site (BAS) and undergoes the S_EAr reaction to attack the aromatic

Received: December 3, 2021
Published: February 3, 2022

https://doi.org/10.1021/jacsau.1c00544

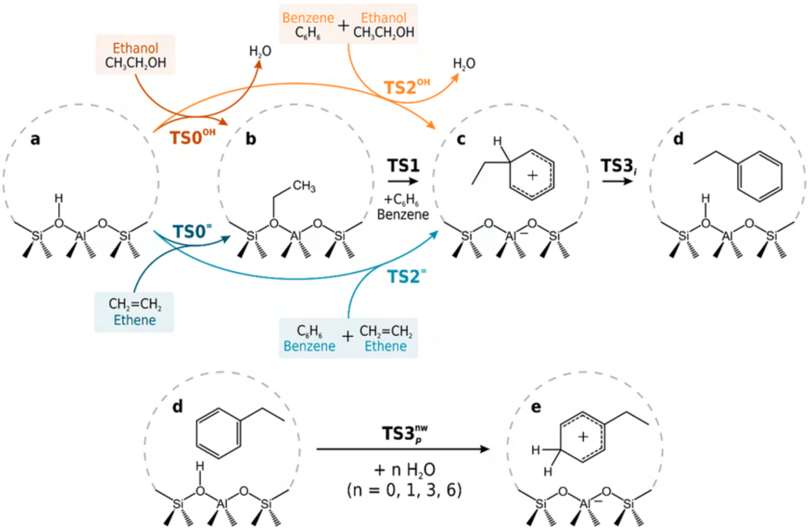


Figure 1. Schematic depiction of all reactions investigated in this work. (Top) Possible mechanistic pathways for the ethylation of benzene with ethanol (orange arrows) and ethene (blue arrows) are shown. From left to right, the pristine zeolite BAS (a) can directly react with ethene or ethanol (TS0) to give a chemisorbed surface ethoxide species (b) which, in its turn, can attack benzene (TS1) to possibly form the *ipso*-protonated Wheland complex (c). Alternatively, benzene and ethene or ethanol can directly react to again form the *ipso*-protonated Wheland complex (TS2). The latter can then deprotonate ($TS3_i$), leading to the final ethylbenzene product (d). Ethylbenzene can also be further reprotonated by the framework on the *para* carbon, forming a second Wheland complex (e); this last step was investigated in the presence of n = 0, 1, 3, 6 water molecules ($TS3_p^{nw}$).

substrate, again forming the Wheland complex. The latter can then transfer a proton back to the zeolite framework, giving the final ethylbenzene product (Figure 1d).

Interestingly, the formation of the Wheland complex—not only during benzene ethylation but in all S_EAr reactions—is still a matter of debate.[16] While in homogeneous catalysis its existence has been recently questioned,[16] the microporous environment of the zeolite is known to substantially stabilize charged species,[17,18] thus making its presence more likely. From a computational perspective, initial studies on cluster models indicated that the deprotonation step was occurring spontaneously and the Wheland complex was not a relevant reaction intermediate.[8,9] However, later studies showed that its stability is strongly dependent on the size of the cluster adopted in the calculation, becoming an effective minimum on the potential energy surface (PES) once the zeolite environment was better accounted for through the use of bigger cluster models.[11] As could be expected, with the subsequent passage to more realistic periodic models of the zeolite framework, the Wheland complex naturally emerged again as a stable product state of the S_EAr step, although its stability has not been deeply investigated so far.[12−14] On the other hand, while studying the effect of the zeolite topology on ethylbenzene transalkylation, Corma and co-workers found that the reaction can start from the *ipso* protonation of the substrate with an expected free energy increase of 44 kJ·mol^{-1} in the channel intersection of H-ZSM-5 at 573 K.[19,20] The authors also showed a remarkable dependence of the Wheland complex stability with respect to the chosen framework topology and location of the substrate within it.

Some additional insights can also be derived from the methanol-to-hydrocarbons process, where methylation of the aromatic hydrocarbon pool species also proceeds through a S_EAr mechanism.[21] Using static DFT calculations with higher order corrections, Fečík et al.[22] recently found the Wheland intermediate as a minimum in the PES when benzene is methylated with methanol or dimethyl ether. The deprotonation was found to proceed with a small barrier of 47 kJ·mol^{-1} and a large energetic gain of 127 kJ·mol^{-1}, indicating a very short lifetime for the intermediate.

Experimentally, the existence of the Wheland complex has been recently supported by Chowdhury and co-workers.[23] Through a series of *operando* UV−vis and NMR spectroscopy measurements, they were able to identify the Wheland complex and the SES as key intermediates present during the alkylation of benzene with ethanol over H-ZSM-5.

To thoroughly investigate the reaction mechanism at *operando* conditions, i.e., at elevated temperatures and in the presence of water, it is necessary to go beyond the standard static approach where only a few points on the potential energy surface are considered. In addition, the simple harmonic approximation normally adopted to evaluate free energies has proven to be often inadequate to correctly describe the large anharmonic motions of the adsorbates in the zeolite pores.[24−28] This especially reflects in the quality of the reaction entropy estimate and the mobility of the reactive species whose weight rapidly grows with the reaction temperature.[29,30]

In this work, we used enhanced sampling molecular dynamics (MD) simulations to overcome the intrinsic limitations of static methodologies and investigate the alkylation of benzene with ethene and ethanol at realistic operating conditions. First, two-dimensional multiple-walkers[31] well-tempered[32] metadynamics was used to freely explore the free energy landscape of the S_EAr reaction with ethene and the SES as ethylating agents, proving that the Wheland intermediate spontaneously appears as a minimum at *operando* conditions. Second, we fully characterized all of the alkylation transition states with umbrella sampling (US) and used classical transition-state theory to retrieve accurate kinetic constants for all reaction steps (which are graphically shown in Figure 1 and explicitly listed in Table 1). In that way, we showed that the use of enhanced sampling can remarkably change the activation energy of mobile transition states and possibly alter the mechanistic conclusions derived from static calculations. Finally, we investigated the *para* protonation of ethylbenzene (being energetically more favorable than the *ipso* one in the S_EAr) in the presence of various loadings of water in the zeolite unit cell. 0 ($TS3_p^{0w}$), 1 ($TS3_p^{1w}$), 3 ($TS3_p^{3w}$), and 6 ($TS3_p^{6w}$) water molecules were considered. In this way, we showed that water can act as a proton-shuttling agent and assist the protonation reaction. At the lowest coverage, this effect can increase the rate of proton exchange by about 1 order of magnitude. When higher loadings are considered, on the other hand, the BAS gets solvated by the water cluster as a hydronium ion. This effect tends to remarkably stabilize it, making the proton transfer kinetics slower and similar again to the anhydrous case at the higher considered loading.

Table 1. Nomenclature of the Reaction Steps Investigated in This Work

name	description	reactants	products
$TS0^{OH}$	SES formation from ethanol	ethanol	SES + water
$TS0^{=}$	SES formation from ethene	ethene	SES
TS1	SES electrophilic attack on benzene	SES + benzene	*ipso*-protonated Wheland intermediate
$TS2^{OH}$	ethanol electrophilic attack on benzene	ethanol + benzene	*ipso*-protonated Wheland intermediate + water
$TS2^{=}$	ethene electrophilic attack on benzene	ethene + benzene	*ipso*-protonated Wheland intermediate
$TS3_i$	*ipso*-protonated Wheland intermediate deprotonation	*ipso*-protonated Wheland intermediate	ethylbenzene
$TS3_p^{nw}$	*para* protonation of ethylbenzene with water	ethylbenzene + *n* water	*para*-protonated Wheland intermediate + *n* water

2. COMPUTATIONAL DETAILS

Catalyst Model

To represent the three-dimensional MFI pore structure of the H-ZSM-5 zeolite catalyst, we adopted a periodic model with a unit cell containing 96 tetrahedral Si atoms connected by oxygen bridges (Figure S1). As is commonly done,[33,34] the catalytic active site was modeled by substituting a Si^{4+} atom at the T12 position (located at the channel intersection) with Al^{3+} to give an Si/Al ratio of 95 and, therefore, isolated active sites. The negative charge created by the substitution was compensated with the addition of a proton on one of the oxygen atoms adjacent to the defective site ($O_{zeo,1}$ in Figure S1). In this way, the proton, i.e., the actual active BAS for the alkylation reaction, finds itself located at the intersection between the straight and sinusoidal channels of the MFI pore structure, ensuring maximal accessibility for the substrate. The initial location of the BAS is of limited importance as all oxygens in the first coordination sphere of the Al defect are equivalent in the collective variable definition (vide infra), allowing the BAS to spontaneously interact with the energetically most favorable site.

To model the reactions, a single molecule of each involved species was manually placed in the proximity of the BAS as initial structure for the MD simulations, using our in-house developed software Zeobuilder.[35] This means, for instance, a benzene and an ethene molecule were used for the reactant state of $TS2^{=}$ or a single ethanol molecule was used for the reactant state of $TS0^{OH}$ (Table 1).

Ab Initio Molecular Dynamics

All of the simulations performed in this work are based on enhanced sampling techniques, which rely on biased ab initio MD to derive the free energy profile for a chemical reaction (vide infra). This allows us to properly account for entropic effects by accurately simulating the dynamic behavior of the adsorbate molecules as well as accounting for the catalyst flexibility at operating conditions. Born−Oppenheimer MD simulations were performed using the CP2K software[36] (version 5.1) within the density functional theory (DFT) framework. We adopted the PBE exchange-correlation functional[37] in its parametrization revPBE[38] (due to its improved performance for reaction energies calculations[39]), coupled with Grimme's D3 dispersion correction[40] to account for long-range dispersive interactions. A triple-ζ quality basis set including valence and polarization functions was adopted for all atoms in the atom-centered Gaussian orbitals + plane waves (GPW)[41,42] basis set approach used by CP2K. GTH pseudopotentials[43] were used to smooth the electron density in the proximity of the nuclei, and the plane waves energy cutoff was set to 350 Ry. The time step for the integration of the equations of motion was set to 0.5 fs. All simulations were conducted in the NPT ensemble using a chain of five Nosé−Hoover thermostats[44,45] and an MTK barostat[46] to control temperature and pressure, respectively, that were set to 573 K and 1 atm to reproduce the experimental conditions.[23] The NPT ensemble was chosen to allow maximum catalyst flexibility, but due to the rigid nature of the H-ZSM-5 structure, the small fluctuations in the unit cell parameters are expected to have little influence on the results.

Collective Variables Definition

Since in regular MD simulations the chances of sampling activated events with a barrier larger than few times k_BT (like most chemical reactions) is extremely small, enhanced sampling techniques are used to bias the system along a set of properly chosen collective variables (CVs) and retrieve the activation energy of the process of interest. A CV is selected to monotonically vary in the function of the reaction progression from a value associated with the reactant state to a different value associated with the product state. All CVs considered in this work are function of the atomic coordinates of the system. Being interested in chemical reactivity, an ideal CV should be able to describe the formation/rupture of bonds while, at the same time, allow to account for the chemical equivalence of some atoms (e.g., the six aromatic carbons of benzene). Both goals can be achieved by using a linear combination of coordination numbers (CNs). The CN between two groups of atoms α and β is defined as

$$\mathrm{CN}(\alpha;\beta) = \sum_{i\in\alpha}\sum_{j\in\beta}\frac{1-\left(\frac{r_{ij}}{r_0}\right)^{NN}}{1-\left(\frac{r_{ij}}{r_0}\right)^{MM}} \tag{1}$$

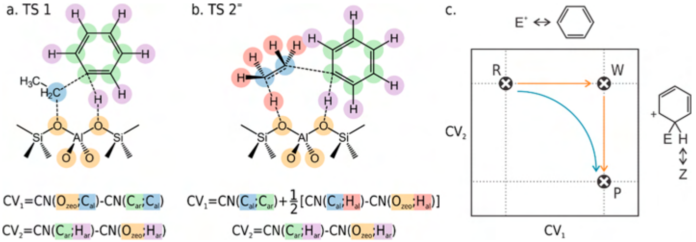


Figure 2. (a, b) Definition of the coordination numbers (CNs) and collective variables (CVs) used in the metadynamics simulations of TS1/TS3$_i$ (a) and TS2$^=$/TS3$_i$ (b). (c) Schematic depiction of the expected features of the two-dimensional MTD free energy surface as a function of CV$_1$ (describing the electrophilic attack on benzene) and CV$_2$ (describing the deprotonation of the Wheland complex). The location of reactants (R), products (P), and Wheland complex (W) is shown. The orange arrows follow a hypothetical path going through the formation of a stable Wheland complex, while the blue one shows a concerted conversion from R to P.

where r_{ij} is the distance between atom i, belonging to group α, and atom j, belonging to group β. Unless differently specified, NN = 6 and MM = 2NN. The function argument of the double summation tends to 1 if $r_{ij} < r_0$ and quickly drops to 0 otherwise. In our simulations, the parameter r_0 is selected to be roughly equal to the length that a bond between two atoms of the α and β groups assumes in the transition state region. In that way, there will be a contribution of ~1 to CN(α; β) if i and j are bounded ($r_{ij} < r_0^{TS}$) and of ~0 otherwise, thereby making CN(α; β) approximately equal to the number of bonds between group α and group β.

Almost all the reactions investigated in this work involve the simultaneous formation and rupture of bonds between at least three (sometimes four) group of atoms. However, obtaining converged free energy surfaces (FESs) in more than two dimensions is extremely costly and often unfeasible.[47] To reduce the dimensionality of the problem, we selected our final collective variables to be a linear combination of CNs. In practice, two types of linear combinations were used: the first one (CV$_A$) is the simplest and is used to describe a reaction in which one atom (whether or not it belongs to a larger molecular fragment) is transferred from a group of atoms to another. This is the case, for instance, in TS3$_i$, where a single proton is transferred between the *ipso* aromatic carbon and the zeolite oxygens. For this kind of process, a simple difference of CNs can appropriately describe the reaction (eq 2).

$$CV_A = CN(\alpha;\, \beta) - CN(\beta;\, \gamma) \tag{2}$$

In practice, an atom belonging to group β (simply the proton in the TS3$_i$ example) is transferred from group α (the *ipso* aromatic carbon) to group γ (the four oxygen atoms around the Al site) or vice versa, adding 1 to one of the CNs while removing 1 to the other. Therefore, this type of CV will vary of about two units between reactants and products (see Figure S2 and CV$_2$ in Figure 2).

More complex is the specific case in which ethene is reacting, either to form the SES intermediate (TS0$^=$) or directly attacking benzene (TS2$^=$). In this case, two processes are occurring simultaneously: a proton is transferred from the zeolite to one of the ethene carbon atoms while the other carbon attacks the nucleophilic moiety (either another zeolite oxygen or the benzene molecule). To avoid an excessive increase in the number of dimensions, the following linear combination of CNs was used as collective variable:

$$CV_B = CN(\alpha;\, \beta) + \frac{1}{2}[CN(\beta;\, \gamma) - CN(\gamma;\, \delta)] \tag{3}$$

TS2$^=$ can be taken as an example to explain this choice (see Figure 2b). First, ethene is activated by the zeolite BAS, that protonates it. Similarly to the previous example with TS3$_i$, the reaction can be described with a difference in coordination numbers (CN(β; γ) − CN(γ; δ) in eq 3). In this case, β includes the two aliphatic carbons while γ the four aliphatic hydrogens and the BAS, as they become indistinguishable once the protonation has occurred. CN(β; γ) is then going to vary from about 4 (the number of C−H bonds in ethene) to 5 (the number of C−H bonds in an ethyl fragment). δ includes the four oxygen atoms surrounding the Al site of the framework, and therefore, CN(γ; δ) goes from 1 (BAS on the framework) to 0 (BAS on the ethyl fragment). Overall, CN(β; γ) − CN(γ; δ) is then equal to about 3 in the reactant region and 5 in the product region. If α is chosen to include the six aromatic carbons, CN(α; β) can describe the C−C bond formation between benzene and the ethyl fragment, with an expected variation from 0 to 1. The dimensionality of the FES can then be reduced by summing the two partial CVs but, since CN(β; γ) - CN(γ; δ) varies of two units while CN(α; β) only of one, the former is first multiplied by 0.5. This should allow for a smoother variation of the CV from reactants to products. A graphical depiction of the aforementioned steps can be seen in Figure S3.

Every collective variable used in the advanced sampling simulations either falls in one of the two previous categories or simply is a single coordination number. More specific information is reported in the following sections.

Well-Tempered Metadynamics

Well-tempered metadynamics (MTD) and umbrella sampling (US) simulations are the two advanced sampling techniques that were chosen to explore the FES of benzene ethylation by biasing the system in a one- or two-dimensional CV space. For both techniques, the bias was applied to the MD simulations using PLUMED[48] (version 2.4.0) as dependency of CP2K. Initially, MTD was used to explore the FES of the S$_E$Ar step in benzene ethylation with a SES (TS1 in Figure 1) and with ethene (TS2$^=$) and assess whether the Wheland complex spontaneously appears as reaction intermediate. To make the formation of the Wheland complex possible—but not mandatory—a two-dimensional CV space is needed in which the first CV describes the electrophilic attack on the benzene ring while the second the deprotonation of the Wheland complex (Figure 2c). In the case of TS1, the electrophilic attack can be described with a difference of coordination numbers as shown in eq 2, by defining $\alpha = C_{ar}$ (the six aromatic carbons), $\beta = C_{al}$ (the reactive alkyl carbon), and $\gamma = O_{zeo}$ (the four oxygen atoms in the first coordination sphere of the Al defect, Figure 2a). For $TS2^{=}$, the more complex linear combination of CNs described by eq 3 must be used with, as explained before, $\alpha = C_{ar}$, $\beta = C_{al}$ (this time considering both the alkyl carbons), $\gamma = H_{al}$ (the alkyl hydrogens including the BAS), and $\delta = O_{zeo}$ (Figure 2b). For $TS3_i$, a difference of CNs ($\alpha = C_{ar}$, $\beta = H_{ar}$, i.e., the six aromatic hydrogens and $\gamma = O_{zeo}$) can again be used. By using MTD to explore this 2-dimensional CV space, the stability of the Wheland complex can be elucidated. Indeed, if its formation is spontaneous during the reaction, a minimum in the FES will be sampled at the corresponding CVs values (orange arrows in Figure 2c), while a direct interconversion between reactants and products will be observed otherwise (blue arrow in Figure 2c).

Concerning the methodology, two multiple-walkers[31] well-tempered[32] metadynamics[49,50] simulations were performed. In this approach, the bias potential is constructed on the fly by spawning a Gaussian-shaped hill at regular time intervals, centered on the average CV position explored by the walker in a certain number of previous simulation steps. Every walker feels, at every simulation step, an overall potential given by the sum of all walkers' spawned hills up to that moment. Six walkers were run in parallel, three initially located in the reactant basin (benzene + ethene for $TS2^{=}$ or benzene + SES for TS1, see Table 1) and three in the product (ethylbenzene) basin. The hills were spawned in the CVs space every 50 fs, with an initial height of 5 $kJ\cdot mol^{-1}$. Such height was then gradually rescaled according to the well-tempered recipe, with an initial BIASFACTOR—as defined by PLUMED—of 10, subsequently increased to 15 to improve convergence. The mass of all hydrogen atoms was set to 2 to improve the stability of the simulations. The simulations were stopped once the hills height dropped below 1 $kJ\cdot mol^{-1}$, and a few barrier recrossings were observed among the various walkers. The final estimate of the FES was then obtained by inverting the sum of the hills spawned by all walkers. A set of walls was required to improve barrier recrossing or avoid the occurrence of side reactions, the details of which are given in the Supporting Information, Section S2.4.

Umbrella Sampling

Reaching a satisfactory convergence of the FES estimate in two-dimensional MTD is extremely difficult and requires a prohibitively long computational time. For this reason, we subsequently used one-dimensional umbrella sampling[51,52] (US) simulations to derive the FES of all the benzene ethylation reaction steps, both with ethene and ethanol (see Figure 1). Moreover, US was also used to investigate the role of water in the *para* protonation of ethylbenzene when various loadings are considered ($TS3_p^{nw}$, $n = 0, 1, 3, 6$). The adopted CVs are fundamentally analogous or precisely the same to the one presented for MTD in Figure 2, belonging to one of the two categories described by eqs 2 and 3. An exception is represented by $TS3_p^{nw}$ ($n \neq 0$), in which a single CN between the ethylbenzene *para* carbon and all hydrogen atoms involved in the protonation process was sufficient to sample the reaction. A full list of the US CVs is reported in Table S1 and graphically depicted in Figure S4.

In a US simulation, various quadratic potentials (the "umbrellas") are set along the selected CV from the reactants to the products basin and a MD simulation is then run in each of them. The bias has the form

$$V_i(\mathrm{CV}) = \frac{\kappa_i}{2}(\mathrm{CV} - \mathrm{CV}_{0,i})^2 \tag{4}$$

where $V_i(\mathrm{CV})$ is the bias of the ith umbrella, κ_i is its spring constant, and $\mathrm{CV}_{0,i}$ is the collective variable value at which it is centered. The κ_i and $\mathrm{CV}_{0,i}$ choice was guided by previous literature reports[30,53] and tuned to balance a uniform sampling between reactant and products while keeping low the number of required umbrellas. Being an equilibrium technique and, therefore, intrinsically more stable than MTD, the hydrogen mass in US was set back to 1. The FES estimate was obtained by combining the simulations through the weighted histogram analysis method[54,55] (WHAM), as implemented in our in-house developed ThermoLIB library.[56] A full list of the umbrella parameters used in the simulations for all transition states is reported in the Supporting Information, Tables S2−S9.

While using a linear combination of coordination numbers allows us to perform 1-dimensional US simulations, which is much faster to converge than multidimensional US, the risk of poorly exploring some important regions of the phase space increases the more degrees of freedom are "squeezed" in a single linear combination. To ensure that an appropriate sampling of the relevant phase space regions in every reaction was achieved, the 1-dimensional free energy profiles have been expanded in terms of their constituting coordination numbers using the statistical analysis tools of ThermoLIB,[56] to obtain 2-dimensional FESs. More details on the procedure can be found in the Supporting Information, Section S2.3. While for most reactions a good sampling of the reaction path was observed, it was noticed that some possibly important regions around the transition state of $TS0^{=}$ were poorly sampled. To solve the problem, some extra 2-dimensional umbrellas were added to reach a satisfactory coverage of the transition state region and the complete 2D FES was then reprojected onto the original 1D CV (more details are reported in Section S2.3 of the Supporting Information).

As for the MTD case, most of the simulations required some walls to prevent undesired side reactions or facilitate convergence by improving barrier recrossing. Full details are reported in the Supporting Information, Section S2.4.

Calculation of the Phenomenological Reaction Barriers

As recently shown by some of us,[53,57] the information derived from an US simulation can be combined to classical transition state theory (TST) to remove the dependency on the chosen CV and retrieve reliable kinetic constant for the reaction of interest. In this framework, the reaction kinetic constant can be written as

$$k_{\mathrm{TST}} = \sqrt{\frac{1}{2\pi\beta}}\,\langle|\vec{\nabla}_x \mathrm{CV}|\rangle_{\mathrm{CV}^{\ddagger}} \frac{\mathrm{e}^{-\beta F(\mathrm{CV}^{\ddagger})}}{\int_{-\infty}^{\mathrm{CV}^{\ddagger}} \mathrm{e}^{-\beta F(\mathrm{CV})}\,\mathrm{dCV}} \tag{5}$$

in which $\beta = 1/k_B T$ and $\mathrm{CV}^{\ddagger}$ is the value of the collective variable corresponding to the reaction transition state. $\langle|\vec{\nabla}_x \mathrm{CV}|\rangle_{\mathrm{CV}^{\ddagger}}$ is the ensemble average of the CV gradient with respect to the mass-weighted coordinates of the system, computed when the CV is restricted at the transition state value. From the kinetic constant, it is then possible to use Eyring's equation in order to retrieve a so-called phenomenological reaction barrier no longer dependent on the CV choice

$$\Delta F_f^{\ddagger} = -\frac{1}{\beta}\ln(k_{\mathrm{TST}}\beta \mathrm{h}) \tag{6}$$

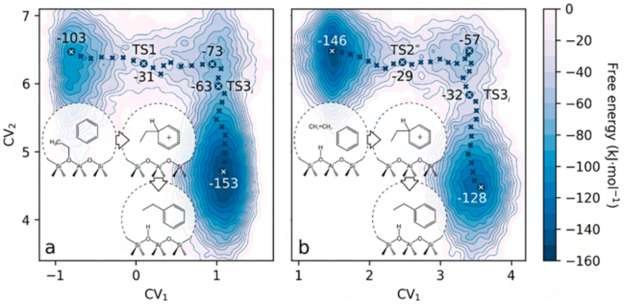


Figure 3. Free energy surfaces related to the alkylation of benzene with an SES (a) and with ethene (b), as obtained from MTD simulations. Isoenergetic lines are placed every 10 kJ·mol^{-1}. Black crosses show the minimum free energy path (MFEP) connecting reactants and products, obtained according to the procedure of ref 58. Free energy values for the stationary points along the MFEP are reported in kJ·mol^{-1}. For the CVs definition, see Figure 2.

with *h* being Planck's constant. Statistical analysis was used to compute the 95% confidence interval for the kinetic constant and the phenomenological barrier, more details can be found in Section S2.5 of the Supporting Information.

3. RESULTS AND DISCUSSION

Probing the Wheland Complex Formation within the Zeolite Channels

To assess whether the Wheland complex is part of the reaction intermediates during the ethylation of benzene at *operando* conditions, we investigated the S_EAr step with 2-dimensional multiple-walkers MTD considering both the SES (TS1/$TS3_i$) and ethene ($TS2^=$/$TS3_i$) as alkylating agents to probe the effect of a mobile vs chemisorbed electrophilic agent.

By adopting two specific CVs, one describing the electrophilic attack while the other the deprotonation step (vide supra, Figure 2c), the system can evolve from reactants to products with the formation of the Wheland complex being possible but not mandatory. With each of the 6 parallel walkers in the two simulations running for about 70 ps, a total of more than 420 ps of total simulation time per mechanism was necessary. The final FESs are shown in Figure 3.

Both profiles look qualitatively very similar. Interestingly, it is clearly visible how the S_EAr proceeds in two distinct steps: first, the electrophile attacks the benzene ring (TS1 and $TS2^=$ in Figure 3a,b, respectively) with the formation of the Wheland complex. The raw reaction barrier, obtained from the minimum free energy path,[58] is higher for $TS2^=$ than for TS1 (117 vs 72 kJ·mol^{-1}, respectively). This can be expected due to the preactivated nature of the SES with respect to ethene. The differences in activation energy will be further discussed in the following paragraph while analyzing the US results.

The Wheland intermediate is rather elusive and quickly deprotonates to ethylbenzene, restoring its aromaticity. Indeed, its corresponding minimum (top right of Figure 3a,b) is very shallow and a small barrier (10−25 kJ·mol^{-1}, depending on the considered profile) is associated with its deprotonation, which is also largely exergonic (80−71 kJ·mol^{-1}).

The differences observed in the $TS3_i$ barriers for the two cases are to be expected, as converging 2-dimensional free energy surfaces with MTD is extremely expensive. On the other hand, the purpose of these simulations was to freely explore the free energy surface of the S_EAr reaction, from which it clearly emerged that the Wheland complex has to be considered as reaction intermediate. For this reason, 1D umbrella sampling simulations were subsequently performed on all the reaction steps, to retrieve accurate barriers and unravel the full reaction mechanism when ethene and ethanol are used as alkylating agents.

Reaction Mechanism of Benzene Ethylation

An overview of all the reactions investigated with US is shown in Figure 1 and in Table 1. Six separate US simulations were performed, two for the formation of the SES ($TS0^=$ and $TS0^{OH}$) and one for its electrophilic attack on the benzene ring (TS1), two for the direct ethylation of benzene ($TS2^=$ and $TS2^{OH}$), and finally, one for the Wheland complex deprotonation ($TS3_i$). As previously stated, a full list of the adopted umbrellas and the walls adopted in each simulation can be found in the Supporting Information, Sections S2.2 and S2.4.

As previously explained, all reaction profiles were deprojected as a function of the single CNs constituting the final CV to ensure that the path connecting reactants and products was adequately sampled. All of the raw free energy profiles and the respective 2-D expansion are shown in Section S3.1 of the Supporting Information. Only in the case of $TS0^=$ was it found that the transition state region had possibly relevant portions of the phase space not properly sampled by the 1D umbrellas. For this reason, extra 2-dimensional umbrellas were added to improve the sampling and the final 2D FES reprojected on one dimension to obtain the final phenomenological barrier (Section S2.3).

A full overview of the reaction free energy profile can be seen in Figure 4a. Note that the alignment between reactant and product state of different reactions is done for graphical purposes only, as even for reactions with the same reactant/product state (for instance the *ipso*-protonated Wheland intermediate without coadsorbed molecules is formed both in TS1 and $TS2^=$) there is no absolute guarantee that the same phase space was explored in both simulations.[59]

Starting from the stepwise path (TS0−TS1), the formation of the SES is the rate-determining step for both ethene and ethanol, which present a barrier of 101 ± 7 and 115 ± 3 kJ·mol^{-1}, respectively. The formation of the SES from ethanol is therefore more difficult than from ethene as, even if the opposite extremes of the confidence interval are considered, a minimum difference of 4 kJ·mol^{-1} is present between the two barriers. At the reaction

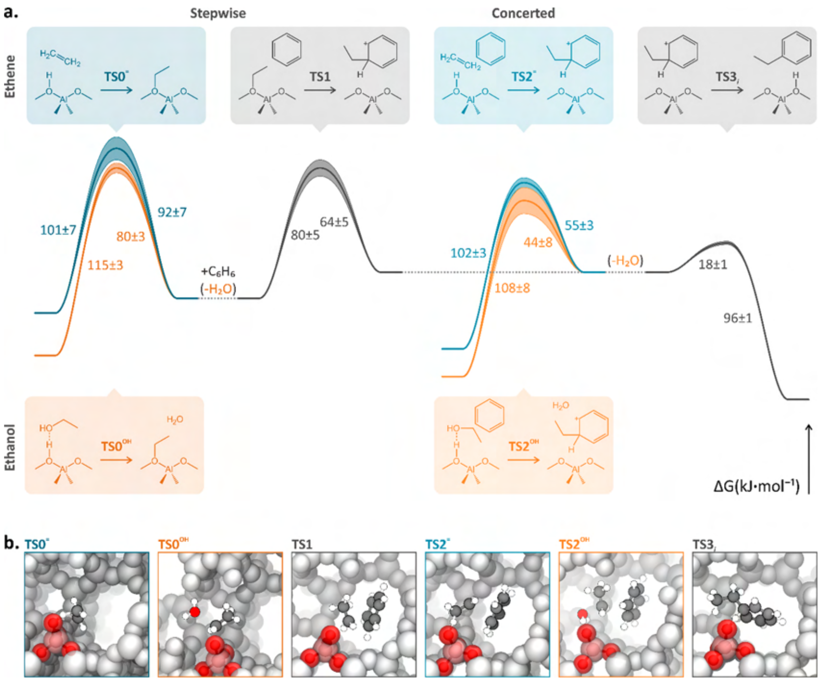


Figure 4. (a) Complete free energy profile for the ethylation of benzene with ethene and ethanol, as derived from ab initio umbrella sampling simulations. The numbers close to the curves report the respective reaction phenomenological barrier (in kJ·mol^{-1}) together with the respective 95% confidence interval (also graphically shown by the shaded regions around the curves). The alignment of reactant and product states between different reactions is for graphical reasons only (see main text). (b) Cross-section of the zeolite model for the various reactions as seen along the sinusoidal channel, showing a representative snapshot arbitrarily extracted from the umbrella on top of the transition state. All zeolite atoms, except for the Al tetrahedra, are in white for the sake of clarity.

conditions, such difference would still imply that the formation of the SES from ethene is more than twice as fast as from ethanol. This is in line with previous reports in the literature,[13] in which it was shown that ethanol can adsorb more strongly on the BAS thanks to the formation of an H bond with its hydroxyl group, thereby stabilizing the reactant state. A strong interaction between the BAS and the hydroxyl group of the ethanol is also observed in the umbrella corresponding to the reactant state of TS0^{OH}, where a hydrogen bond between ethanol and the BAS persists for the whole simulation time (Figure S14). The hydroxyl group hydrogen can also form a second hydrogen bond with the zeolite oxygens that is, however, significantly weaker (Figure S15). While potential walls are present in the simulations to explicitly prevent the molecules from leaving the active site, ethene appears to be quite more mobile and prone to diffuse away (Figure S16). Once formed, the SES can readily react with benzene, with a much lower barrier of 80 ± 5 kJ·mol^{-1}, to produce the *ipso*-protonated Wheland intermediate as shown in the MTD simulations.

In the direct ethylation (TS2), ethene or ethanol are directly activated by the BAS and attack the benzene ring. Also in this case ethanol presents a slightly higher barrier than ethene (108 ± 8 vs 102 ± 3 kJ·mol^{-1}, respectively), however, the difference is rather small certainly given the large error bars. These results are potentially counterintuitive, as both TS0 and TS2 consist in an activation of the electrophile (ethene or ethanol) with concerted attack on an electron-rich moiety (the zeolite oxygens or the benzene carbons). Therefore, one could expect a proportional difference in activation energies between ethanol and ethene, while for the former attacking benzene seems to be easier than attacking the framework. An explanation for this can be found in the transition-state geometries. Indeed, to form the SES (TS0^{OH} in Figure 4b), ethanol is completely protonated by the BAS forming an $EtOH_2^+$ cationic species, which must then rotate to expose the electrophilic carbon toward the framework and undergo the S_N2 reaction. On the other hand, when reacting with benzene, the proton transfer to the hydroxyl group can occur gradually while the reaction proceeds, as no large reorientation is needed to form the C−C bond (TS2^{OH} in Figure 4b). This can also be seen by expanding the free energy profile in term of a new collective variable, encoding the proton transfer from the zeolite to the hydroxyl group (Figure S17). In the case of TS0^{OH}, the BAS has been fully transferred from the zeolite to the hydroxyl group when the transition state is reached, while for TS2^{OH} the transfer is still progressing and the partially positive H_2O moiety remains in interaction with the zeolite framework. Similarly to what has already been observed in the methanol-to-hydrocarbon process, the presence of extra protic molecules could assist the proton transfer from the framework to the reacting ethanol molecule, facilitating the

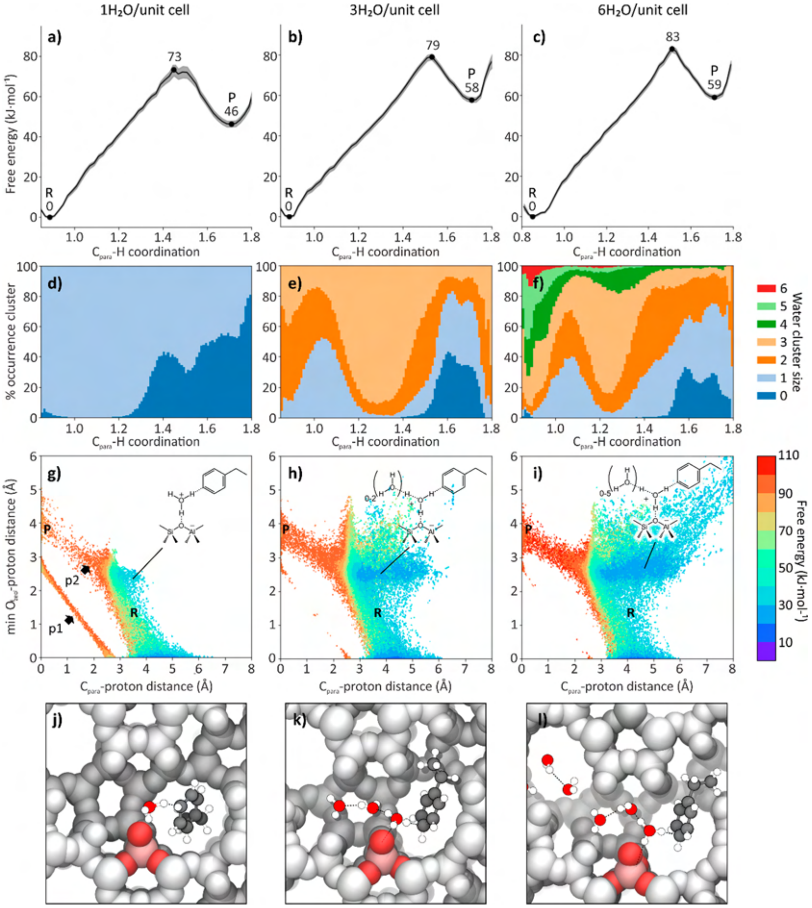


Figure 5. (a−c) One-dimensional free energy profiles obtained from the umbrella sampling simulations of the *p*-ethylbenzene protonation in the presence of 1 (a), 3 (b), and 6 (c) water molecules per unit cell. (d−f) Percentage of occurrence of a certain water cluster size in proximity of the active site as a function of the reaction collective variable. (g−i) Two-dimensional free energy surfaces obtained by expansion of the original one-dimensional profile showing the free energy as a function of the distance between the *para* carbon and the proton location (see main text) and the minimum distance between the first-coordination sphere O atoms around the Al defect and the proton location. (j−l) Cross-section of the zeolite model, as seen along the sinusoidal channel, showing a representative snapshot arbitrarily extracted from the umbrella in proximity of the transition state. All zeolite atoms, except for the AlO_4 tetrahedra, are in white for the sake of clarity.

formation of the SES.[60,61] Therefore, the differences between ethanol and ethene for the two mechanisms could significantly change at high ethanol or water loadings.

The reaction barriers are in line with previous literature reports (for a full comparison see Table S13). While it is known that activation energies obtained with improved level of theory are in general higher,[12] an interesting comparison can be made with the static results reported by Wang et al.,[13] where PBE-D3 on a periodic H-ZSM-5 model is also used. The barriers for $TS0^{=}$, $TS0^{OH}$, and TS1 are in very good agreement, while a significant deviation can be seen for $TS2^{=}$ (∼−17 $kJ{\cdot}mol^{-1}$) and $TS2^{OH}$(∼+18 $kJ{\cdot}mol^{-1}$). This can be explained considering that TS0 and TS1 are quite rigid transition states, involving either the formation of a bond with the zeolite framework (TS0) or the transfer of a framework-bounded species to the hindered benzene molecule (TS1). On the other hand, in TS2, the ethylating agent interacts contemporarily with the benzene and the BAS and quite a large mobility of the species is observed (Figure S18). Since the geometries explored by the molecules during the dynamic simulations do not dramatically deviate from the optimized transition state geometries of Wang et al.,[13] the relatively large difference in the computed barrier could be likely attributed to entropic effects arising from the large anharmonic motions of the adsorbates in the zeolite pores.

The change in activation energy between static and dynamic simulations can have an important effect on the preferred mechanism. While it is known that microkinetic modeling is needed to properly include all the reaction variables,[13,62] this is outside the scope of this work. On the other hand, the ratio between the forward kinetic constants of the stepwise and concerted pathways can already provide some interesting mechanistic insights. By considering the activation energies computed by Wang et al.,[13] $TS2^{=}$ is about an order of magnitude faster than $TS0^{=}$ while the opposite is true for ethanol (in line with the full results of the microkinetic model). On the other hand, by considering our barriers, we learn that both the stepwise and concerted mechanisms proceed with a similar speed when the same ethylating agent is considered, with ethanol likely favoring the concerted path (despite the partial overlap between the confidence intervals). It must be pointed out that Chowdhury et al.[23] observed the SES formation using an equimolar amount of benzene and ethanol in the reacting mixture. This is a quite high ethylating agent concentration, which for normal industrial applications is kept about 5 times lower than benzene to suppress the formation of polyalkylated products.[63] According to our results, the formation of the SES in the reaction environment does not derive from a strong energetic difference between the two mechanisms—as found by Wang et al.—but more likely from the statistical unlikeliness of having benzene and ethanol simultaneously coadsorbed on the BAS, which becomes more significant at low benzene concentrations. This highlights the importance of the use of advanced sampling methods in zeolite catalysis, where certain reactive events can exhibit significant entropic effects that are not possible to fully capture with the traditional static approach.[64]

Once the alkylation has occurred, the newly formed *ipso* protonated Wheland complex can easily deprotonate to give the final ethylbenzene product, with a barrier of only 18 ± 1 kJ·mol^{-1} and a large energetic gain (78 kJ·mol^{-1}). These results point toward an extremely short lifetime for the Wheland complex, making it present in minimal concentrations in the reaction environment. It must be pointed out, on the other hand, that the ethylbenzene product carries an electron-donating substituent which should, in principle, facilitate the protonation of the *ortho* and *para* positions on the aromatic ring. Moreover, when ethanol is used as ethylating agent, water is formed during the reaction. Its effect on the protonation kinetics is rather unexplored and, therefore, we also investigated the *para* protonation reaction of ethylbenzene with various water loadings in the zeolite (0, 1, 3, and 6 molecules per unit cell). In this way we assessed whether−under specific conditions−the formation of Wheland complexes could become more favorable.

Water Influence on the Wheland Complex Formation

It is well-known that when water adsorbs in the H-ZSM-5 pores it can strongly interact with the BAS. When the number of water molecules per BAS exceeds 3, the proton can even become fully solvated as hydronium ion.[65−67] It goes without saying that this phenomena can have an impact on proton transfer reactions and, therefore, on the formation of the Wheland complex in the zeolite pores.

To assess the effect of water on the formation kinetics and stability of the Wheland complex we performed four separate US simulations with varying amount of water in the zeolite unit cell, namely 0, 1, 3, and 6 water molecules. Considering that the maximal amount of water that can adsorb on the BAS before condensation at 298 K is around 7−8 molecules in H-ZSM-5,[68] the chosen coverages should be representative for a wide range of water partial pressures. As previously stated, we focused on the *para* protonation of the ethylbenzene product, which should be significantly more favorable than the *ipso* protonation constituting the last step of the ethylation reaction.

The raw reaction profiles for the ethylbenzene *para* protonation in the presence of various amounts of water are shown in Figure 5a−c. The forward and backward phenomenological reaction barriers are listed in Table 2. Starting from the anhydrous case, the protonation on the *para* carbon exhibits—as expected—a significantly lower activation energy than the *ipso* one, going from 96 to 76 kJ·mol^{-1}. The backward barrier, on the other hand, is computed to be exactly the same (18 kJ·mol^{-1}), thereby making the *para*-protonated Wheland complex 20 kJ·mol^{-1} more stable than the *ipso*-protonated one. By looking at the effect of water, the forward barrier experiences a strong drop, from 76 ± 2 to 62 ± 3 kJ·mol^{-1}, when one water molecule is introduced in the unit cell. By increasing the amount of water to 3 molecules per unit cell, the barrier increases to 68 ± 1 kJ·mol^{-1}, reapproaching then the anhydrous one with the highest considered loading of 6 water molecules per unit cell (72 ± 2 kJ·mol^{-1}). The backward barrier, on the other hand, experiences a much less pronounced variation, from a minimal value of 11 kJ·mol^{-1} for 3 water molecules per unit cell to a maximum of 18 kJ·mol^{-1} for 0 and 1 water molecules per unit cell.

Table 2. Forward and Backward Phenomenological Barriers (kJ·mol^{-1}) for the *Para* Protonation of Ethylbenzene in the Presence of Various Amounts of Water in the Unit Cell

	no. of water/unit cell			
	0 H_2O	1 H_2O	3 H_2O	6 H_2O
$\Delta F^{\ddagger}_{f}$	76 ± 2	62 ± 3	68 ± 2	72 ± 2
$\Delta F^{\ddagger}_{b}$	18 ± 2	18 ± 3	11 ± 2	14 ± 2

To better understand the reasons underlying these variations in the forward barriers, we analyzed the size of the water cluster being actively involved in the reaction as a function of the reaction CV. In other words, we determine for each simulation step in all umbrellas the amount of water molecules that are in close proximity of the active site, thereby giving an indication if the water—which is free to move around in the simulation cell—is actively taking part to the reaction. More information about the definition of the cluster size can be found in Section S5.1 of the Supporting Information. The cluster sizes were collected as a function of the reaction CV, and their fraction of occurrence is shown in Figure 5d−f. Some interesting observations can be made: first, it can be noticed that the cluster size tends to be significantly smaller in the product region than in the reactant one. For instance, in the simple case of 1 water molecule per unit cell, the fraction of samples in which the water is far from the active site (cluster size of 0) tends to increase going toward the products and the same can be said for both the 3 and 6 water molecules cases (dark blue bars in Figure 5d−f). This is caused by the strong affinity between the water molecules and the BAS which, in the reactant state, are free to interact. When the proton is transferred to the ethylbenzene molecule, on the other hand, its ability to form hydrogen bonds with water is strongly inhibited and the latter tends then to diffuse away from the Al defect. With water poorly interacting or even leaving the active site region in the product state, it can be expected that not much influence will be seen in the reactivity of the Wheland complex toward deprotonation, in line with the low variation in the backward phenomenological barriers (Table 2).

Interestingly, it is also possible to notice how the transition state region for 3 water molecules (Figure 5e) is associated with a preference for larger cluster sizes than the reactant state. In the latter, cluster sizes of 1, 2, and 3 water molecules are similarly sampled in the simulation, but in the central region a large prevalence of a 3 molecules cluster is observed. In the reactant state, the BAS interacts strongly with one water molecule but, at the relatively high simulation temperature, interwater interactions are not strong enough to keep the cluster together. By pushing the system to react, on the other hand, the proton is forced toward the ethylbenzene and a transient hydronium ion is formed when water act as proton transfer bridge (Figure 5k). This hydronium ion is a strong H-bond donor and can be stabilized by the partial solvation offered by the remaining two water molecules, which remain tightly bounded over most of the simulation time.

For the 6 water molecules per unit cell the situation is quite different. In the reactant state there are enough water molecules to fully solvate the BAS as hydronium ion, thereby creating a reactive cluster mainly consisting of 3−5 water molecules (Figure 5f). By forcing the system toward the transition state, the hydronium ion is forced at the edge of the water cluster in order to interact with the ethylbenzene (Figure 5l). This reduces the possibility of being surrounded by the remaining water molecules and a smaller predominant cluster size of 3 is observed in the central region of the CV range.

While analyzing the size of the reactive water cluster gives valuable information about the chemistry of the system, we also adopted a second—more quantitative—approach to investigate the role of water on the reaction energetics. The 1-dimensional free energy profile was expanded in terms of two new collective variables, namely the distance between the ethylbenzene *para* carbon and the atom carrying the extra proton in the system and the minimum distance between the oxygen atoms in the first coordination sphere of the Al defect and, again, the atom carrying the extra proton. The latter is defined using a combination of the definitions proposed by Pérez de Alba Ortíz et al.[47] and Grifoni et al.[69] (more detailed information can be found in the Supporting Information, Section S5.2). The two-dimensional FESs for the various water loadings can be seen in Figure 5g−i. Starting from the lower loading of 1 water molecule per unit cell (Figure 5g), two separate paths can be seen going from the bottom (min O_{zeo}−proton distance = 0, proton on the zeolite) to the left (C_{para}−proton distance = 0, proton on the ethylbenzene *para* carbon). The first one (p1 in the figure) corresponds to the anhydrous protonation, in which the BAS directly jumps on the ethylbenzene without water mediation (corresponding to the states with a cluster size of 0 around the transition state region in Figure 5d). The second path (p2), on the other hand, indicates that the water molecule can also act as proton transfer medium between the zeolite and the ethylbenzene (Figure 5j). The opening of this second possible protonation path which, as visible, is quite wide and similar in energy to the anhydrous protonation, is the reason for the large decrease in forward barrier going from 0 to 1 water molecule per unit cell.

Going toward higher water loadings, two things are visible. First, the anhydrous protonation path becomes less prominent as, with more water molecules in the catalyst unit cell, it is less and less likely for all of them to diffuse away from the active site, especially in the reactant region (compare with Figure 5e,f). Second, the number of states where the proton is solvated by the water as hydronium ion increases drastically, and in the case of 6 water molecules per unit cell, the proton can diffuse quite far from the active site while jumping from water to water. The creation of more and more stable states in the reactant region—due to the large mobility of the proton at high water loadings, is responsible for its increased stabilization and, ultimately, for the progressive increase of the forward reaction barrier. The transition state region, on the other hand, is qualitatively independent of the water loading. There, the hydronium ion must move toward the edge of the water cluster to interact with ethylbenzene (Figure 5k,l). This limits the possible interactions between the other water molecules and the hydronium ion and only 2−3 of them remain close by, as highlighted by the lack of cluster sizes greater than 4 in the transition state region of Figure 5f. Apart from the possibility of forming a few extra hydrogen bonds with the hydronium ion, the presence of extra water molecules in the catalyst does not significantly change the main features of the protonation transition state. Therefore, the changes in the reaction barrier can be mostly attributed to the stabilization of the reactant state deriving from the solvation of the hydronium ion when the number of adsorbed water molecules goes beyond one.

According to our simulations, water is expected to play a key role in modulating protonation reaction kinetics in zeolites. When looking at the forward reaction rate constant as a function of the water content (Figure 6), an increase of more than 1 order of magnitude is expected when one water molecule is introduced in the reaction environment. The rate then progressively decreases with increasing amounts of water as the solvation of the BAS stabilizes the reactant state. An interesting comparison can be made with the rate of H/D exchange for benzene, which should proceed through a Wheland-like transition state. Chen et al. found through NMR spectroscopy measurements that the presence of trace amounts of water can increase the speed of benzene H/D exchange at room temperature, but only for the relatively low Si/Al ratio of 15 and not a higher ones of 40.[70] At higher—but still substoichiometric—water concentrations the exchange rate was found to be slower in all cases. Our

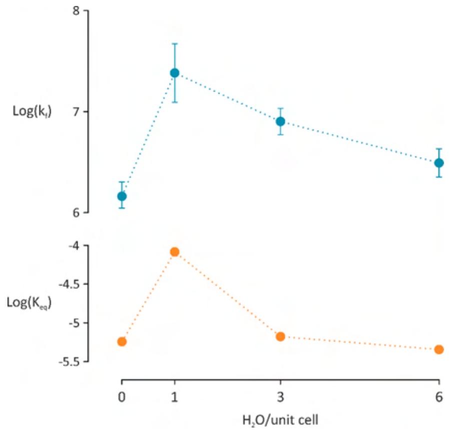


Figure 6. Forward kinetic constant (k_f) and equilibrium constant (K_{eq}) for the *para* protonation of ethylbenzene as a function of the number of water molecules in the zeolite unit cell. Dotted lines are present to guide the reader's eye only.

computational model is certainly more in line with the Si/Al = 40 case consisting of isolated active sites, where a slowing down of the rate was experimentally observed for any amount of adsorbed water. It must be pointed out, however, that water is known to not uniformly distribute on the zeolite active sites and rather form heterogeneous clusters where some of the BAS are solvated while other basically anhydrous.[71] This means that the boost in activity that calculations suggest for 1 water molecule per BAS is unlikely to be measurable experimentally and rates similar or lower to the anhydrous case are going to be observed instead. At lower Si/Al ratios, cooperative effects between proximal BAS could significantly change the chemistry of the system and such effect is certainly worth of future investigation.

Since, as shown before, the reverse phenomenological barrier remains quite similar for all water loadings, the equilibrium constant of the protonation reaction follows a similar trend as the forward kinetic constant (Figure 6). In practice, water can speed up the protonation while interacting with the BAS, but then diffuses away once the proton has been transferred, thereby not changing the deprotonation kinetics. This reflects in an increase of the equilibrium constant from 6×10^{-6} in the anhydrous case to 8×10^{-5} in the presence of one water molecule. With 3 and 6 water molecules in the unit cell, the equilibrium constant decreases back to 7×10^{-6} and 4×10^{-6}, respectively. Even with the most favorable conditions, the equilibrium constant remains always in strong favor of neutral ethylbenzene, suggesting that the Wheland complex will exist in traces inside the catalyst. The Wheland intermediate is thus suggested to exist as an elusive intermediate; however, it remains an open question in how it is sensitive to UV−vis in such low amounts. The predicted lifetimes found in our study are moreover an upper bound for its actual concentration, as pure GGA functionals, like the revPBE-D3 employed here, tend to overstabilize charged species.[72,73] As final remark, it is also possible that other types of active sites, for instance Extra Framework ALuminum (EFAL) species, could significantly increase the BAS acidity through cooperative effects[74] and, with it, the concentration of Wheland intermediates.

4. CONCLUSIONS

In this work, we report for the first time a complete mechanistic investigation of the benzene ethylation reaction with both ethene and ethanol using enhanced sampling molecular dynamics techniques to capture realistic operating conditions. Moreover, we thoroughly investigated the effect of various water loadings on the formation of Wheland complexes in the catalyst. At the chosen reaction conditions (623 K, 1 atm) the ethylation reaction is shown to proceed with analogous rates both through the stepwise mechanism (with the intermediate formation of a surface ethoxide species) and the concerted mechanism. The latter is, however, made unlikely by the restrain that both benzene and the ethylating agents must find themselves in proximity of the active site at the same time. In both cases, a transient Wheland complex is formed as reaction intermediate, which is however rather short living.

To understand the role of water on the protonation kinetics, we considered the more favorable *para* protonation of the ethylbenzene product and we analyzed it in the presence of 0, 1, 3, and 6 water molecules per zeolite unit cell. We showed that water can actively act as proton transferring agent, lowering the activation energy for the protonation reaction and increasing the rate of about 1 order of magnitude when 1 molecule per BAS is considered. At higher coverages, the BAS is partially or fully solvated by the water cluster. Such solvation strongly stabilizes the reactant state and is associated with a decrease in the protonation rate, which becomes basically as low as the anhydrous case with 6 water molecules per unit cell. The insights presented here provide further mechanistic details on the role and effect of water as a proton-transfer agent in zeolite-catalyzed reactions.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacsau.1c00544.

Catalyst model, extended computational details on the enhanced sampling simulations, two-dimensional expansion of the free energy profiles, mobility analysis of the reacting substates, comparison with previous literature reports, extended details on the analysis of the simulations with extra water molecules (PDF)

Cartesian coordinates of the initial structures for all simulations; CP2K and PLUMED input file examples (ZIP)

AUTHOR INFORMATION

Corresponding Author

Veronique Van Speybroeck − *Center for Molecular Modeling, Ghent University, 9052 Zwijnaarde, Belgium;* orcid.org/0000-0003-2206-178X; Email: veronique.vanspeybroeck@ugent.be

Authors

Massimo Bocus − *Center for Molecular Modeling, Ghent University, 9052 Zwijnaarde, Belgium;* orcid.org/0000-0001-9474-6644

Louis Vanduyfhuys − *Center for Molecular Modeling, Ghent University, 9052 Zwijnaarde, Belgium;* orcid.org/0000-0001-6747-3388

Frank De Proft − *Eenheid Algemene Chemie (ALGC), Vrije Universiteit Brussel, 1050 Brussels, Belgium;* orcid.org/0000-0003-4900-7513

Bert M. Weckhuysen − *Inorganic Chemistry and Catalysis Group, Debye Institute for Nanomaterials Science, Utrecht University, 3584 CG Utrecht, The Netherlands;* orcid.org/0000-0001-5245-1426

Complete contact information is available at:
https://pubs.acs.org/10.1021/jacsau.1c00544

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge the Fund for Scientific Research - Flanders (FWO) as well as the Research Board of Ghent University (BOF). The computational resources and services used were provided by Ghent University (Stevin Supercomputer Infrastructure) and the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO).

REFERENCES

(1) Vogt, E. T. C.; Whiting, G. T.; Dutta Chowdhury, A.; Weckhuysen, B. M. Zeolites and Zeotypes for Oil and Gas Conversion.

Advances in Catalysis; Elsevier, Inc.: Amsterdam, 2015; Vol. *58*, pp 143−314.

(2) Busca, G. Acid Catalysts in Industrial Hydrocarbon Chemistry. *Chem. Rev.* **2007**, *107* (11), 5366−5410.

(3) Al-Khattaf, S.; Ali, S. A.; Aitani, A. M.; Žilková, N.; Kubička, D.; Čejka, J. Recent Advances in Reactions of Alkylbenzenes over Novel Zeolites: The Effects of Zeolite Structure and Morphology. *Catal. Rev. - Sci. Eng.* **2014**, *56* (4), 333−402.

(4) Venuto, P. B.; Hamilton, L. A.; Landis, P. S.; Wise, J. J. Organic Reactions Catalyzed by Crystalline Aluminosilicates. I. Alkylation Reactions. *J. Catal.* **1966**, *5* (1), 81−98.

(5) Sun, J.; Wang, Y. Recent Advances in Catalytic Conversion of Ethanol to Chemicals. *ACS Catal.* **2014**, *4* (4), 1078−1090.

(6) Schutyser, W.; Renders, T.; Van Den Bosch, S.; Koelewijn, S. F.; Beckham, G. T.; Sels, B. F. Chemicals from Lignin: An Interplay of Lignocellulose Fractionation, Depolymerisation, and Upgrading. *Chem. Soc. Rev.* **2018**, *47* (3), 852−908.

(7) Liao, Y.; Koelewijn, S. F.; van den Bossche, G.; van Aelst, J.; van den Bosch, S.; Renders, T.; Navare, K.; Nicolaï, T.; van Aelst, K.; Maesen, M.; et al. A Sustainable Wood Biorefinery for Low-Carbon Footprint Chemicals Production. *Science* **2020**, *367* (6484), 1385−1390.

(8) Vos, A. M.; Schoonheydt, R. A.; De Proft, F.; Geerlings, P. Reactivity Descriptors and Rate Constants for Acid Zeolite Catalyzed Ethylation and Isopropylation of Benzene. *J. Phys. Chem. B* **2003**, *107* (9), 2001−2008.

(9) Arstad, B.; Kolboe, S.; Swang, O. Theoretical Investigation of Arene Alkylation by Ethene and Propene over Acidic Zeolites. *J. Phys. Chem. B* **2004**, *108* (7), 2300−2308.

(10) Namuangruk, S.; Pantu, P.; Limtrakul, J. Alkylation of Benzene with Ethylene over Faujasite Zeolite Investigated by the ONIOM Method. *J. Catal.* **2004**, *225* (2), 523−530.

(11) Hansen, N.; Brüggemann, T.; Keil, F. J.; Bell, A. T. Theoretical Investigation of Benzene Alkylation with Ethene over H-ZSM-5. *J. Phys. Chem. C* **2008**, *112*, 15402−15411.

(12) Hansen, N.; Kerber, T.; Sauer, J.; Bell, A. T.; Keil, F. J. Quantum Chemical Modeling of Benzene Ethylation over H-ZSM-5 Approaching Chemical Accuracy: A Hybrid MP2:DFT Study. *J. Am. Chem. Soc.* **2010**, *132* (33), 11525−11538.

(13) Wang, D.; Wang, C. M.; Yang, G.; Du, Y. J.; Yang, W. M. First-Principles Kinetic Study on Benzene Alkylation with Ethanol vs. Ethylene in H-ZSM-5. *J. Catal.* **2019**, *374*, 1−11.

(14) Wang, D.; Wang, C. M.; Yang, W. M. Three-Dimensional Kinetic Trends in Zeolites Catalyzed Benzene Ethylation Reaction: A Descriptor-Based DFT Study Coupled with Microkinetic Modeling. *ACS Catal.* **2020**, *10* (3), 1652−1662.

(15) Acharya, D.; Chen, W.; Yuan, J.; Liu, Z.; Yi, X.; Xiao, Y.; Zheng, A. Stepwise or Concerted Mechanisms of Benzene Ethylation Catalyzed by Zeolites? Theoretical Analysis of Reaction Pathways. *Catal. Lett.* **2021**, *151*, 3048−3056.

(16) Galabov, B.; Nalbantova, D.; Schleyer, P. V. R.; Schaefer, H. F. Electrophilic Aromatic Substitution: New Insights into an Old Class of Reactions. *Acc. Chem. Res.* **2016**, *49* (6), 1191−1199.

(17) Fang, H.; Zheng, A.; Xu, J.; Li, S.; Chu, Y.; Chen, L.; Deng, F. Theoretical Investigation of the Effects of the Zeolite Framework on the Stability of Carbenium Ions. *J. Phys. Chem. C* **2011**, *115* (15), 7429−7439.

(18) Cnudde, P.; De Wispelaere, K.; Van Der Mynsbrugge, J.; Waroquier, M.; Van Speybroeck, V. Effect of Temperature and Branching on the Nature and Stability of Alkene Cracking Intermediates in H-ZSM-5. *J. Catal.* **2017**, *345*, 53−69.

(19) Margarit, V. J.; Osman, M.; Al-Khattaf, S.; Martínez, C.; Boronat, M.; Corma, A. Control of the Reaction Mechanism of Alkylaromatics Transalkylation by Means of Molecular Confinement Effects Associated to Zeolite Channel Architecture. *ACS Catal.* **2019**, *9*, 5935−5946.

(20) Li, C.; Ferri, P.; Paris, C.; Moliner, M.; Boronat, M.; Corma, A. Design and Synthesis of the Active Site Environment in Zeolite Catalysts for Selectively Manipulating Mechanistic Pathways. *J. Am. Chem. Soc.* **2021**, *143*, 10718−10726.

(21) Van Speybroeck, V.; De Wispelaere, K.; Van Der Mynsbrugge, J.; Vandichel, M.; Hemelsoet, K.; Waroquier, M. First Principle Chemical Kinetics in Zeolites: The Methanol-to-Olefin Process as a Case Study. *Chem. Soc. Rev.* **2014**, *43* (21), 7326−7357.

(22) Fecík, M.; Plessow, P. N.; Studt, F. A Systematic Study of Methylation from Benzene to Hexamethylbenzene in H-SSZ-13 Using Density Functional Theory and Ab Initio Calculations. *ACS Catal.* **2020**, *10* (15), 8916−8925.

(23) Chowdhury, A. D.; Houben, K.; Whiting, G. T.; Chung, S. H.; Baldus, M.; Weckhuysen, B. M. Electrophilic Aromatic Substitution over Zeolites Generates Wheland-Type Reaction Intermediates. *Nat. Catal.* **2018**, *1* (1), 23−31.

(24) Van Speybroeck, V.; Hemelsoet, K.; Joos, L.; Waroquier, M.; Bell, R. G.; Catlow, C. R. A. Advances in Theory and Their Application within the Field of Zeolite Chemistry. *Chem. Soc. Rev.* **2015**, *44* (20), 7044−7111.

(25) De Wispelaere, K.; Ensing, B.; Ghysels, A.; Meijer, E. J.; Van Speybroeck, V. Complex Reaction Environments and Competing Reaction Mechanisms in Zeolite Catalysis: Insights from Advanced Molecular Dynamics. *Chem.—Eur. J.* **2015**, *21* (26), 9385−9396.

(26) De Wispelaere, K.; Bailleul, S.; Van Speybroeck, V. Towards Molecular Control of Elementary Reactions in Zeolite Catalysis by Advanced Molecular Simulations Mimicking Operating Conditions. *Catal. Sci. Technol.* **2016**, *6* (8), 2686−2705.

(27) Collinge, G.; Yuk, S. F.; Nguyen, M. T.; Lee, M. S.; Glezakou, V. A.; Rousseau, R. Effect of Collective Dynamics and Anharmonicity on Entropy in Heterogenous Catalysis: Building the Case for Advanced Molecular Simulations. *ACS Catal.* **2020**, *10* (16), 9236−9260.

(28) Piccini, G.; Alessio, M.; Sauer, J. Ab-Initio Calculation of Rate Constants for Molecule-Surface Reactions with Chemical Accuracy. *Angew. Chem., Int. Ed.* **2016**, *55* (17), 5235−5237.

(29) De Wispelaere, K.; Vanduyfhuys, L.; Van Speybroeck, V. Entropy Contributions to Transition State Modeling. *Modelling and Simulation in the Science of Micro- and Meso-Porous Materials*; Elsevier, Inc.: Amsterdam, 2018; pp 189−228.

(30) Cnudde, P.; De Wispelaere, K.; Vanduyfhuys, L.; Demuynck, R.; Van der Mynsbrugge, J.; Waroquier, M.; Van Speybroeck, V. How Chain Length and Branching Influence the Alkene Cracking Reactivity on H-ZSM-5. *ACS Catal.* **2018**, *8*, 9579−9595.

(31) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. *J. Phys. Chem. B* **2006**, *110* (8), 3533−3539.

(32) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100* (2), 020603.

(33) Bhan, A.; Joshi, Y. V.; Delgass, W. N.; Thomson, K. T. DFT Investigation of Alkoxide Formation from Olefins in H-ZSM-5. *J. Phys. Chem. B* **2003**, *107* (38), 10476−10487.

(34) Van Der Mynsbrugge, J.; Hemelsoet, K.; Vandichel, M.; Waroquier, M.; Van Speybroeck, V. Efficient Approach for the Computational Study of Alcohol and Nitrile Adsorption in H-ZSM-5. *J. Phys. Chem. C* **2012**, *116* (9), 5499−5508.

(35) Verstraelen, T.; Van Speybroeck, V.; Waroquier, M. ZEOBUILDER: A GUI Toolkit for the Construction of Complex Molecular Structures on the Nanoscale with Building Blocks. *J. Chem. Inf. Model.* **2008**, *48*, 1530−1541.

(36) Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schütt, O.; Schiffmann, F.; et al. CP2K: An Electronic Structure and Molecular Dynamics Software Package -Quickstep: Efficient and Accurate Electronic Structure Calculations. *J. Chem. Phys.* **2020**, *152*, 194103.

(37) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77* (18), 3865−3868.

(38) Zhang, Y.; Yang, W. Comment on "Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1998**, *80* (4), 890.

(39) Yang, K.; Zheng, J.; Zhao, Y.; Truhlar, D. G. Tests of the RPBE, RevPBE, τ -HCTHhyb, Ωb97X-D, and MOHLYP Density Functional Approximations and 29 Others against Representative Databases for Diverse Bond Energies and Barrier Heights in Catalysis. *J. Chem. Phys.* **2010**, *132* (16), 1−10.

(40) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(41) Lippert, G.; Hutter, J.; Parrinello, M. A Hybrid Gaussian and Plane Wave Density Functional Scheme. *Mol. Phys.* **1997**, *92* (3), 477−487.

(42) Lippert, G.; Hutter, J.; Parrinello, M. The Gaussian and Augmented-Plane-Wave Density Functional Method for Ab Initio Molecular Dynamics Simulations. *Theor. Chem. Acc.* **1999**, *103*, 124−140.

(43) Goedecker, S.; Teter, M.; Hutter, J. Separable Dual-Space Gaussian Pseudopotentials. *Phys. Rev. B - Condens. Matter Mater. Phys.* **1996**, *54* (3), 1703−1710.

(44) Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52* (2), 255−268.

(45) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé-Hoover Chains: The Canonical Ensemble via Continuous Dynamics. *J. Chem. Phys.* **1992**, *97* (4), 2635−2643.

(46) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101* (5), 4177−4189.

(47) Pérez De Alba Ortíz, A.; Tiwari, A.; Puthenkalathil, R. C.; Ensing, B. Advances in Enhanced Sampling along Adaptive Paths of Collective Variables. *J. Chem. Phys.* **2018**, *149* (7), 072320.

(48) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604−613.

(49) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (20), 12562−12566.

(50) Laio, A.; Gervasio, F. L. Metadynamics: A Method to Simulate Rare Events and Reconstruct the Free Energy in Biophysics, Chemistry and Material Science. *Rep. Prog. Phys.* **2008**, *71*, 126601.

(51) Torrie, G. M.; Valleau, J. P. Monte Carlo Free Energy Estimates Using Non-Boltzmann Sampling: Application to the Sub-Critical Lennard-Jones Fluid. *Chem. Phys. Lett.* **1974**, *28* (4), 578−581.

(52) Torrie, G. M.; Valleau, J. P. Monte Carlo Study of a Phase-Separating Liquid Mixture by Umbrella Sampling. *J. Chem. Phys.* **1977**, *66* (4), 1402−1408.

(53) Bailleul, S.; Dedecker, K.; Cnudde, P.; Vanduyfhuys, L.; Waroquier, M.; Van Speybroeck, V. Ab Initio Enhanced Sampling Kinetic Study on MTO Ethene Methylation Reaction. *J. Catal.* **2020**, *388*, 38−51.

(54) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The Weighted Histogram Analysis Method for Free-energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13* (8), 1011−1021.

(55) Souaille, M.; Roux, B. Extension to the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations. *Comput. Phys. Commun.* **2001**, *135* (1), 40−57.

(56) ThermoLIB | Center for Molecular Modeling. https://molmod.ugent.be/software/thermolib (accessed 2021-11-19).

(57) Van Den Broeck, E.; Verbraeken, B.; Dedecker, K.; Cnudde, P.; Vanduyfhuys, L.; Verstraelen, T.; Van Hecke, K.; Jerca, V. V.; Catak, S.; Hoogenboom, R.; et al. Cation-Π Interactions Accelerate the Living Cationic Ring-Opening Polymerization of Unsaturated 2-Alkyl-2-Oxazolines. *Macromolecules* **2020**, *53* (10), 3832−3846.

(58) Ensing, B.; Laio, A.; Parrinello, M.; Klein, M. L. A Recipe for the Computation of the Free Energy Barrier and the Lowest Free Energy Path of Concerted Reactions. *J. Phys. Chem. B* **2005**, *109* (14), 6676−6687.

(59) Bailleul, S.; Rogge, S. M. J.; Vanduyfhuys, L.; Van Speybroeck, V. Insight into the Role of Water on the Methylation of Hexamethylbenzene in H-SAPO-34 from First Principle Molecular Dynamics Simulations. *ChemCatChem.* **2019**, *11* (16), 3993−4010.

(60) Van Der Mynsbrugge, J.; Moors, S. L. C.; De Wispelaere, K.; Van Speybroeck, V. Insight into the Formation and Reactivity of Framework-Bound Methoxide Species in h-Zsm-5 from Static and Dynamic Molecular Simulations. *ChemCatChem.* **2014**, *6* (7), 1906−1918.

(61) Nastase, S. A. F.; Cnudde, P.; Vanduyfhuys, L.; De Wispelaere, K.; Van Speybroeck, V.; Catlow, C. R. A.; Logsdail, A. J. Mechanistic Insight into the Framework Methylation of H-ZSM-5 for Varying Methanol Loadings and Si/Al Ratios Using First-Principles Molecular Dynamics Simulations. *ACS Catal.* **2020**, *10* (15), 8904−8915.

(62) De Wispelaere, K.; Martínez-Espín, J. S.; Hoffmann, M. J.; Svelle, S.; Olsbye, U.; Bligaard, T. Understanding Zeolite-Catalyzed Benzene Methylation Reactions by Methanol and Dimethyl Ether at Operating Conditions from First Principle Microkinetic Modeling and Experiments. *Catal. Today* **2018**, *312*, 35−43.

(63) Weissermel, K.; Arpe, H.-J. *Industrial Organic Chemistry*, 4th ed.; Wiley-VCH: Weinheim, Germany, 2003.

(64) Chizallet, C. Toward the Atomic Scale Simulation of Intricate Acidic Aluminosilicate Catalysts. *ACS Catal.* **2020**, *10* (10), 5579−5601.

(65) Liu, P.; Mei, D. Identifying Free Energy Landscapes of Proton-Transfer Processes between Brønsted Acid Sites and Water Clusters Inside the Zeolite Pores. *J. Phys. Chem. C* **2020**, *124* (41), 22568−22576.

(66) Grifoni, E.; Piccini, G. M.; Lercher, J. A.; Glezakou, V. A.; Rousseau, R.; Parrinello, M. Confinement Effects and Acid Strength in Zeolites. *Nat. Commun.* **2021**, *12*, 2630.

(67) Bocus, M.; Neale, S. E.; Cnudde, P.; Van Speybroeck, V. Dynamic Evolution of Catalytic Active Sites within Zeolite Catalysis. In *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*; Elsevier, Inc.: Amsterdam, 2021.

(68) Eckstein, S.; Hintermeier, P. H.; Zhao, R.; Baráth, E.; Shi, H.; Liu, Y.; Lercher, J. A. Influence of Hydronium Ions in Zeolites on Sorption. *Angew. Chem., Int. Ed.* **2019**, *58* (11), 3450−3455.

(69) Grifoni, E.; Piccini, G. M.; Parrinello, M. Microscopic Description of Acid-Base Equilibrium. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116* (10), 4054−4057.

(70) Chen, K.; Gumidyala, A.; Abdolrhamani, M.; Villines, C.; Crossley, S.; White, J. L. Trace Water Amounts Can Increase Benzene H/D Exchange Rates in an Acidic Zeolite. *J. Catal.* **2017**, *351*, 130−135.

(71) Vjunov, A.; Wang, M.; Govind, N.; Huthwelker, T.; Shi, H.; Mei, D.; Fulton, J. L.; Lercher, J. A. Tracking the Chemical Transformations at the Brønsted Acid Site upon Water-Induced Deprotonation in a Zeolite Pore. *Chem. Mater.* **2017**, *29* (21), 9030−9042.

(72) Sauer, J. Ab Initio Calculations for Molecule-Surface Interactions with Chemical Accuracy. *Acc. Chem. Res.* **2019**, *52* (12), 3502−3510.

(73) Goncalves, T. J.; Plessow, P. N.; Studt, F. On the Accuracy of Density Functional Theory in Zeolite Catalysis. *ChemCatChem.* **2019**, *11* (17), 4368−4376.

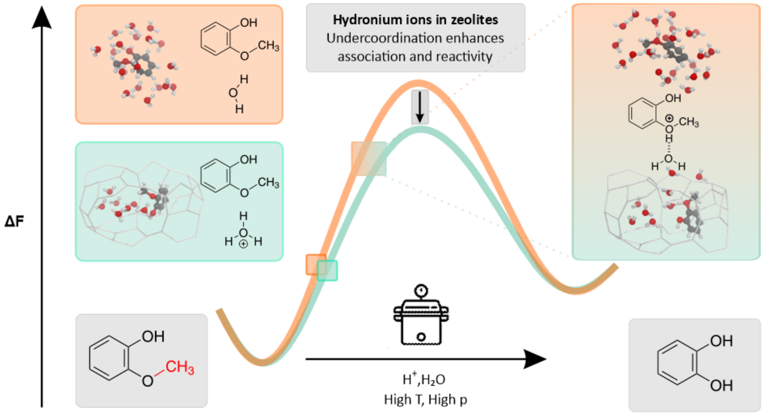
(74) Li, G.; Pidko, E. A. The Nature and Catalytic Function of Cation Sites in Zeolites: A Computational Perspective. *ChemCatChem.* **2019**, *11* (1), 134−156.

Paper IV

Undercoordinated confined water in Brønsted acidic zeolites speeds up the *O*-activated demethylation of guaiacol in hot-pressurized water

Elias Van den Broeck[‡], Massimo Bocus[‡], Xian Wu[‡], Mathias Bal[‡],
Jeroen Bomon, Louis Vanduyfhuys, Bert F. Sels*, Bert U. W. Maes*
and Veronique Van Speybroeck*

Submitted



---

*Corresponding authors

‡Authors contributed equally

M. Bocus performed the simulations of the zeolite-catalyzed reaction and prepared the initial draft of the manuscript together with E. Van Den Broeck.

Undercoordinated confined water in Brønsted acidic zeolites speeds up the *O*-activated demethylation of guaiacol in hot-pressurized water

Elias Van Den Broeck[‡,a], Massimo Bocus[‡,a], Xian Wu[‡,b], Mathias Bal[‡,c], Jeroen Bomon[c], Louis Vanduyfhuys[a], Bert F. Sels[*,b], Bert U. W. Maes[*,c], Veronique Van Speybroeck[*,a]

[a] Center for Molecular Modeling, Ghent University,Technologiepark 46, 9052 Zwijnaarde, Belgium

[b] Center for Sustainable Catalysis and Engineering, KU Leuven, Celestijnenlaan 200F, 3001 Leuven, Belgium

[c] Organic Synthesis Division, Department of Chemistry, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium

‡ *authors contributed equally*

Corresponding authors

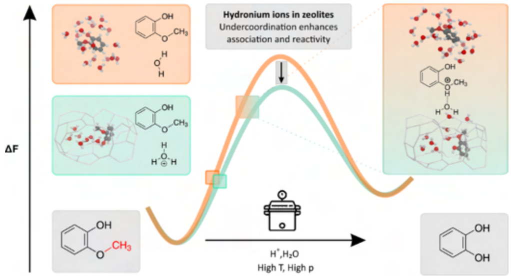
* Veronique Van Speybroeck, e-mail: Veronique.vanspeybroeck@ugent.be

* Bert Sels, e-mail: bert.sels@kuleuven.be

* Bert Maes, e-mail: bert.maes@uantwerpen.be

Abstract

New biorefinery technologies are paramount to more efficient biomass processing such as the emerging lignin conversion into platform chemicals. Herein, we study the Brønsted acid catalysed *O*-demethylation of guaiacol in hot-pressurized water in presence of soluble (HCl) and microporous solid (Beta zeolite) acids as a prototype model reaction with lignin-derived aromatic substrates. Operando molecular modelling, supported by experimental data, unravels the mechanism and reveals a great effect of the molecular environment on the catalytic activity. Irrespective of the acid catalyst, the reaction follows a similar, though rather surprising concerted, one-step *O*-activated demethylation reaction according to an S_N2 mechanism involving hydrogen bonding of guaiacol with the hydronium ion. Strong confinement effects within the micropores of the zeolite benefit the association between the hydronium ion and guaiacol, due to the higher activity of the low-coordinated hydronium ions. These proof-of-concept results show how acid-catalysed bimolecular reactions can be affected by the solvent structuring in subcritical water.



Introduction

With the sixth major climate assessment report by the IPCC published, it is crystal clear that effective mitigation strategies for climate change are vital to save the planet.[1] Hence, a paradigm shift is required to move from a ‘take-make-consume-dispose’ to a ‘take-make-consume-recycle’ flow of materials, i.e. switching to circular processes, that inherently implies a change from fossil carbon resources to renewable (e.g., bio-based) carbon.[2,3] Abundantly available lignocellulosic biomass, consisting of lignin and (hemi)cellulose, shows the potential of such sustainable feedstock, provided biorefineries can efficiently upgrade plant material, i.e. both the lignin and carbohydrate fractions, to viable product streams.[4–6] To this end, various lignin conversion technologies are under development to investigate full exploitation and utilization of the lignocellulosic biomass feedstock.[6–8]

Lignin, the largest renewable source of aromatics, consists of *p*-coumaryl, coniferyl and sinapyl alcohol units that are linked through *C-O* and *C-C* interlinkages.[8–12] Selective cleavage in the biorefinery can depolymerise lignin, providing lignin oils rich in for instance guaiacols. *O*-demethylation of guaiacol is an elegant way to form catechol (**Figure 1A**) with large market potential.[13] Indeed, catechol is known for a variety of applications such as antioxidant, anticorrosion or chelating agent, as detergent, or as precursor for fine chemicals (e.g. pesticides, perfumes and pharmaceuticals) and new (biodegradable) polymeric materials.[6,12–19]

A variety of synthetic procedures have been reported for the *O*-demethylation of guaiacol (and anisoles in general). However, they typically require harsh reaction conditions and (super)stoichiometric amount(s) of hazardous, non-sustainable reagent(s) in undesired solvents, producing problematic by-products.[20,21] A sustainable and scalable *O*-demethylation reaction, should be selective and high yielding, and preferably performed in a non-toxic solvent.[4,21] To this end, recent studies have shown the potential of Brønsted acid catalysis for *O*-demethylation in hot-pressurized water.[21–25] At high temperatures, water behaves similarly to an organic solvent, due to its lower dielectric constant, offering great opportunities in the context of green organic chemistry. Recently, we showed for instance that guaiacol in 20 mol% HCl in subcritical water yields 97% catechol.[21] The main drawback of the process lies in the high corrosivity of aqueous homogeneous Brønsted acids at high temperature, requiring Hastelloy reactor materials with an increased CapEx.

Solid acids in water provide a non-corrosive alternative to mineral acids, besides the classical technological advantages of hetero- *vs* homogeneous catalysts. Among them, acidic zeolites - crystalline microporous aluminosilicates - show exceptional catalytic activity, selectivity and stability in traditional fossil fuels conversion and may also become the workhorse of the emerging biorefinery industry, albeit modified to ensure their stability in more polar circumstances. [4,26,27] There are currently more than a hundred synthesized zeolites with different crystal structure, pore diameter and acidic properties. Among them, we recently identified commercial Beta-zeolites (BEA topology) as a suitable acid catalyst for the *O*-demethylation of 4-propylguaiacol yielding 90% 4-propylcatechol in hot-pressurized water, *en route* to catechol after subsequent propyl cleavage.[27]

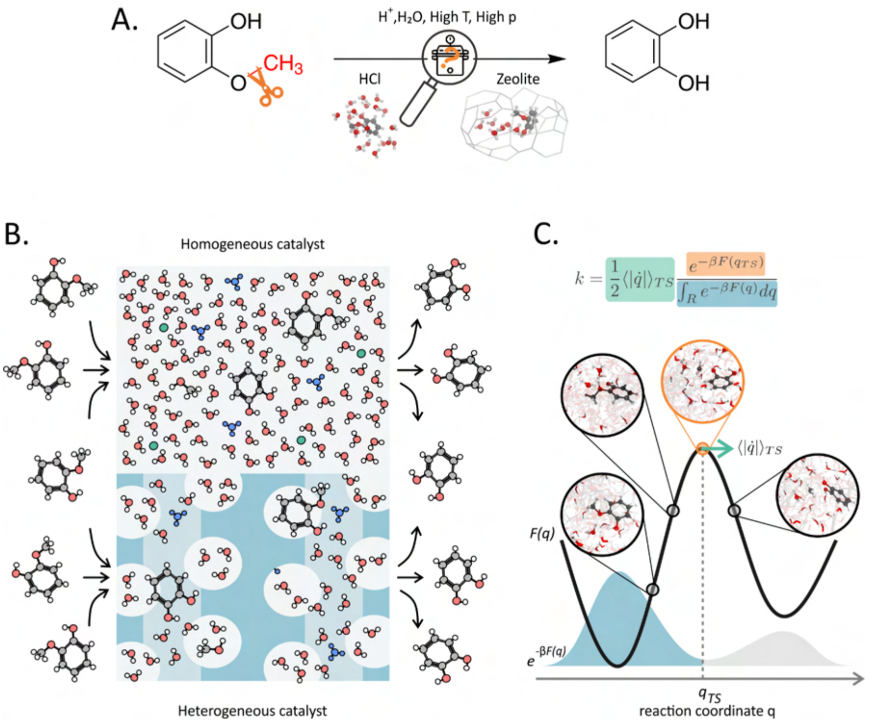


Figure 1. Enhanced molecular dynamics simulations are used to investigate the homo- and heterogeneous catalysed *O*-demethylation of guaiacol. A.) General reaction scheme for the Brønsted acid catalysed *O*-demethylation of guaiacol with formation of catechol. B.) Schematic representation of the investigated systems: the homogeneous Brønsted acid catalysed conversion (upper side) and the heterogeneous Brønsted acid catalysed conversion (bottom side), i.e. HCl versus zeolites, both in subcritical water. C.) Enhanced molecular dynamics simulations are combined with classical transition state theory to investigate the *O*-demethylation of guaiacol, properly accounting for the operating conditions and entropic contributions, snapshots for the *O*-demethylation of guaiacol are included.

The presence of water remarkably affects the catalytic properties of protons in zeolites, as demonstrated in the case of monomolecular dehydration reactions.[28–30] Lercher and co-workers have shown that protons are more active for cyclohexanol dehydration in intrazeolitic water than in bulk solution.[28] The experimental evidence suggests the creation of an ionic environment in the micropores by small protonated water clusters and the negatively charged zeolite framework, that are present in proximity of the defect site created by Al substitution. This affects the chemical potential of chemisorbed species and transition states facilitating reactions between charged reaction partners.[28,31–33] Despite the experimental evidence, atomistic insights regarding the activity and behavioural differences between such hydronium ions in bulk and confined water environments are yet to be explored. The ability to tune the molecular environment opens up a large potential to steer the selectivity, activity for other types of acid-catalysed reaction in the presence of water.

Therefore, we present a combined computational and experimental study on the acid-catalysed *O*-demethylation of guaiacol into catechol as model for a bimolecular reaction with water as reaction partner, with particular emphasis on the differences between the homogeneous (HCl) and heterogeneous (Beta-zeolite) catalysis (**Figure 1B**). After establishing the acid-catalysed mechanism of the catalytic cycle, we aim to provide a molecular understanding on how the confinement effects influence on the hydronium ion activity for this particular reaction. To effectively simulate the dynamic state of protonated water in the complex molecular environment – either in bulk solvent or confined space – enhanced sampling molecular dynamics (MD) simulations at operating conditions were carried out to properly account for thermal effects (**Figure 1C**).[34] Experimental kinetic data provided access to reaction orders, rates, activation energies and kinetic isotope effects (KIEs) for both the homogeneous and heterogeneous catalysis. Microkinetic modelling reconciles the experimental and modelling results, providing complete mechanistic and kinetic details for the reaction under study. In essence, this work elucidates a different solvation state of the hydronium ions depending on its molecular surrounding, either present in fully coordinated large water clusters in bulk or part of undercoordinated surface-like water structures which are consequence of the zeolite pore confinement. This explains the large difference in proton activity for the otherwise similar catalytic *O*-demethylation pathway in bulk and in zeolite pores. While the reaction can be seen as a prototype case study for acid-catalysed reactions with water as reaction partner in subcritical water conditions, the findings are of utmost importance for the development of other biomass-related reactions, usually processed in aqueous conditions.

Results and discussion

Structural characterization of the proton state in a bulk and confined reaction environment

Prior to an in-depth discussion of the governing reaction mechanism, the behaviour and activity of the hydronium ion at reaction conditions were investigated for both the homogeneous and heterogeneous acid catalysed systems. From a modelling point of view, both systems were set-up to mimic as close as possible the experimental conditions, as schematically illustrated in **Figure 2A** and **B.** To account for the true reaction temperature and the mobility of reactive species at these conditions, fully *ab initio* MD simulations were performed. More details about the system preparation and the technical details can be found in the *Methods Section* and the *Supplementary information Section **S1.1***.

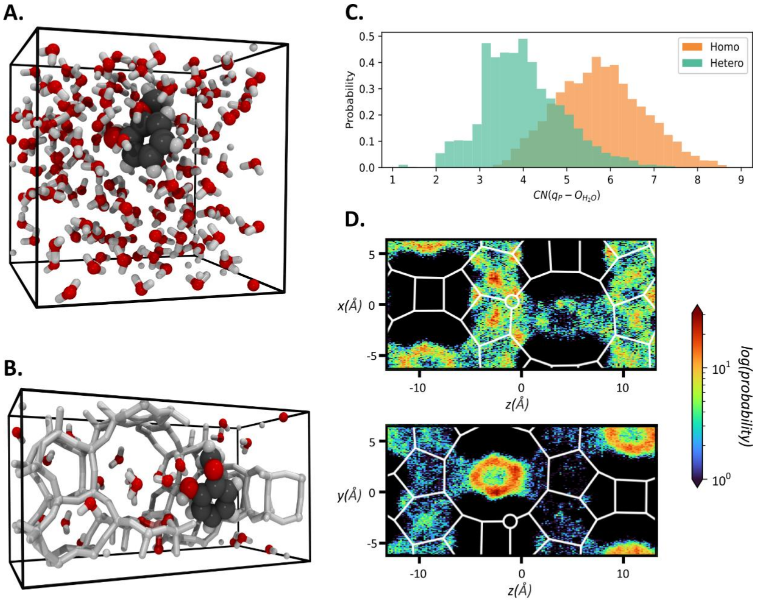


Figure 2. The hydronium ions reside in an undercoordinated surface-like state in zeolites compared to the homogeneous bulk phase. A) The unit cell of the homogeneous system containing 170 water molecules, 1 proton and 1 guaiacol molecule; B) The beta zeolite unit cell of the heterogeneous system containing 22 water molecules, 1 Al substitution (Si/Al=63) and 1 guaiacol molecule; C) Probability distribution of the coordination number (r_0=3.6, n=16, m=28) between the protonated-site q_P (based on ref. 40) and the water molecules, only Eigen-like structures are accounted for (see Supplementary information S2.1.1); D) Probability for O_{H_2O} within the Beta zeolite framework (white) as seen along the y (top) and x (bottom) axis (circle represents the Al defect location). (Level of theory: BLYP-D3, NVT, see Methods Section).

At first instance, we analysed the proton location from the MD based trajectories using the definition presented in the work of Ensing *et al.*[35] The hydronium ion is shown to be in an Eigen-like state ([$H_3O^+(H_2O)_n$]) around 93% of the time for both the homogeneous and heterogeneous catalytic systems, while the remaining ~7% is in a Zundel-like state ([H_2O••H••OH_2]$^+$, see *Supplementary Information* ***S2.1.1***, **Figure S8B**). In the zeolite, the Brønsted acid site (BAS) is readily solvated, forming a hydronium ion at the beginning of the simulation, in agreement with previous literature reports.[29] By tracking its location, it was found relatively mobile along the channel containing the Al substitution, whereas diffusion in other regions of the unit cell is difficult (see **Figure S8A** in the *Supplementary Information*). When analysing the coordination number (CN, see *Methods section and Supplementary information* ***S.1.1.3.1***) between the protonated site (q_p) with respect to nearby H_2O-oxygen atoms (O_{H_2O}), $CN(q_P - O_{H_2O})$, undercoordination of the hydronium ion is observed in the zeolite w.r.t. the bulk case (**Figure 2C**). While water molecules in bulk are fully solvated, the molecular-sized channels in the zeolite constrain the water molecules into an undercoordinated surface-like state (see **Figure S7**).[36] This is caused by the formation of tube-like water structures within the channels of the zeolite, as revealed by the probability distribution of O_{H_2O} within the framework, see **Figure 2D** (and confirmed by the difference in $O_{H_2O} - H_{H_2O}$ and $H_{H_2O} - H_{H_2O}$ radial distribution functions (rdfs), see *Supplementary Information Section* ***S2.1.1***). A similar phenomenon has been reported by Bukowski *et al.*, albeit in the case of Lewis acid Sn-Beta when studying the impact of defects on the water structure in the material.[37]

From a reactivity point of view, this structural analysis pinpoints an undercoordinated and, therefore, possibly less stabilized hydronium ion in the confined environment. Because of such a reduced solvation sphere, the hydronium ion electrophilicity is enhanced, leading to an increased activity towards proton-transferring steps. This is also supported by the average Mulliken's charges computed on 100 random snapshots extracted from the MD trajectories of both systems with the hydronium ion in an Eigen-like state. We found that the average charge on the hydronium ion goes from -0.379 to -0.357 in the homo- and heterogeneous case, respectively, in line with an increased electrophilicity. Moreover, the water cluster as a whole (*i.e.* all water molecules in the cell excluding the H_3O^+ moiety) presents an increased average charge in the zeolite (-0.534/H_2O) compared to bulk (-0.590/H_2O), which can easily be attributed to the lower coordination and, therefore, to a lower charge delocalization.

Reaction mechanism of guaiacol O-demethylation

A full mechanistic characterization of the guaiacol *O*-demethylation reaction was performed by means of enhanced sampling MD simulations (see *Methods* section). Two distinct catalytic routes have been proposed in literature for the Brønsted acid-catalysed *O*-demethylation of guaiacol towards catechol: an *O*- and an arene-activated route (**Figure 3A**), with the former being the generally accepted mechanism.[21–25] The two-step *O*-activated demethylation through an S_N2 mechanism (*O*2-S_N2) is assumed to proceed via the formation of a stable oxonium intermediate (**VI**) that subsequently undergoes an S_N2 reaction with water, providing methanol as by-product. The stepwise arene-activated *O*-demethylation involves a *C*-protonated intermediate (**II**) and was recently proposed by us as a potential competing path.[25] After protonation, the reaction can proceed again with an S_N2 mechanism towards intermediate **IV**, i.e. a two-step arene-activated *O*-demethylation through an S_N2 mechanism (A2-S_N2), or reach it after a further step with formation of a hemiacetal intermediate **III**, i.e. the hemiacetal path. **IV** can then be converted in the final catechol through a tautomerization step via intermediate **VII**. For a full overview of the computational results of each reaction step, the interested reader is referred to the *Supplementary Information Section **S2.1.2***. All steps involving methanol formation are assumed to be irreversible (i.e. $\mathrm{k}_{-2} = \mathrm{k}_{-4} = \mathrm{k}_{-7} = 0$) suggested by the much higher concentration of water (solvent) and the higher volatility of methanol disfavouring the backward reactions.

When investigating the *O*2-S_N2 path, we never observed the formation of a (meta)stable oxonium species **VI** (*vide infra*), shaded grey in **Figure 3A**. Surprisingly, a concerted mechanism is found instead, with a one-step *O*-activated demethylation through an S_N2 mechanism (*O*1-S_N2) involving a strong hydrogen bond with the hydronium ion, in contrast with previous mechanistic studies.[21–25] The calculated reaction rates and corresponding phenomenological barriers for each elementary step of the reaction scheme depicted in **Figure 3A** are summarized in **Table 1**. A graphical depiction of the phenomenological barriers can also be seen in **Figure 3C**.

Based on the assumption that each elementary step follows first-order kinetics w.r.t. each of the reactants, the set of differential equations S1.1-S1.6 (*Section **S2.1.2***) was solved to obtain a concentration *versus* time plot for the homogeneous case (**Figure 3B**). Such an assumption is in line with the experimentally observed first-order kinetics for the Brønsted acid in both the homo- and heterogeneous catalysis (*Section **S2.2.2.2*** and ***S2.3.1.1***). Interestingly near zero-order kinetics for guaiacol were observed in the heterogeneous case (*Section **S2.3.1.1**),* whereas a first-order was noticed in the homogeneous case (*Section **S2.2.2.1***).[28] This is in line with a strong sorption of guaiacol in water by zeolites, leading to an intra-pore concentration that is independent of the concentration in solution, as seen before for cyclohexanol.[28] Remark that, given its irrelevance for the reaction mechanism, the final conversion from **VI** to **V** was not investigated computationally in the heterogeneous case.

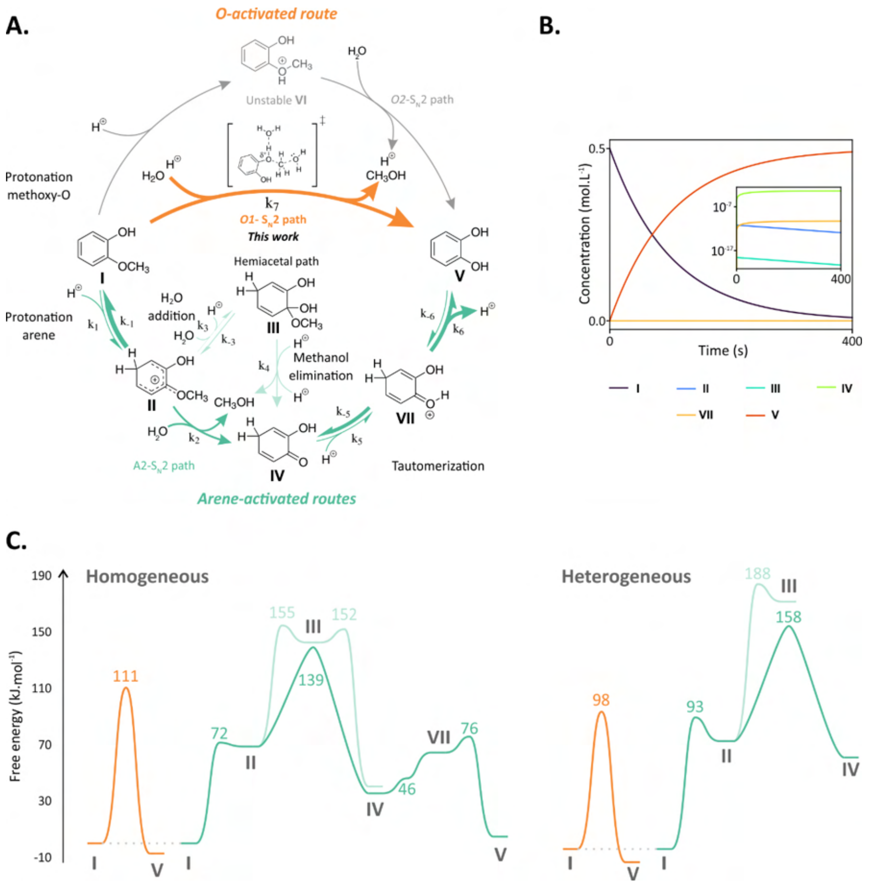


Figure 3. The O1-S$_N$2 path dominates the conversion of guaiacol to catechol. A) Complete mechanistic overview of the reaction paths considered in this work. The arrows' thickness is proportional to the kinetic constant of each reaction step. H^+ refers to the solvated hydronium ions [$H_3O^+(H_2O)_n$]; B) Time evolution of the intermediates concentrations based on the theoretical kinetic constants for the homogeneous case (with the irreversible-step approximation, see Equations S1.1-S1.6, Section **S2.1.2**); C) Complete free energy profiles of the homo- (*left*) and heterogeneously (*right*) catalysed *O*-demethylation of guaiacol, based on the phenomenological barriers summarized in **Table 1**. Note that, because of the uncertainty on the computed free energy differences, the alignment between states is not necessarily perfect, as in the case of **IV** and **V** for the homogeneous case (Level of theory: BLYP-D3, NVT, see Methods Section).

As it can be seen in **Figure 3B**, the concentrations of intermediates **II** to **VII** remain very low during the reaction. Despite the lower activation energy of the arene-activated S$_N$2 with respect to the *O*-activated one, the unfavourable equilibrium between **I** and **II** ($K_{1,Homo} = 1.4 \times 10^{-7}$and $K_{1,Hetero} = 2.3 \times 10^{-8}$) leads to a high energy for the arene-activated paths. The guaiacol conversion proceeds relatively fast (see black curve in

Figure 3B) through the *O*1-S$_N$2 path, with a phenomenological barrier of 110.6 and 97.5 kJ·mol^{-1} for the homo- and heterogeneous case, respectively (**Figure 3A** and **C**).

Calculations therefore suggest a clear enhancement of the reaction rate (per proton) when going from the homo- to the heterogeneous case. In particular, the 13.6 kJ·mol^{-1} decrease in the phenomenological barrier (for StepID 7, **Table 1**) corresponds to an increase by a factor of 20 in the reaction rate This observation is in qualitative agreement with the experimental results, where an increase in turn-over frequency (TOF) by one order of magnitude (523K, see *Section **S2.3.1.4***) is noted for the heterogeneously catalysed system with respect to the homogeneous one. Experimental Arrhenius activation energies of 147 ± 8 kJ·mol^{-1} and 113 ± 9 kJ·mol^{-1} were obtained for the homo- and heterogeneous case respectively (**Figure 4A**), suggesting an enthalpic enhancement of the rate in the zeolite. These results are in line with earlier reports of an increased hydronium ion activity in constrained environments,[28] but are here shown for the first time in the case of a bimolecular reaction with water as reaction partner.

Prior to proceed in the results discussion, we first analysed the impact of the water loading in the zeolite unit cell on the computed kinetics by recomputing the *O*1-S$_N$2 path with 17 and 27 molecules in the unit cell (that is a variation of ± 5 w.r.t. the original 22, see *Supplementary Information **S2.1.3***). This is because a reliable estimate of water loadings in the zeolite pores is extremely challenging to obtain. The results indicate that an increase in water loading enhances the reaction kinetics. For instance, phenomenological barriers of 111.1, 97.5 and 90.7 kJ·mol^{-1} were found for 17, 22 and 27 molecules per unit cell, respectively. This increase in activity is attributed to entropic effects. Since water acts both as solvent and reagent, a larger loading makes it likelier for the *O*-methyl group to find itself in proximity of a nucleophilic water molecule during the simulation (**Figure S33**).

Although the *A*2-S$_N$2 path is negligible for the formation of catechol, it is worth noting that the S$_N$2 mechanism (StepID 2) dominates over the previously reported hemiacetal mechanism involving **III** (StepID 3), in contrast with previous reports relying on static calculation with a limited amount of solvent molecules.[25] Therefore, accounting for an explicit solvent environment is of paramount importance to capture the entropic effects and thus describe the reaction kinetics and the corresponding mechanism as correctly as possible. It can also be noticed that the *C*(sp^2)-(de)protonation of guaiacol (**I** - **II**) has a lower or comparable activation energy as the *O*1-S$_N$2 path. Despite not leading to catechol formation, the calculations suggest that the aromatic ring quickly exchanges protons with the environment, which is in line with the experimental observation for the reaction conducted in heavy water (*vide infra* and **Section 2.2.2.4** of the *Supplementary Information*).

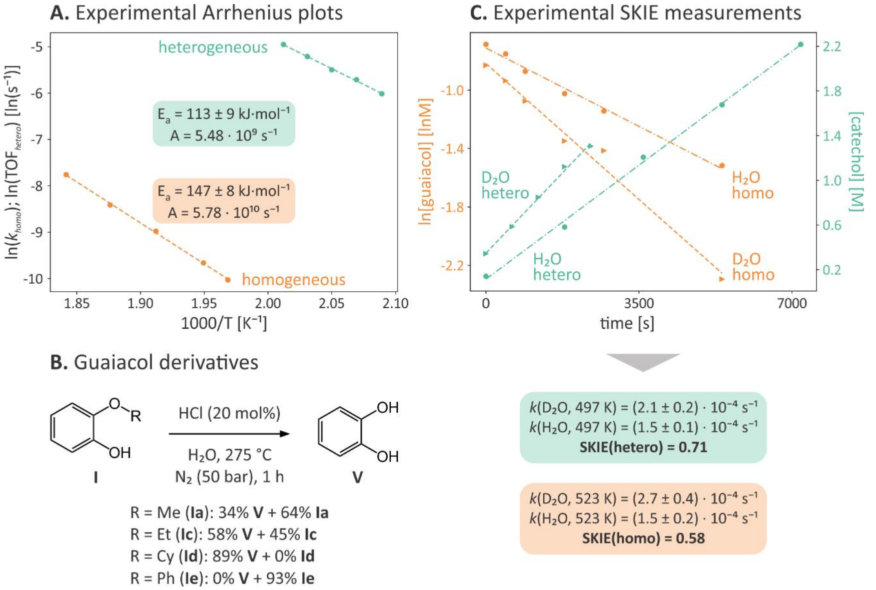


Figure 4. Experimental evidences support a faster *O1*-S_N2 mechanism in zeolite. A) Experimental Arrhenius plots for the homogeneously (orange, HCl) and heterogeneously (green, Beta-zeolite) catalysed *O*-demethylation of guaiacol in hot-pressurized water; B) Yields for the *O*-demethylation of guaiacol derivatives featuring different alkyl chains; C) Experimental reaction progress plots in H_2O and D_2O to determine the solvent kinetic isotope effect. Note that for the heterogeneous case reaction is zero-order in guaiacol (*Section S3.3.1.1* of the *Supplementary Information*).

Finally, additional catalytic experiments were performed in the homogeneous case with various guaiacol derivatives featuring a variety of *O*-alkyl chains, to provide further support to an S_N2-type mechanism (for a complete overview of the results the interested reader is referred to the *Supplementary information*, *Section* **S2.2.1.2**). It was found that increasing the number of substituents on the reactive *O*-carbon, *i.e.* going from a methyl to an ethyl or cyclohexyl group, leads on the one hand to an increase in the reaction rate (**Figure 4B**). While seemingly counterintuitive, we attribute this effect to the increased nucleophilicity of the leaving guaiacol oxygen, which may improve its association with the hydronium ions in solution (*vide infra*). On the other hand, with a phenyl group that cannot undergo the S_N-type reaction, no dealkylation reaction is observed, providing additional support to the *O*1-S_N2 path proposed from the calculation.

Unique protonation states in the O1-S_N2 mechanism

Solvent kinetic isotope effect (SKIE) was studied to get further insight in the activating effect of hydronium on the *O*1-S_N2 pathway. This comprises effects on the rate constant by replacing the water solvent with its deuterated counterpart. Reactions in which the solvent acts both as solvent and reactant, as in our case of guaiacol *O*-demethylation, are prone to both primary and secondary kinetic isotope effects (KIEs) and hence the observed SKIE values include the net effect of both:

$$\left(\frac{k_H}{k_D}\right)_{obs} = \left(\frac{k_H}{k_D}\right)_{sec}\left(\frac{k_H}{k_D}\right)_{pri}$$

Experimentally, the rate constants were determined both in deuterated water ($k_{D,obs}$) and non-deuterated water ($k_{H,obs}$) for both homogeneous and zeolite catalysis. The results are shown in **Figure 4C** (see also *Supplementary Information*, *Section **S2.2.2.4*** and ***S2.3.1.3*** for a complete results overview). Interestingly, a similar reverse SKIE of 0.58 and 0.71 was measured in both the homo- and heterogeneous systems. In order to understand this unexpected observation due to solvent isotope labelling, we analysed theoretically the proton-transfer progress along the *O*1-S_N2 reaction. To this end, both primary and secondary KIE were rationalized by reconstructing the free energy surfaces along new collective variables, to be able to monitor the protonation states of relevant oxygen atoms during the *O*1-S_N2 reaction (more details on the procedure are reported in the *Supplementary Information*, *Section **S2.1***). The results for both catalytic systems are shown in **Figure 5**.

On the one hand, primary KIE is governed by the reacting hydrogens. The white diamonds in **Figure 5A** and **B** highlight that at reaction transition state (TS) the proton transfer from the solvent to $O_{methoxy}$ is already complete $(CN(O_{methoxy} - H_{water}) \approx$ 0.8, see *Methods Section*). The primary KIE will in this case be close to unity or even inverse $\left(\frac{k_H}{k_D}\right)_{pri} \lesssim$ 1.[38,39]

Secondary KIE, on the other hand, involves the non-reacting hydrogen atoms. The protonation state of the water molecule closest to the guaiacol methyl group – i.e. involved in the S_N2 reaction – was therefore monitored (**Figure 5C** and **D**). The nucleophile in this case is exclusively represented by a neutral water molecule in the TS region (with $CN(H_2O_{closest} - H_{water}) \approx 1.7$) and deprotonation of the product methanol occurs only after complete C-O bond breaking. This indicates that two non-reacting hydrogens are strongly contributing to the SKIE. Previous reports in literature highlighted that such secondary KIE for non-reacting hydrogen atoms are strongly inverse, e.g. with reported values of $\left(\frac{k_H}{k_D}\right)_{sec} = 0.6$.[40]

The simulations thus suggest that the proton transfer to the guaiacol methoxy oxygen is already completed at the *O*1-S_N2 TS, leading to a negligible primary KIE, while the two non-reacting hydrogens of the attacking water molecule are responsible for the inverse SKIE, as observed experimentally in both catalytic systems (**Figure 4C**).

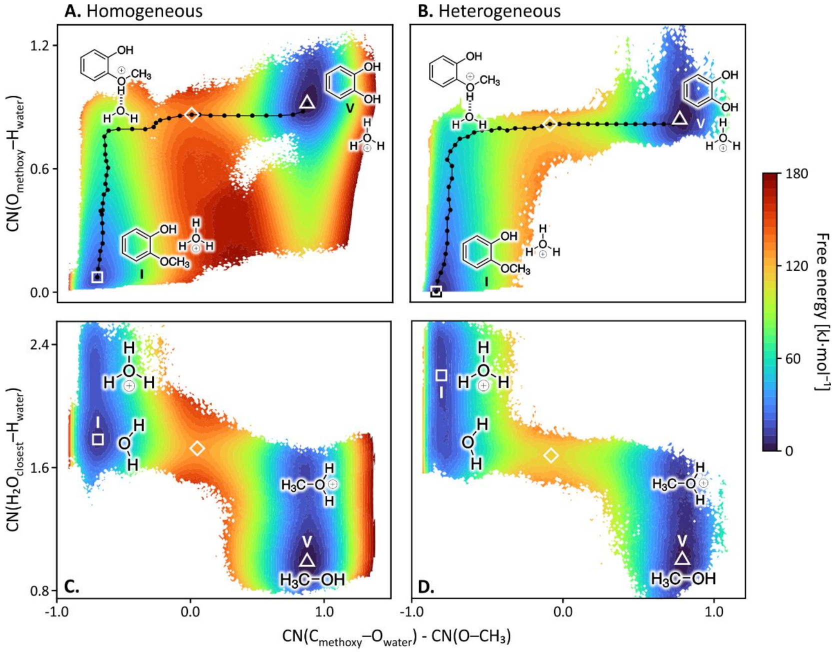


Figure 5. The proton transfer to the substrate is not involved in the rate-determining step. A and B) 2D free energy surfaces for the O1-SN2 path of the homogeneously (A) and heterogeneously (B) catalysed O-demethylation of guaiacol. The horizontal axis is the original collective variable while the vertical axis follows the protonation state of the methoxy oxygen. The minimal free energy path is shown by the dotted line. C,D) Same as in A and B, but with the vertical axis monitoring the protonation state of the closest water oxygen to the methyl group. Reactant, product and transition states are highlighted by squares, triangles and diamonds respectively. The reader is referred to the *Methods Section* for more details on the collective variable definitions used.

In contrast to previously reported mechanisms,[22–24] the proton transfer to the oxygen of the $O_{methoxy}$ group does not result in a (meta)stable oxonium ion intermediate when correctly accounting for solvent, temperature and entropic effects. Instead, an oxonium-water contact pair forms when approaching the reaction TS (**Figure 5A** and **B**). To better clarify the atomistic reasons underlying the reduced free energy requirement to reach such TS in the zeolite, we took a closer look to the proton transfer from the solvent to the guaiacol molecule by monitoring the protonation state of the water molecule closest to $O_{methoxy}$ at various stages of the protonation process (**Figure 6**). As previously mentioned, the proton-transfer reaction to $O_{methoxy}$ is completed before reaching the SN2 TS. Therefore, only states belonging to the reactant region are considered, i.e. the left part of **Figure 5A** and **B**. The protonation state of the water molecule closest to $O_{methoxy}$ was monitored within these states and at discrete values of $CN(O_{methoxy} - H_{water})$, using the coordination number between the closest water oxygen and all water hydrogen atoms.

Starting with the homogeneous case (**Figure 6A**), a neutral water molecule is most often in the closest proximity to $O_{methoxy}$ near the reactant minimum (bottom profile of **Figure 6A**). As the reaction proceeds, the interaction gradually shifts in favour of H_3O^+ as proton-donating species. Finally, at the onset of the S_N2 step (at $CN(O_{methoxy} - H_{water}) \approx 0.8$), the oxonium-water contact pair is formed. Indeed, it can be seen how the final coordination number lies in between a neutral H_2O molecule and a genuine hydronium ion H_3O^+, showing that the donating water remains very close to the protonated $O_{methoxy}$.

The zeolite-catalysed system presents a significantly different behaviour (**Figure 6B**, also apparent from **Figure 5C** and **D**). Already near the reactant state, the guaiacol $O_{methoxy}$ interacts similarly with both H_2O and H_3O^+, which is clear from the existence of two free energy minima in the lower panel of **Figure 6B**. As soon as the system gets biased towards higher free energy regions, the interaction shifts – very differently from the gradual shift in the homogeneous case – exclusively involving H_3O^+. Since the hydronium ion is undercoordinated (i.e. less solvated) in the zeolite, its electrophilicity is higher (*vide supra*). In turn, the hydronium ion is more reactive and thus more prone to transfer a proton to the substrate, explaining the stronger association between the hydronium ion and the guaiacol. Finally, at the onset of the S_N2 reaction, the system reaches an oxonium-water contact pair state similar to the homogeneous case.

These results clearly show that an enhanced association between the hydronium ion and guaiacol is present in the zeolite compared to the homogeneous catalyst. The simulations reveal that the undercoordination of the hydronium ion and its increased electrophilicity, are the driving force to coordinate more easily to guaiacol, as the undercoordinated hydronium ions seeks reaction partners to stabilize itself. A beneficial effect of the constrained environment on the association step was also proposed, based on experimental measurements, for cyclohexanol dehydration,[28] indicating that these conclusions should be of general applicability within zeolite-catalysed reactions in the condensed phase.

These results clearly demonstrate the strong interaction between the hydronium ion and guaiacol in the zeolite, as a consequence of the undercoordination-induced higher electrophilicity of the former. This is not observed in bulk water, when considering the homogeneous catalyst. A beneficial effect of the constrained environment on the association step equilibrium constant was also proposed – experimentally – for the monomolecular cyclohexanol dehydration,[28] indicating that the conclusions may be more general applicable for condensed water related reactions in presence of acid zeolites

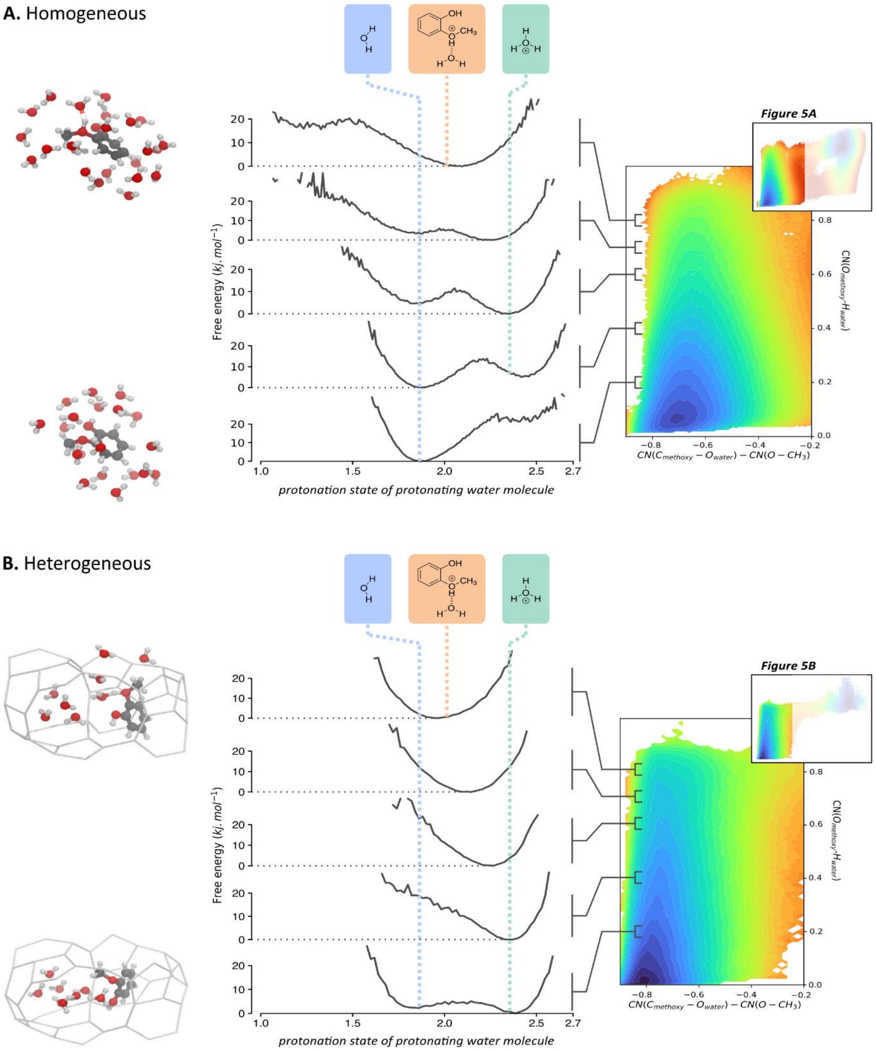


Figure 6. The association between guaiacol and the hydronium ion is associated in the constrained environment. Protonation state of the water molecule closest to $O_{methoxy}$, $CN(O_{closest} - H_{water})$, in the homogeneously (A) and heterogeneously (B) catalysed system for discrete regions along the protonation reaction of guaiacol (**Figure 5A** and **B**), *i.e.* at $CN(O_{methoxy} - H_{water}) = 0.2, 0.4, 0.6, 0.7, 0.8$ when progressing from the reactant state to the TS.

Conclusion

Herein, an experimentally verified computational study was performed to unravel the mechanism of the Brønsted acid catalysed *O*-demethylation of guaiacol in hot-pressurized water and uncover the impact of the molecular reaction environment on the proton activity.

Detailed inspection of the water structure reveals a confinement-induced reactivity enhancement of the hydronium ion that is rationalized by its lower coordination with surrounding water molecules inside the zeolite when compared to the bulk water. This translates into the formation of tube-like water structures within the zeolite channels, where most water molecules are exposed to the hydrophobic walls of the zeolite channels and hence behave very differently from the highly solvated ones in bulk water. The confined hydronium ion is intrinsically more prone to associate with guaiacol – as it is less stabilised, more electrophilic and thus more active – leading to reaction rate acceleration.

The reaction mechanism of guaiacol *O*-demethylation in presence of Brønsted acids proceeds via a concerted S_N2 mechanism – referred to as *O1*-S_N2 – regardless of the proton's molecular environment. Kinetic experiments show first order in the Brønsted acid concentration – both in the homogeneous as in the zeolite case – whereas the reaction is first-order in guaiacol in the bulk and near zeroth-order in the zeolite, due to the the intra-pore concentration that is nearly independent of the concentration in solution under experiment conditions. The theoretical results exclude the formation of a stable oxonium-intermediate in the *O1*-S_N2 mechanism and, in contrast to earlier proposals, an elusive oxonium-water contact-pair is observed instead. The latter can react with a neutral water molecule leading to catechol and methanol in a concerted manner.[22–24] This mechanistic proposal is in line with the inverse SKIE observed experimentally.

The *O1*-S_N2 reaction free energy profile shows the lowest energetic cost with zeolite to bring the hydronium ion close to guaiacol, forming the reactive contact pair. The higher electrophilicity of the hydronium ion, as a result of the undercoordination, facilitate the transfer of the excess proton to guaiacol, leading to an overall increase of the reaction rate in the zeolite environment. We believe that the reported findings are not exclusive to the guaiacol *O*-demethylation reaction but are of general applicability within the field of Brønsted acid catalysed reactions in hot liquid water.

Methods

Both the homo- and heterogeneous catalytic systems were investigated by means of periodic *ab initio* molecular dynamics (AIMD), providing an accurate description of the entropic effects at operating conditions and a more accurate derivation of reaction kinetics based on classical transition state theory. Hence, in comparison to typical static approaches, the molecular environment is more realistically accounted for. The general systems setup is shown in **Figure 2A** and **B**. A combination of well-tempered Metadynamics and Umbrella Sampling (*vide infra*) was employed in combination with AIMD to describe the reactive events.[41] The computational details are summarized below. We refer the interested reader to the *Supplementary information, Section **S1.1.*** for an extensive discussion. Details about the experimental methodology are reported in detail in *Section **S1.2*** of the *Supplementary information.*

System set up

The simulation unit cells of both investigated systems are shown in **Figure 2A** and **B.** The homogeneous system accounts for two solvation layers (171 water molecules), whereas the number of solvent molecules in the heterogeneous system was determined from grand canonical Monte Carlo (MC) simulations (22 water molecules, see *Supplementary Information* **S1.1.1**). Each of the systems accounts for a single active site, i.e. 1 excess proton. While it is known that zeolites can easily undergo reversible framework bonds breaking and formation when exposed to liquid water,[42] we here focus on a pristine BAS to obtain some general reactivity trends while leaving the arguably much more complex effect of framework-associated and extra-framework aluminium species for future investigations. System preparation and equilibration were performed using a combination of classical molecular dynamics simulations with in-house developed force fields for the solutes and *ab initio* MD simulations at the Density Functional Theory (DFT) level in the appropriate thermodynamic ensembles.

For the homogeneous catalysed system initial system configurations were prepared by means of molecular mechanics simulations in the *isobaric-isothermal (NPT)* ensemble, whereas the heterogeneous catalysed system was prepared by means of MC/MD simulations, i.e. a combination of grand canonical Monte Carlo and molecular dynamics. More details on the system preparation can be found in *Section **S1.1*** of the *Supporting Information.*

Enhanced and regular Ab initio molecular dynamics simulations

Subsequently, AIMD simulations were performed in the canonical ensemble by means of CP2K 5.1 with a unit cell measuring 18.7 Å x 18.7 Å x 18.7 Å and 12.7 Å x 12.7 Å x 26.7 Å (angles = 90°) for the homogeneous and heterogeneous catalysed systems respectively.[43,44] The BLYP-D3 functional was chosen in combination with the TZVP-GTH basis set as the level of theory.[45–48] The Nosé-hoover thermostat was used for temperature control (523 K) in combination with a timestep of 0.5 fs for the integration of the equations of motion.[49,50] The convergence criterion for the self-consistent field method was fixed at 10^{-5} Hartree.

All enhanced sampling AIMD simulations were performed by means of PLUMED, interfaced with CP2K as MD engine.[51] In order to account for the indistinguishability of the reacting water molecules, coordination numbers or linear combinations of coordination numbers are employed as reaction coordinates and as collective variables for post-processing purposes, see *Supplementary information **S1.1.3.1***. As an example, the collective variables used to describe the governing mechanism (O1-S_N2, see **Figure 5A** and **B**) are schematically depicted in **Figure 7**.

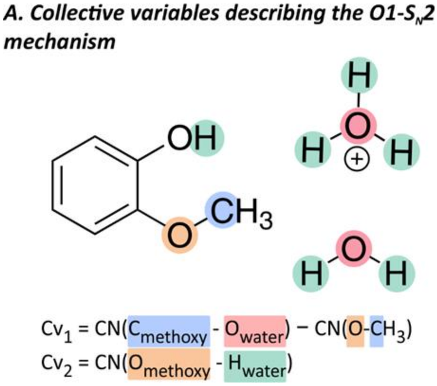


Figure 7. Definition of the collective variables and corresponding coordination numbers describing the O1-S_N2 mechanism. Cv_1 describes the S_N2 mechanism and Cv_2 describes the proton transfer from the water to the guaiacol methoxy oxygen.

In this study, a two-step enhanced sampling AIMD approach was used to obtain a converged free energy profile (FEP), combining the speed of well-tempered multiple walker metadynamics (wt MTD) for an efficient reaction coordinate screening and the accuracy of umbrella sampling (US) to obtain a final qualitative and quantitative estimate of the free energy surface.[51–55] The methodology was benchmarked with respect to previous work by some of the authors and a more elaborate discussion on the applied approach is given in *Section **S3*** of the *Supplementary information*.

wt MTD was performed using 4 to 8 walkers, a bias factor of 10-12, an initial hill height of 2 to 4 kJ·mol^{-1} and a hill width σ of 0.03. Remark that the choice of these settings was guided by our previous static DFT results.[25] The deposition rate and walkers stride was set to 100 steps, i.e. every 50 fs a hill was deposited and every 50 fs the hill deposition files are flushed. The sum_hills utility of PLUMED was used during post-processing for the construction of the FEP.

US simulations were performed to obtain an accurate estimate of the real FEP employing a harmonic bias potentials centred around the equilibrium value q_0:

$$U_b(q) = \frac{\kappa}{2}(q - q_0)^2$$

With κ the force constant typically amounting to 250 kJ·mol^{-1}. For each reaction step ‘umbrellas’ were positioned along the reaction coordinate with a typical increment of 0.1. For each of these umbrellas a biased MD simulation was performed, i.e. biased using an harmonic potential and a potential equal to the inverse of the ‘crude’ FEP (obtained via wt MTD). In some cases, extra umbrellas were added, or force constants were increased to improve the sampling of difficult regions along the free energy surface. Subsequent construction of the FEP is performed by means of the weighted histogram analysis method (WHAM) accounting for both the harmonic and additional bias (originating from the wt MTD simulations).[56–58] For certain reaction steps the 2-dimensional (2D) variant is used, i.e. instead of using a single reaction coordinate two collective variables are used to describe the desired transformation. In this case 2D harmonic potentials and a 2D-wham analysis are employed. Subsequent error estimates are extracted by a bootstrapping procedure with replacement of the sampled CV-space, for a total of 1000 generated samples. The minimal free energy path is extracted by means of the MEPSA python package.[59] Mechanistic features for each elementary reaction step are extracted by means of free energy transformations using the ThermoLIB python package, for more details the reader is referred to the *Supplementary information*.[58]

Acknowledgements

The authors acknowledge financial support from the Research Foundation Flanders (FWO) (BioFact Excellence of Science project Grant No. 30902231). The computational resources and services used were provided by Ghent University (Stevin Supercomputer Infrastructure), the VSC (Flemish Supercomputer Center), funded by Ghent University, the Research Foundation - Flanders (FWO) and the Flemish Government. E.V.D.B, M. Bocus, L.V. and V.V.S acknowledge funding from FWO. V.V.S. and L.V. acknowledge the Research Board of the Ghent University. B.U.W.M. thanks the Francqui Foundation for an appointment as Collen-Francqui professor and UAntwerp for BOF GOA support. We thank Ruben Goeminne for his assistance with the GCMC simulations. X.W. thanks the Chinese Scholarship Council (No. 201806890016) for financial support.

Author contributions

E.V.D.B., V.V.S., B.U.W.M. and B.F.S. initiated the discussion and designed the scope of the project. E.V.D.B. conceived and coordinated the research under the supervision of V.V.S. E.V.D.B. and M. Bocus performed all DFT-based calculations related to the homogeneous catalysed and heterogeneous catalysed systems, respectively, under the supervision of V.V.S. M. Bal performed all the kinetic experiments for the homogeneously catalysed *O*-dealkylation under the supervision of B.U.W.M. X.W. performed all the kinetic experiments for the heterogeneously catalysed *O*-dealkylation under the supervision of B.F.S. J.B. performed the initial *O*-dealkylation reactions to unravel its S_N-character and the labelled water experiments under the supervision of B.U.W.M. L.V.D. was responsible for the development of the thermoLIB library utilised for the analysis of all enhanced sampling simulations and provided guidance throughout. E.V.D.B., M. Bocus and V.V.S wrote the first version of the manuscript revised with input from all other co-authors. The supporting information was compiled by E.V.D.B., M. Bocus, J.B., M.Bal and X.W.

Data Availability

An extensive description of the performed simulations and experiments is available in the Supplementary Information. Any unavailable data that support the findings of this study can be requested from the authors.

Code Availability

CP2K (https://github.com/cp2k/cp2k) and PLUMED (https://github.com/plumed/plumed2) are open-source software and freely available at the provided links. ThermoLIB is available upon request at https://molmod.ugent.be/software/thermolib.

References

(1) IPCC; Masson-Delmotte, V.; Zhai, P.; Pirani, A.; Connors, S. L.; Péan, C.; Berger, S.; Caud, N.; Chen, Y.; Goldfarb, L.; et al. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; 2021.

(2) Sheldon, R. A. Biocatalysis and Biomass Conversion: Enabling a Circular Economy. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2020**, *378* (2176), 20190274. http://doi.org/10.1098/rsta.2019.0274.

(3) Mazmanian, D. A.; Jurewitz, J.; Nelson, H. T. A Governing Framework for Climate Change Adaptation in the Built Environment. *Ecol. Soc.* **2013**, *18* (4), art56. http://doi.org/10.5751/es-05976-180456.

(4) Liao, Y.; Koelewijn, S.-F.; Van den Bossche, G.; Van Aelst, J.; Van den Bosch, S.; Renders, T.; Navare, K.; Nicolaï, T.; Van Aelst, K.; Maesen, M.; et al. A Sustainable Wood Biorefinery for Low–Carbon Footprint Chemicals Production. *Science* **2020**, *367* (6484), 1385–1390. http://doi.org/10.1126/science.aau1567.

(5) Graglia, M.; Kanna, N.; Esposito, D. Lignin Refinery: Towards the Preparation of Renewable Aromatic Building Blocks. *ChemBioEng Rev.* **2015**, *2* (6), 377–392. http://doi.org/10.1002/cben.201500019.

(6) Schutyser, W.; Renders, T.; Van Den Bosch, S.; Koelewijn, S. F.; Beckham, G. T.; Sels, B. F. Chemicals from Lignin: An Interplay of Lignocellulose Fractionation, Depolymerisation, and Upgrading. *Chem. Soc. Rev.* **2018**, *47* (3), 852–908. http://doi.org/10.1039/c7cs00566k.

(7) Bartling, A. W.; Stone, M. L.; Hanes, R. J.; Bhatt, A.; Zhang, Y.; Biddy, M. J.; Davis, R.; Kruger, J. S.; Thornburg, N. E.; Luterbacher, J. S.; et al. Techno-Economic Analysis and Life Cycle Assessment of a Biorefinery Utilizing Reductive Catalytic Fractionation. *Energy Environ. Sci.* **2021**, *14* (8), 4147–4168. http://doi.org/10.1039/d1ee01642c.

(8) Ha, J.-M.; Hwang, K.-R.; Kim, Y.-M.; Jae, J.; Kim, K. H.; Lee, H. W.; Kim, J.-Y.; Park, Y.-K. Recent Progress in the Thermal and Catalytic Conversion of Lignin. *Renew. Sustain. Energy Rev.* **2019**, *111*, 422–441. http://doi.org/10.1016/j.rser.2019.05.034.

(9) Zakzeski, J.; Bruijnincx, P. C. A.; Jongerius, A. L.; Weckhuysen, B. M. The Catalytic Valorization of Lignin for the Production of Renewable Chemicals. *Chem. Rev.* **2010**, *110* (6), 3552–3599. http://doi.org/10.1021/cr900354u.

(10) Vanholme, R.; Demedts, B.; Morreel, K.; Ralph, J.; Boerjan, W. Lignin Biosynthesis and Structure. *Plant Physiol.* **2010**, *153* (3), 895–905. http://doi.org/10.1104/pp.110.155119.

(11) Liu, X.; Bouxin, F. P.; Fan, J.; Budarin, V. L.; Hu, C.; Clark, J. H. Recent Advances in the Catalytic Depolymerization of Lignin towards Phenolic Chemicals: A Review. *ChemSusChem* **2020**, *13* (17), 4296–4317. http://doi.org/10.1002/cssc.202001213.

(12) Sheldon, R. A. Green and Sustainable Manufacture of Chemicals from Biomass: State of the Art. *Green Chem.* **2014**, *16* (3), 950–963. http://doi.org/10.1039/c3gc41935e.

(13) Fiege, H.; Voges, H.-W.; Hamamoto, T.; Umemura, S.; Iwata, T.; Miki, H.; Fujita, Y.; Buysch, H.-J.; Garbe, D.; Paulus, W. Phenol Derivatives. In *Ullmann's Encyclopedia of Industrial Chemistry*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2000. http://doi.org/10.1002/14356007.a19_313.

(14) Sun, Z.; Fridrich, B.; De Santi, A.; Elangovan, S.; Barta, K. Bright Side of Lignin Depolymerization: Toward New Platform Chemicals. *Chem. Rev.* **2018**, *118* (2), 614–678. http://doi.org/10.1021/acs.chemrev.7b00588.

(15) Zhang, L.; Hu, G. Supply Chain Design and Operational Planning Models for Biomass to Drop-in Fuel Production. *Biomass and Bioenergy* **2013**, *58*, 238–250. http://doi.org/10.1016/j.biombioe.2013.08.016.

(16) Cheng, F.; Brewer, C. E. Producing Jet Fuel from Biomass Lignin: Potential Pathways to Alkyl-Benzenes and Cycloalkanes. *Renew. Sustain. Energy Rev.* **2017**, *72*, 673–722. http://doi.org/10.1016/j.rser.2017.01.030.

(17) Zhao, S.; Abu-Omar, M. M. Renewable Thermoplastics Based on Lignin-Derived Polyphenols. *Macromolecules* **2017**, *50* (9), 3573–3581. http://doi.org/10.1021/acs.macromol.7b00064.

(18) Bohnet, M. editor. *Ullmann's Encyclopedia of Industrial Chemistry*, 6th, compl ed.; Weinheim, 2003.

(19) Siragusa, F.; Van Den Broeck, E.; Ocando, C.; Müller, A. J.; De Smet, G.; Maes, B. U. W.; De Winter, J.; Van Speybroeck, V.; Grignard, B.; Detrembleur, C. Access to Biorenewable and CO 2 -Based Polycarbonates from Exovinylene Cyclic Carbonates. *ACS Sustain. Chem. Eng.* **2021**, *9* (4), 1714–1728. http://doi.org/10.1021/acssuschemeng.0c07683.

(20) Yang, L.; Seshan, K.; Li, Y. A Review on Thermal Chemical Reactions of Lignin Model Compounds. *Catal. Today* **2017**, *298*, 276–297. http://doi.org/10.1016/j.cattod.2016.11.030.

(21) Bomon, J.; Bal, M.; Achar, T. K.; Sergeyev, S.; Wu, X.; Wambacq, B.; Lemière, F.; Sels, B. F.; Maes, B. U. W. Efficient Demethylation of Aromatic Methyl Ethers with HCl in Water. *Green Chem.* **2021**, *23* (5), 1995–2009. http://doi.org/10.1039/d0gc04268d.

(22) Wu, X.; Fu, J.; Lu, X. Kinetics and Mechanism of Hydrothermal Decomposition of Lignin Model Compounds. *Ind. Eng. Chem. Res.* **2013**, *52* (14), 5016–5022. http://doi.org/10.1021/ie302898q.

(23) Yang, L.; Zhou, W.; Seshan, K.; Li, Y. Green and Efficient Synthesis Route of Catechol from Guaiacol. *J. Mol. Catal. A Chem.* **2013**, *368–369*, 61–65. http://doi.org/10.1016/j.molcata.2012.11.024.

(24) Li, Z.; Sutandar, E.; Goihl, T.; Zhang, X.; Pan, X. Cleavage of Ethers and Demethylation of

Lignin in Acidic Concentrated Lithium Bromide (ACLB) Solution. *Green Chem.* **2020**, *22* (22), 7989–8001. http://doi.org/10.1039/d0gc02581j.

(25) Bomon, J.; Van Den Broeck, E.; Bal, M.; Liao, Y.; Sergeyev, S.; Van Speybroeck, V.; Sels, B. F.; Maes, B. U. W. Brønsted Acid Catalyzed Tandem Defunctionalization of Biorenewable Ferulic Acid and Derivates into Bio-Catechol. *Angew. Chem. Int. Ed.* **2020**, *59* (8), 3063–3068. http://doi.org/10.1002/anie.201913023.

(26) Ennaert, T.; Van Aelst, J.; Dijkmans, J.; De Clercq, R.; Schutyser, W.; Dusselier, M.; Verboekend, D.; Sels, B. F. Potential and Challenges of Zeolite Chemistry in the Catalytic Conversion of Biomass. *Chem. Soc. Rev.* **2016**, *45* (3), 584–611. http://doi.org/10.1039/c5cs00859j.

(27) Wu, X.; Liao, Y.; Bomon, J.; Tian, G.; Bai, S.-T.; Van Aelst, K.; Zhang, Q.; Vermandel, W.; Wambacq, B.; Maes, B. U. W.; et al. Lignin-First Monomers to Catechol: Rational Cleavage of C−O and C−C Bonds over Zeolites. *ChemSusChem* **2022**, *15* (7). http://doi.org/10.1002/cssc.202102248.

(28) Liu, Y.; Vjunov, A.; Shi, H.; Eckstein, S.; Camaioni, D. M.; Mei, D.; Baráth, E.; Lercher, J. A. Enhancing the Catalytic Activity of Hydronium Ions through Constrained Environments. *Nat. Commun.* **2017**, *8*, 14113. http://doi.org/10.1038/ncomms14113.

(29) Stanciakova, K.; Weckhuysen, B. M. Water–Active Site Interactions in Zeolites and Their Relevance in Catalysis. *Trends Chem.* **2021**, *3* (6), 456–468. http://doi.org/10.1016/j.trechm.2021.03.004.

(30) Bates, J. S.; Bukowski, B. C.; Greeley, J.; Gounder, R. Structure and Solvation of Confined Water and Water–Ethanol Clusters within Microporous Brønsted Acids and Their Effects on Ethanol Dehydration Catalysis. *Chem. Sci.* **2020**, *11* (27), 7102–7122. http://doi.org/10.1039/d0sc02589e.

(31) Eckstein, S.; Hintermeier, P. H.; Zhao, R.; Baráth, E.; Shi, H.; Liu, Y.; Lercher, J. A. Influence of Hydronium Ions in Zeolites on Sorption. *Angew. Chemie Int. Ed.* **2019**, *58* (11), 3450–3455. http://doi.org/10.1002/anie.201812184.

(32) Pfriem, N.; Hintermeier, P. H.; Eckstein, S.; Kim, S.; Liu, Q.; Shi, H.; Milakovic, L.; Liu, Y.; Haller, G. L.; Baráth, E.; et al. Role of the Ionic Environment in Enhancing the Activity of Reacting Molecules in Zeolite Pores. *Science* **2021**, *372* (6545), 952–957. http://doi.org/10.1126/science.abh3418.

(33) Shi, H.; Eckstein, S.; Vjunov, A.; Camaioni, D. M.; Lercher, J. A. Tailoring Nanoscopic Confines to Maximize Catalytic Activity of Hydronium Ions. *Nat. Commun.* **2017**, *8* (1), 15442. http://doi.org/10.1038/ncomms15442.

(34) De Wispelaere, K.; Bailleul, S.; Van Speybroeck, V. Towards Molecular Control of Elementary Reactions in Zeolite Catalysis by Advanced Molecular Simulations Mimicking Operating Conditions. *Catal. Sci. Technol.* **2016**, *6* (8), 2686–2705. http://doi.org/10.1039/c5cy02073e.

(35) Pérez De Alba Ortíz, A.; Tiwari, A.; Puthenkalathil, R. C.; Ensing, B. Advances in Enhanced Sampling along Adaptive Paths of Collective Variables. *J. Chem. Phys.* **2018**, *149* (7), 072320. http://doi.org/10.1063/1.5027392.

(36) Bregante, D. T.; Chan, M. C.; Tan, J. Z.; Ayla, E. Z.; Nicholas, C. P.; Shukla, D.; Flaherty, D. W. The Shape of Water in Zeolites and Its Impact on Epoxidation Catalysis. *Nat. Catal.* **2021**, *4* (9), 797–808. http://doi.org/10.1038/s41929-021-00672-4.

(37) Bukowski, B. C.; Bates, J. S.; Gounder, R.; Greeley, J. Defect-Mediated Ordering of Condensed Water Structures in Microporous Zeolites. *Angew. Chemie Int. Ed.* **2019**, *58* (46), 16422–16426. http://doi.org/10.1002/anie.201908151.

(38) Melander, L. Isotope Effects in Proton-Transfer Reactions. *ACTA Chem. Scand.* **1971**, *25* (10), 3821–3826.

(39) Bigeleisen, J. Correlation of Kinetic Isotope Effects with Chemical Bonding in Three-Centre Reactions. *Pure Appl. Chem.* **1964**, *8* (3–4), 217–224. http://doi.org/10.1351/pac196408030217.

(40) Kresge, A. J.; Onwood, D. P. The Secondary Isotope Effect on Proton Transfer from the Hydronium Ion in Aqueous Solution. *J. Am. Chem. Soc.* **1964**, *86* (22), 5014–5016. http://doi.org/10.1021/ja01076a063.

(41) Ensing, B.; Laio, A.; Parrinello, M.; Klein, M. L. A Recipe for the Computation of the Free Energy Barrier and the Lowest Free Energy Path of Concerted Reactions. *J. Phys. Chem. B* **2005**, *109* (14), 6676–6687. http://doi.org/10.1021/jp045571i.

(42) Heard, C. J.; Grajciar, L.; Uhlík, F.; Shamzhy, M.; Opanasenko, M.; Čejka, J.; Nachtigall, P. Zeolite (In)Stability under Aqueous or Steaming Conditions. *Adv. Mater.* **2020**, *32* (44), 2003264. http://doi.org/10.1002/adma.202003264.

(43) Vandevondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. Quickstep: Fast and Accurate Density Functional Calculations Using a Mixed Gaussian and Plane Waves Approach. *Comput. Phys. Commun.* **2005**, *167* (2), 103–128. http://doi.org/10.1016/j.cpc.2004.12.014.

(44) Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schütt, O.; Schiffmann, F.; et al. CP2K: An Electronic Structure and Molecular Dynamics Software Package -Quickstep: Efficient and Accurate Electronic Structure Calculations. *J. Chem. Phys.* **2020**, *152*, 194103. http://doi.org/10.1063/5.0007045.

(45) VandeVondele, J.; Hutter, J. Gaussian Basis Sets for Accurate Calculations on Molecular Systems in Gas and Condensed Phases. *J. Chem. Phys.* **2007**, *127* (11), 114105. http://doi.org/10.1063/1.2770708.

(46) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15), 154104. http://doi.org/10.1063/1.3382344.

(47) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38* (6), 3098–3100. http://doi.org/10.1103/physreva.38.3098.

(48) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37* (2), 785–789. http://doi.org/10.1016/0009-2614(89)85118-8.

(49) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé-Hoover Chains: The Canonical Ensemble via Continuous Dynamics. *J. Chem. Phys.* **1992**, *97* (4), 2635–2643. http://doi.org/10.1063/1.463940.

(50) Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52* (2), 255–268. http://doi.org/10.1080/00268978400101201.

(51) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604–613. http://doi.org/10.1016/j.cpc.2013.09.018.

(52) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (20), 12562–12566. http://doi.org/10.1073/pnas.202427399.

(53) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. *J. Phys. Chem. B* **2006**, *110* (8), 3533–3539. http://doi.org/10.1021/jp054359r.

(54) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100* (2), 020603. http://doi.org/10.1103/physrevlett.100.020603.

(55) Grubmüller, H.; Heymann, B.; Tavan, P. Ligand Binding: Molecular Mechanics Calculation of the Streptavidin-Biotin Rupture Force. *Science* **1996**, *271* (5251), 997–999. http://doi.org/10.1126/science.271.5251.997.

(56) Souaille, M.; Roux, B. Extension to the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations. *Comput. Phys. Commun.* **2001**, *135* (1), 40–57. http://doi.org/10.1016/s0010-4655(00)00215-0.

(57) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13* (8), 1011–1021. http://doi.org/10.1002/jcc.540130812.

(58) ThermoLIB | Center for Molecular Modeling https://molmod.ugent.be/software/thermolib (accessed Nov 19, 2021).

(59) Marcos-Alcalde, I.; Setoain, J.; Mendieta-Moreno, J. I.; Mendieta, J.; Gómez-Puertas, P. MEPSA: Minimum Energy Pathway Analysis for Energy Landscapes. *Bioinformatics* **2015**, *31* (23), 3853–3855. http://doi.org/10.1093/bioinformatics/btv453.

Tables

Table 1. Summary of the rate constants and corresponding phenomenological barriers for all the investigated guaiacol demethylation reaction steps.

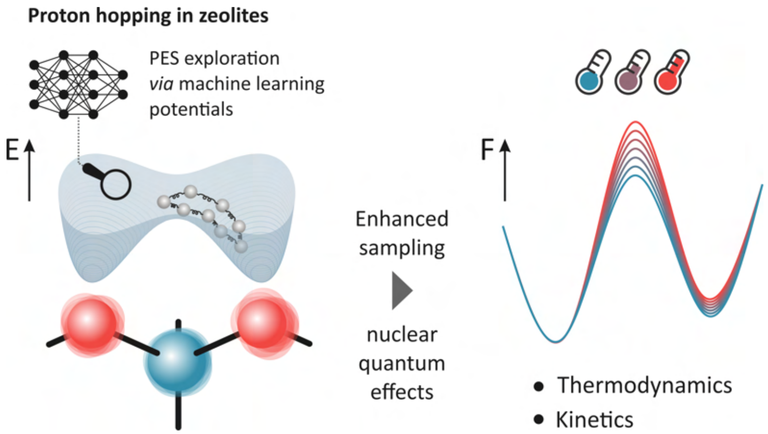
StepID	Catalyst		Rate constant k [s^{-1}]		K [$k_{forward}$/ $k_{backward}$]	Phenomenological barrier ΔF [$kJ \cdot mol^{-1}$]	
	Homo	Hetero	Forward	Backward		Forward	Backward
1 (**I** → **II**)	x		$7.9 \times 10^{5} \pm 0.4 \times 10^{5}$	$5.8 \times 10^{11} \pm 0.2 \times 10^{11}$	**1.4×10^{-7}**	71.5 ± 0.2	2.7 ± 0.2
		x	$5.2 \times 10^{3} \pm 0.4 \times 10^{3}$	$2.3 \times 10^{11} \pm 0.1 \times 10^{11}$	**2.3×10^{-8}**	93.4 ± 0.3	16.7 ± 0.2
2 (**II** → **IV**)	x		$1.01 \times 10^{6} \pm 0.04 \times 10^{6}$	$4.9 \times 10^{2} \pm 0.2 \times 10^{2}$	2.2×10^{3}	70.4 ± 0.2	103.6 ± 0.2
		x	$7.8 \times 10^{4} \pm 0.5 \times 10^{4}$	$5.2 \times 10^{3} \pm 0.3 \times 10^{3}$	1.5×10^{1}	81.6 ± 0.3	93.3 ± 0.3
3 (**II** → **III**)	x		$2.83 \times 10^{4} \pm 0.08 \times 10^{4}$	$6.8 \times 10^{11} \pm 0.2 \times 10^{11}$	4.1×10^{-8}	86.0 ± 0.1	12.1 ± 0.1
		x	$8.3 \times 10^{1} \pm 0.5 \times 10^{1}$	$6.2 \times 10^{11} \pm 0.2 \times 10^{11}$	1.3×10^{-10}	111.4 ± 0.3	12.5 ± 0.2
4 (**III** → **IV**)	x		$1.2 \times 10^{12} \pm 0.05 \times 10^{12}$	$7.6 \times 10^{1} \pm 0.3 \times 10^{1}$	1.6×10^{10}	9.4 ± 0.2	111.7 ± 0.2
		-	-	-	-	-	-
5 (**IV** → **VII**)	x		$9.5 \times 10^{11} \pm 0.4 \times 10^{11}$	$7.4 \times 10^{14} \pm 0.3 \times 10^{14}$	1.3×10^{-3}	10.6 ± 0.2	-18.3 ± 0.2
		-	-	-	-	-	-
6 (**VII** → **V**)	x		$7.9 \times 10^{11} \pm 0.9 \times 10^{11}$	$8.8 \times 10^{5} \pm 1.0 \times 10^{5}$	9.0×10^{5}	11.4 ± 0.5	71.0 ± 0.5
		-	-	-	-	-	-
7 (**I** → **V**)	x		$9.9 \times 10^{1} \pm 0.3 \times 10^{1}$	$1.92 \times 10^{1} \pm 0.06 \times 10^{1}$	**5.2×10^{0}**	110.6 ± 0.1	117.7 ± 0.1
		x	$2.0 \times 10^{3} \pm 0.1 \times 10^{3}$	$2.3 \times 10^{2} \pm 0.2 \times 10^{2}$	**8.7×10^{0}**	97.5 ± 0.3	106.8 ± 0.3

Paper V

Nuclear quantum effects on zeolite proton hopping kinetics explored with machine learning potentials and path integral molecular dynamics

Massimo Bocus‡, Ruben Goeminne‡, Aran Lamaire, Maarten Cools-Ceuppens, Toon Verstraelen and Veronique Van Speybroeck*

Nature Communications, **2023**, *14*, 1008



*Corresponding author
‡Authors contributed equally
M. Bocus performed the DFT simulations used to train and validate the MLP and prepared the initial draft of the manuscript.

nature communications

Article

https://doi.org/10.1038/s41467-023-36666-y

Nuclear quantum effects on zeolite proton hopping kinetics explored with machine learning potentials and path integral molecular dynamics

Received: 26 July 2022

Accepted: 10 February 2023

Massimo Bocus[1,2], Ruben Goeminne[1,2], Aran Lamaire[1],
Maarten Cools-Ceuppens[1], Toon Verstraelen[1] & Veronique Van Speybroeck[1] ✉

Proton hopping is a key reactive process within zeolite catalysis. However, the accurate determination of its kinetics poses major challenges both for theoreticians and experimentalists. Nuclear quantum effects (NQEs) are known to influence the structure and dynamics of protons, but their rigorous inclusion through the path integral molecular dynamics (PIMD) formalism was so far beyond reach for zeolite catalyzed processes due to the excessive computational cost of evaluating all forces and energies at the Density Functional Theory (DFT) level. Herein, we overcome this limitation by training first a reactive machine learning potential (MLP) that can reproduce with high fidelity the DFT potential energy surface of proton hopping around the first Al coordination sphere in the H-CHA zeolite. The MLP offers an immense computational speedup, enabling us to derive accurate reaction kinetics beyond standard transition state theory for the proton hopping reaction. Overall, more than 0.6 μs of simulation time was needed, which is far beyond reach of any standard DFT approach. NQEs are found to significantly impact the proton hopping kinetics up to ~473 K. Moreover, PIMD simulations with deuterium can be performed without any additional training to compute kinetic isotope effects over a broad range of temperatures.

Brønsted-acidic zeolites are versatile, resistant catalysts that for decades have been recognized as the workhorse of the petrochemical industry[1]. Furthermore, they are also expected to play a vital role in next-generation biorefineries for the conversion of non-fossil feedstocks[2]. From a theoretical point of view, zeolites belong to the most studied materials in the field of heterogeneous computational catalysis[3]. In their ideal, defect-free crystalline form, Brønsted-acidic zeolites are composed of interconnected SiO_4 tetrahedra, where a fraction of the Si^{4+} ions is substituted by Al^{3+}. The excess of negative charge is compensated by the addition of a proton—the Brønsted Acid Site (BAS)—on one of the oxygens in the first coordination sphere of the Al substitution. Interestingly, the BAS is not confined to a specific oxygen of the Al tetrahedron, but it can jump from one oxygen atom to another in what is commonly known as the 'proton hopping' reaction. This process is one of the most fundamental activated events within zeolite chemistry (Fig. 1a) and represents the archetypal proton-transfer reaction which is at the base of any Brønsted acid-catalyzed reaction.

Because of its apparent simplicity, proton hopping is an ideal case study for both experiment and theory, hence various methods have been used to investigate the process kinetics. Experimentally, Nuclear Magnetic Resonance (NMR)[4–6], Impedance Spectroscopy (IS)[7] and

---

[1]Center for Molecular Modeling, Ghent University, Technologiepark 46, 9052 Zwijnaarde, Belgium. [2]These authors contributed equally: Massimo Bocus, Ruben Goeminne. ✉e-mail: Veronique.Vanspeybroeck@Ugent.be

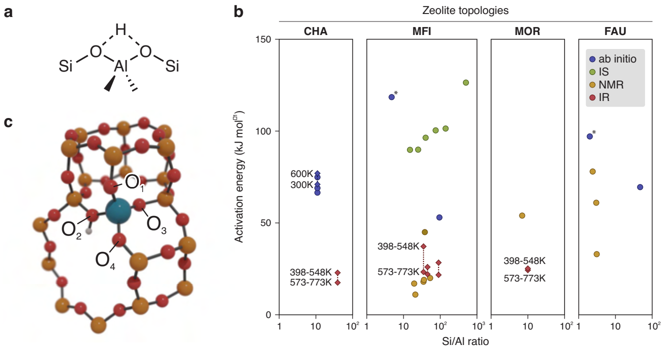


Fig. 1 | Poor agreement is found in the available literature for the activation energy of the proton hopping reaction. a Schematic depiction of the proton hopping reaction. **b** Activation energy for the proton hopping process as function of the Si/Al ratio for multiple zeolites, as derived from the available literature[75]. The data was obtained from ab initio calculations (blue, the points with an asterisk correspond to cluster calculations and, therefore, the Si/Al ratio is meaningless) or IR (red), NMR (yellow) and IS (green) spectroscopies. If more values are available for different temperatures, they are reported with diamonds interconnected by a dotted line. For more details about the reported values and a full list of references, the interested reader is referred to Supplementary Note 1. **c** Part of the H-CHA unit cell, showing the conventional nomenclature of the oxygen atoms in the first coordination sphere of the Al defect adopted herein (Si is in yellow, O in red, Al in blue and H in white).

InfraRed spectroscopy (IR)[8] have been employed to retrieve the activation energies for the proton hopping process. From the theoretical side, the reaction has been tackled with various methodologies ranging from static simulations[9–15] to enhanced-sampling techniques based on molecular dynamics (MD)[16]. Given this plethora of scientific reports, it would be tempting to assume that every detail of the proton hopping reaction is now revealed. However, when surveying the available literature, a huge spread in both the theoretically and experimentally obtained activation energies for proton hopping barriers can be found (Fig. 1b).

In general, activation energies derived from NMR spectroscopy are lower than the theoretical ones. From IR experiments, two different activation energies were obtained for two different temperature ranges (398-548 and 573–773 K, see red diamonds in Fig. 1), a fact that was attributed to the switch from intra-site hopping to inter-site hopping[8]. However, a more recent investigation has disproven such interpretation and indicated active site proximity effects as the main cause for the observed change in activation energy[17]. Inter-site hopping was also suggested to be responsible for the high activation energies retrieved with IS[7].

To understand this lack of consistency, it is important to consider the main possible sources of discrepancy between the proton hopping barriers from literature. First, the residual presence of water in the catalyst is often indicated as the main source for the—in general—low experimental barriers[10], as it is almost impossible to achieve a completely dry material with routine drying procedures[18]. Moreover, the presence of defective sites like extra-framework aluminum species is known to alter the BAS' acidity compared to the pristine material[19]. On the theoretical side, most of the calculations performed so far did not explicitly account for the quantum nature of the hydrogen nucleus. Instead, the nuclei in the system are treated as classical particles moving on an underlying Potential Energy Surface (PES), which is obtained by solving the electronic many-body problem using quantum many-body techniques. This is normally done using Density Functional Theory (DFT) for the sake of computational efficiency. In what follows, the terminology 'classical DFT PES' will be used to refer to nuclei that are treated as classical particles on a DFT-determined PES, thus the electronic degrees of freedom are treated quantum mechanically whereas the nuclei are treated as classical particles. To include Nuclear Quantum Effects (NQEs), approximative methods have been used. For example, tunneling corrections have sometimes been applied to account for NQEs[12]. To more rigorously account for NQEs, one should resort to computationally more expensive methods such as the Path Integral Molecular Dynamics (PIMD) approach, which relies on Feynman's path integral formulation of quantum mechanics. Within PIMD simulations, the statistics of quantum particles are retrieved using a classical ring polymer consisting of *P* replicas of the system[20]. Each replica runs on the classical DFT PES, making PIMD *P* times more expensive than a standard MD simulation. This is because an independent DFT-level energy and force evaluation must be performed every MD step for each replica. Within the field of heterogeneous-catalyzed reactions such simulations have so far been mostly unfeasible due to the high computational cost of each PIMD step, as at least 10 replicas are usually required to achieve converged results—making the simulations prohibitively expensive[21]. Nonetheless, there is clear evidence that NQEs may have a significant impact on the physicochemical properties of systems containing light atoms[20,22–24]. For example, it is well-known that they can significantly affect the strength of hydrogen bonds in a variety of systems[25,26]. NQEs have never been explicitly included in zeolite-related reactions and thus it remains unclear to what extent they would affect the rate of proton hopping and by extension any proton-transfer reactions within the field of zeolite catalysis.

To fill this gap of knowledge, we present in the current contribution a methodology that may allow to systematically include NQEs when investigating activated hydrogen transfer events. To this end, proton hopping in H-CHA with isolated active sites is used as a case study (Fig. 1c), for which we first trained an accurate Machine Learning Potential (MLP) based on an underlying set of high-temperature DFT Umbrella Sampling (US) simulations. The use of enhanced-sampling simulations is essential to explore in an efficient way the less-probable highly energetic regions of the PES, which are typically associated with reactive events. The underlying DFT simulations at finite temperatures serve as input to train a deep neural network MLP (Fig. 2). Once an accurate MLP is constructed, an enormous computational speedup can be achieved, which allowed to: (i) compute the Free Energy Surfaces (FESs) of all possible hoppings around an isolated Al defect in the temperature range 273–873 K with a large number of umbrellas and long simulation times to obtain well-converged results, (ii) explicitly include NQEs through the PIMD approach, (iii) derive accurate kinetic

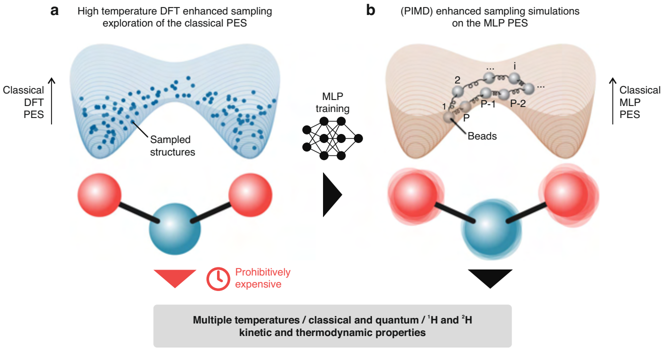


Fig. 2 | Kinetic and thermodynamic properties with and without NQEs over a wide range of temperatures can be computed using an MLP trained on high-temperature enhanced-sampling DFT simulations. The pictures show a simplified schematic representation of the PES experienced by the proton when hopping between two oxygens around the Al site (red and blue spheres, respectively), sampled with DFT simulations (**a**) and subsequently learned with the MLP (**b**). In **b** a schematic view of the hydrogen ring polymer with P beads running on the classical MLP PES is shown.

constants beyond the Transition State Theory (TST) approximation, taking barrier recrossing into account via the reactive flux formalism[27] and (iv) perform an additional set of simulations with deuterium instead of protium to explicitly compute the Kinetic Isotope Effect (KIE) on the reaction.

We show that even at catalytically relevant temperatures (>400 K) NQEs may still be important to consider when computing reaction kinetics and their relevance is not restricted to the absolute low temperature regime. While the work performed here is illustrative for the most basic proton hopping reaction in zeolites, it provides the means to routinely include NQEs and explicitly calculate KIEs when studying any proton-transfer event in heterogeneous catalysis.

Results

Construction of a reactive MLP with DFT accuracy

To train an accurate MLP, a sufficiently large set of DFT datapoints is required, which should cover the relevant regions of the reaction phase space. To this end, high-temperature (873 K) DFT US simulations were performed on a CHA conventional cell containing 36 T Si atoms, where 1 silicon is replaced by Al to give a final Si/Al ratio of 35. The temperature choice of 873 K is arbitrary but, in general, on the higher end of typical zeolite-catalyzed processes[1]. In the CHA topology, all T atoms are equivalent. However, the four O atoms in the first coordination sphere of the Al defect are not (Fig. 1c). This leads to 6 distinct hopping paths, which are all considered in this work. To assess whether any path could be significantly disfavored, activation free energies were initially screened with static calculations. The results suggest that all 6 possible hopping paths have relatively similar activation free energies (within ~20 kJ mol^{-1}) and no single one is strongly (dis)favored (Supplementary Note 2). Therefore, 6 separate DFT US simulations at 873 K were performed to sample all the possible hoppings. A difference in coordination numbers between the proton and the two oxygen atoms involved in the hopping was used as main collective variable to bias the system (see "Methods" section and Supplementary Note 3). One-dimensional umbrellas were used to sample the reaction path and, if needed, additional two-dimensional umbrellas were added to improve the sampling of scarcely visited regions of the phase space (more details are reported in Supplementary Note 3.2). A full overview of the DFT US results is reported in Supplementary Note 4.

Energies and forces were extracted every 5 fs from the DFT US trajectories, yielding a total of ~1,200,000 structures which were used to train an MLP with the SchNetPack package (see "Methods" section and Fig. 2)[28,29]. Performing MD simulations with the MLP provides a dramatic speedup in computational time, going from ~8.3 s/step on 56 Xeon E5-2680v4 CPUs@2.4 GHz cores to ~0.01 s/step on a single NVIDIA Volta V100 GPU. As part of the MLP validation, well-converged 873 K DFT FESs were generated to compare them with the MLP-derived ones within a reasonably small uncertainty. To this end, about 50 ps of simulation time was required for each DFT umbrella. Considering that 19 umbrellas are needed to sample each of the 6 hoppings, this was a computationally demanding task. On the other hand, it also provided us with a very large number of DFT datapoints, hence the large number of structures used to train the MLP. With the acquired knowledge that a mean absolute error on the force of about 40 meV Å^{-1} is sufficient to obtain very accurate FESs (vide infra), we also tested the performance of newer and more data efficient equivariant neural networks[30], where preliminary results indicate that a few hundred fs per umbrella are sufficient to achieve converged results (Supplementary Note 10), providing an enormous computational saving in the DFT data generation.

To further validate the trained MLP, we also tested whether it could reproduce FESs at lower temperatures than the training one. To this end, three additional sets of DFT US simulations were performed. The 2–3 hopping was tested at 573 and 273 K, while the 1–4 hopping was tested at 273 K. With this choice, the hoppings with the smallest activation energies are tested and all four oxygens are considered at the lowest temperature. For the sake of clarity, a detailed comparison between the MLP and DFT results is presented in Supplementary Note 6.1, while here only the 2–3 hopping is discussed in detail. As shown in Fig. 3, the DFT and MLP FESs exhibit an almost perfect overlap, with most variations contained well within the error bars. The free energy barrier exhibits a clear increase with temperature, which is in line with a rigid transition state associated with a negative entropy variation. It must be pointed out that, thanks to the large computational speedup enabled by the MLP, longer simulation times (100 ps vs. 50 ps per umbrella) and a larger number of umbrellas (39 vs. 19 per hopping) were easily achievable. This led to a vastly improved sampling of the reaction PES, thereby obtaining much better converged FESs. Moreover, all MLP simulations were repeated three separate times starting from different initial velocities and the associated results and uncertainties were obtained by averaging over these three independent runs. Initially, the 573 K DFT profile presented a moderate spike in the transition state region, which was not present in the MLP profile. Therefore, 2 additional umbrellas were added in the proximity of the transition state and an additional 40 ps of simulation was performed in every umbrella, for a total of 90 ps. The final DFT profile

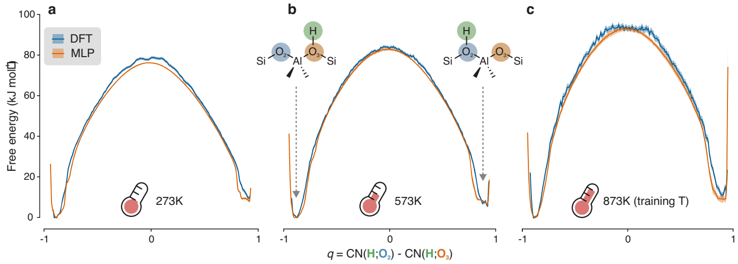


Fig. 3 | The trained MLP can reproduce DFT FESs with high accuracy. The MLP and DFT FESs (in orange and blue, respectively) of the 2–3 hopping at 273 (**a**), 573 (**b**) and 873 K (**c**) are almost perfectly superimposable, even though the MLP model is only trained on snapshots at the highest temperature. A comparison with the other available FESs is shown in Supplementary Fig. 10. For a more detailed definition of the collective variable (q) used to build the FESs, the interested reader is referred to the "Methods" section. Uncertainties (shaded areas around the FESs) on the MLP FESs are obtained by averaging over three independent runs (Supplementary Note 6.1) while for DFT they are estimated according to the procedure summarized in Supplementary Note 3.1.3. Source data are provided as Source Data File.

reaches almost perfect agreement with the MLP one, highlighting how the (small) differences between MLP and DFT FESs are almost certainly caused by sampling issues rather than by significant deviations in the underlying PES. The results show that (i) the trained MLP is effectively capable of encoding chemical reactivity and (ii) high accuracy on the computed FESs is retained also for temperatures lower than the training one, offering thermodynamic transferability in terms of operating conditions.

While directly superimposing FESs provides an intuitive visual means of comparison, the FES itself is not experimentally measurable. The final macroscopic quantity of interest is the kinetic constant of the reaction, which does not depend on the choice of the collective variable used to represent the FES[31,32]. By means of classical (TST), the forward and backward kinetic constants for the 6 high-temperature hoppings, the 2–3 hopping at 573 K and the 1–4 and 2–3 hoppings at 273 K were retrieved (see "Methods" section). Fig 4 reports a graphical comparison between the DFT and MLP rates, where the computed kinetic constants are converted to a corresponding phenomenological barrier using Eyring's equation[32]. The sole purpose of the latter is to provide a more tangible equivalent to the kinetic constant, without comparing values that can span multiple orders of magnitude (more details are provided in the "Methods" section). None of the computed barriers differ more than ~5 kJ mol^{-1} and, for most of the hoppings, the MLP values lie within the error bars of the DFT ones. These results indicate that the MLP accurately reproduces the DFT PES underlying the proton hopping reaction in H-CHA and can therefore be used to compute reaction rates at any temperature of interest and to explicitly introduce NQEs through the PIMD approach (Fig. 2).

Full characterization of the hopping kinetics

Having validated the MLP to faithfully reproduce the proton hopping FESs over a broad temperature range (273-873 K), additional US simulations were performed to retrieve the full reaction kinetics considering all hopping paths. Moreover, NQEs can be systematically included in the reaction investigation as the PIMD formalism becomes accessible, thanks to the large computational efficiency of the MLP. To obtain well-converged FESs, at least 16 system replicas (also referred to as beads) are required in the ring polymer (Supplementary Note 7.1). A graphical visualization of the spread of the beads around the transition state region compared with the classical case is shown in Fig. 5, where it becomes clear that quite some uncertainty is present on the proton's position compared to the classical deterministic trajectories. A full overview of the classical and quantum FESs is reported in Supplementary Notes 6.1 and 7.3, respectively. Introducing NQEs leads to a general decrease of the free energy barriers compared to the case where nuclei are treated classically. This effect tends to lessen with increasing temperatures, in accordance with the expected convergence between the quantum and classical behavior for high temperatures.

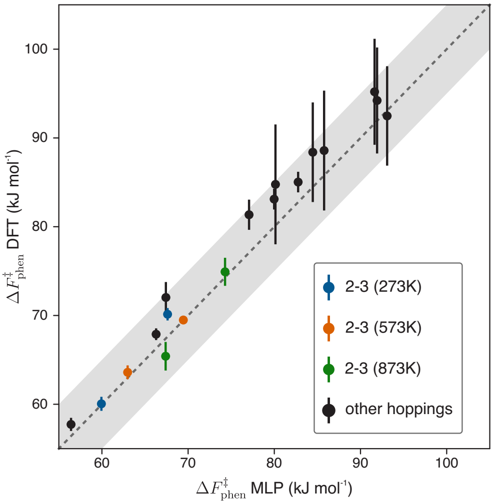


Fig. 4 | The DFT and MLP phenomenological barriers are in excellent agreement. The available DFT phenomenological barriers (12 forward and backward barriers for the 873 K hoppings, the forward and backward barriers of the 2–3 hopping at 573 K and the forward and backward barrier of the 1–4 and 2–3 hoppings at 273 K) are all within about ±5 kJ mol^{-1} from the MLP ones, as highlighted by the gray shaded area. The phenomenological barriers of the 2–3 hopping at various temperatures, corresponding with the free energy profiles of Fig. 3, are highlighted in different colors (273 K in blue, 573 K in orange and 873 K in green). Uncertainties on the MLP barriers are obtained by averaging over three independent runs (although their magnitude is so low that they are barely visible in the figure) while for DFT they are estimated according to the procedure summarized in Supplementary Note 3.3. Source data are provided as Source Data File.

Not only does the MLP allow to include in a reliable—yet computationally feasible—way NQEs, but it also allows to determine reaction kinetics beyond classical TST and explicitly include barrier recrossing through the reactive flux formalism (obtaining the true

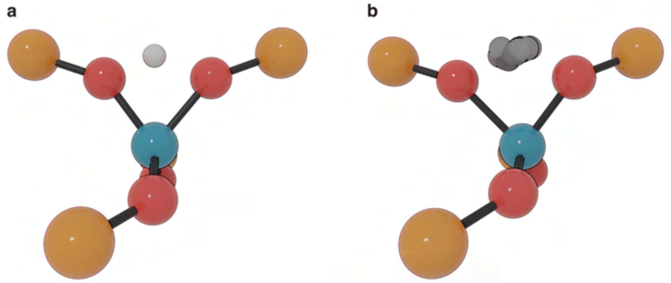


Fig. 5 | The use of PIMD leads to a significant spread in the proton location. These two snapshots, arbitrarily extracted from the transition state umbrella of the 2–3 hopping, highlight how the system beads in PIMD (**b**) can be spread quite significantly in space with respect to the classical case (**a**). For the sake of clarity, only the H-SSZ-13 atoms up to the second coordination sphere around the Al site are shown and in **b** a superposition of all beads is only present for the proton. Si is depicted in yellow, O in red, Al in blue and H in white.

kinetic constant of the reaction)[27,33], which is in its turn part of the Bennett-Chandler reaction rate theory[34,35]. In this approach, multiple unbiased simulations (5000 in this case) are initialized atop the transition state and monitored through time, to retrieve whether they end up in the product basin or whether they recross the barrier towards the reactant basin (see "Methods" section). This approach is most appropriate when NQEs are included, as quantum TST approximations such as ring polymer molecular dynamics (RPMD) TST do not yield a strict upper bound for the quantum rate (more details can be found in Supplementary Note 7.4)[36]. Overall, the MLP US simulations allowed to compute three different kinetic constants for all hopping paths and all temperatures: a classical TST-based one (k^{c}_{TST}), derived from classical MD and the TST approximation, a classical Bennett-Chandler one (k^{c}_{BC}), where barrier recrossing is now explicitly taken into account and, finally, a quantum Bennett-Chandler one (k^{q}_{BC}), analogous to k^{c}_{BC} but derived from the RPMD simulations and thus including NQEs. Remark that the amount of data used to obtain them is well beyond the reach of any pure ab initio methodology where all energy and force evaluations are performed at the DFT level. Even when excluding the thousands of short trajectories required to obtain well-converged k_{BC} values, computing the quantum FESs requires 42 sets of US simulations (6 hopping paths at 7 different temperatures), each consisting of 39 umbrellas simulated with 16 parallel beads—for a total of more than 0.6 µs of simulation time. Such simulation lengths are clearly beyond the limit of any DFT-based MD simulation.

While demonstrating the impact of NQEs on the reaction rate is important to highlight the cases in which NQEs cannot be neglected and should thus be accounted for computationally, the resulting 'quantum speedup' is not experimentally measurable as NQEs are an intrinsic part of nature. What is often measured experimentally, on the other hand, is the KIE—namely the change in rate when the hydrogen atoms in the system are substituted with deuterium (other isotopic substitutions are of course also possible[37]). Interestingly, the MLP trained on ^{1}H simulations can be directly used for KIE calculations, as the reaction PES does not depend on the atomic masses but only on their charge. An additional set of PIMD simulations was therefore performed at 273, 573 and 873 K with the BAS mass set to 2 a.m.u. Given the linear behavior of $\ln(k^{q}_{BC})$ over the whole temperature range (vide infra, Fig. 7), the intermediate temperatures were no longer considered. A full overview of the simulations' results is reported in Supplementary Note 7.5. The error on the MLP forces with respect to DFT remains basically unaffected by the change in the hydrogen mass, confirming that both simulations sample analogous PES regions (Supplementary Fig. 20). The reactive flux kinetic constant for the PIMD simulations with deuterium will be indicated with $k^{q}_{BC}(^{2}H)$, while for protium simulations the isotope label will be omitted.

Using the computed kinetic constants, the equilibrium coverage of the 4 oxygen sites (θ_i, $i = 1 - 4$) was determined as a function of temperature (more details can be found in Supplementary Note 8). The results for ^{1}H are shown in Fig. 6a–c. When considering the k^{c}_{TST} and k^{c}_{BC} kinetic constants, similar equilibrium populations are obtained within the limits of uncertainty (Fig. 6a, b), which is a consequence of the similar recrossing rate between the forward and backward barriers. In general, O_1 and O_3 are the most populated sites at any temperature, followed by O_2 and O_4. In the classical case, O_3 has the largest population up to 373 K, while at higher temperatures its population becomes nearly identical to O_1. Significantly different results are obtained when the quantum kinetic constants (k^{q}_{BC}, Fig. 6c) are considered, where θ_3 remains significantly larger than θ_1 at all temperatures. A similar trend is obtained for the PIMD simulations with deuterium (Supplementary Fig. 19). When the proton is on O_3, it finds itself oriented towards the center of the 6 T atoms ring (Fig. 6d) and can, therefore, interact with the oxygens on the opposite side. To understand more profoundly the impact of these intra-framework interactions, we performed a 273 K classical MD and PIMD simulations of the zeolite with the proton located on O_3. We then analyzed the radial distribution functions (RDFs) of the proton with the 6 oxygens sharing the Si and Al with O_3 ('adjacent' in Fig. 6e) and all the other oxygens in the unit cell ('others'). We found that—as expected— the BAS lies within 2–3 Å from the oxygens on the other side of the 6-Si ring and will therefore interact with them. By comparing the classical and quantum RDFs (Fig. 6e), it can be seen how the maximum in the RDF H-O (others) occurs at slightly shorter distances in the quantum case and, moreover, shorter distances—in the order of ~2 Å—are explored more often. Based on these findings, it appears that when the quantum nature of the hydrogen nucleus is considered the weak interaction between the non-adjacent framework oxygens and the proton becomes stronger and, as a result, O_3 becomes further stabilized with an increase in θ_3. No other site interacts with other framework oxygens within 3 Å (Supplementary Fig. 18). Previous reports in the literature, based on geometrical considerations concerning the crystallographic zeolite unit cell, suggested that none of the four BAS locations are suited to form H-bonds with other oxygens in the framework[38]. This no longer seems to be the case when temperature effects and NQEs are explicitly taken into account. The number of zeolite frameworks presenting this type of intra-framework interaction could thus be higher than previously thought[38] at realistic operando conditions.

Once all the equilibrium coverages as a function of temperature are known, the overall hopping rate can be computed using the

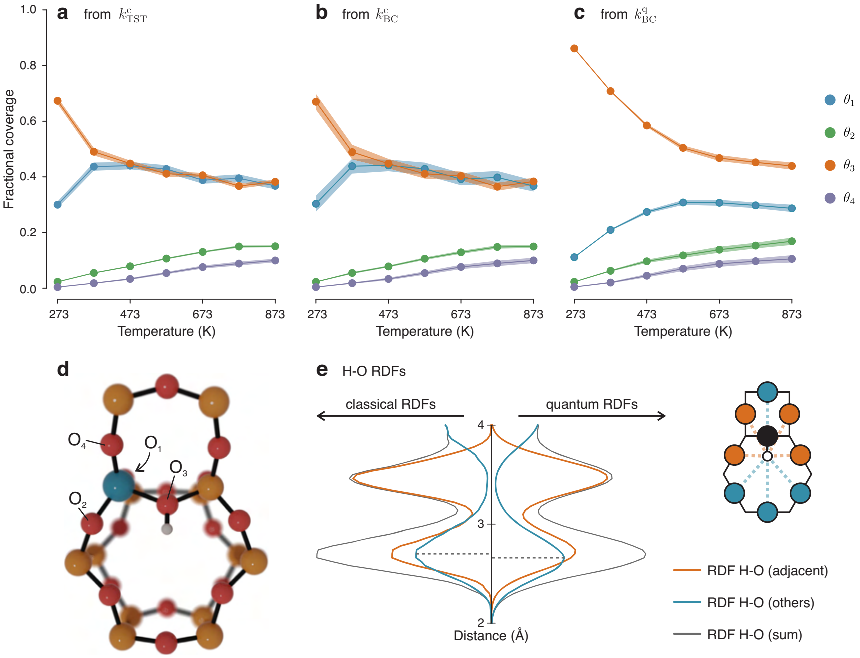


Fig. 6 | Including NQEs can change the computed equilibrium proton coverages. **a**–**c** Equilibrium coverages (θ_i) of the 4 oxygen sites around the Al defect as function of temperature, computed from the k^{c}_{TST}, k^{c}_{BC} and k^{q}_{BC} kinetic constants. **d** Portion of the H-CHA framework as seen along the c cell vector, showing the optimized structure of the BAS on O_3. Si is depicted in yellow, O in red, Al in blue and H in white. **e** RDFs between the BAS lying on O_3 and the 6 oxygens sharing the same Si and Al as O_3 (orange line) and the BAS and all other oxygens in the unit cell (blue line), from classical and NQE simulations. The schematic drawing is seen from the same perspective as d and highlights some of the oxygens belonging to 'adjacent' and 'others'. Source data are provided as Source Data File.

formula:

$$r = \sum_{i=1}^{4} \sum_{j \neq i}^{4} k_{ij} \theta_i, \tag{1}$$

where k_{ij} is the kinetic constant of the hopping from O_i to O_j and θ_i the coverage of O_i. From this, an Arrhenius plot for the hopping rate as a function of the temperature is computed (Fig. 7a), whose activation energy should be comparable with experiment. First, we analyzed in how far each of the hoppings is contributing to the overall rate. In all cases, only two hopping paths dominate the rate kinetics (Fig. 7b), namely the 1 ↔ 4 and 2 ↔ 3 paths, as one could expect based on their low free energy barriers. Note that the forward and backward rates have similar contributions, as a higher free energy of the minimum corresponds to both a lower coverage and a lower free energy barrier to hop towards a stabler minimum. These two factors tend to cancel each other when computing $k_{ij}\theta_i$. Minor contributions are given by the 1 ↔ 2 and 2 ↔ 4 paths, while the remaining two paths (1 ↔ 3 and 3 ↔ 4) only have noticeable contributions at the highest temperatures. In the deuterium case, the 2 ↔ 3 path becomes even more dominant at the expenses of 1 ↔ 4 as it appears that the transition state energy is not shifted consistently by the isotope substitution (Supplementary Fig. 18).

By considering the slope of the best fit lines in the Arrhenius plots (Fig. 7a) it is possible to retrieve an effective activation energy for the proton hopping reaction. The k^{c}_{TST} results yield an activation energy of 67.1 kJ mol^{-1}. Going beyond the TST approximation and explicitly including recrossing (k^{c}_{BC}) does not significantly change the results, with a consistent—but almost negligible—decrease in the rate across the whole temperature range. When NQEs are included (k^{q}_{BC}), in contrast, the activation energy decreases with about 11 kJ mol^{-1} due to the possibility of the proton to tunnel through the potential energy barriers. When analyzing the rates related to a specific hopping (Supplementary Fig. 17), it was noticed that this effect is not constant, and becomes more prominent in the hopping paths with a more sharply peaked FES around the transition state region (Supplementary Figs. 15 and 16). This is because a narrower barrier increases the probability of tunneling, which in practice means that the beads of the ring polymer are easily located on both sides of the potential energy barrier experiencing on average a lower free energy. These results show that the impact of NQEs is not systematic in nature and can therefore be challenging to capture with ad hoc corrections. Indeed, previous investigations which included NQEs through an a posteriori tunneling correction suggested that above room temperature no significant effect should be observed[12]. Our results, on the other hand, show that the reaction proceeds 65 times faster at 273 K if NQEs are included and, even at 373 K, a 16-fold increase in the computed rate is still present (Fig. 7c). At 473 K, the reaction remains 7 times faster while the speedup, as expected, tends to become negligible at higher temperatures. It appears therefore that for zeolite-catalyzed processes conducted at milder conditions, among which the ones related to biomass conversion are a predominant example[2], NQEs might be non-negligible when computing the kinetics of proton-transfer steps. A few examples where these effects might be important are the aqueous cyclohexanol

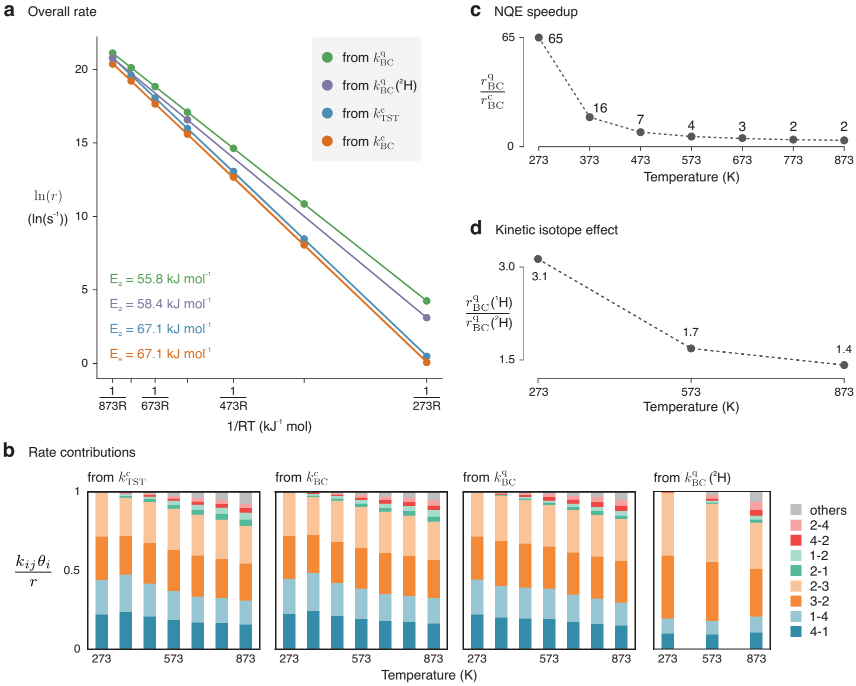


Fig. 7 | NQEs have a significant effect on the computed hopping rate. **a** Arrhenius plot showing the overall hopping rate as function of 1/RT. **b** Contributions to the overall rate from the different hopping paths, as function of temperature and methodology used. **c** Speedup in the rate due to the inclusion of the NQEs. **d** KIE as function of reaction temperature. Source data are provided as Source Data File.

dehydration in H-Beta zeolite at temperatures of about 400 K[39] and the oxidation of methane to methanol in copper-loaded zeolites carried out at ~473 K[40]. These two cases serve as examples of the relevant application area and conditions influenced by NQEs in zeolites. However, it is clear that many more cases could be affected by the inclusion of NQEs.

When protium is substituted by deuterium ($k^{q}_{BC}(^2H)$) the activation energy becomes 58.4 kJ mol^{-1}, yielding a small increase of 2.6 kJ mol^{-1}. This is in line with the modest magnitude of the predicted KIE (given by the ratio between the total rate with protium and with deuterium, Fig. 7d), which equals 3.1 at 273 K and decreases to 1.4 at 873 K, in line with standard Bigeleisen-Mayer KIE theory[41,42]. It can be noted how the deuterium rates remain significantly larger than the classical k^{c}_{BC} ones, indicating that the increase in mass is not sufficient to fully suppress the quantum behavior. Unfortunately, there has not been any experimental attempt so far to measure the KIE for proton hopping in zeolites. Theoretically, an early investigation based on static cluster calculations was performed by Fermann and Auerbach[43]. According to their semiclassical TST model, the differences between ^{1}H and ^{2}H are limited above room temperature, in line with our results.

So far, the study focused on H-CHA, which has a single indistinguishable T-site and a small unit cell. To broaden our scope, it is important to assess the MLP capability of describing other zeolite frameworks on which the MLP was not trained. To this end, the transferability of the MLP to other zeolite topologies was investigated. More specifically, we selected five all-silica frameworks from the international zeolite (IZA) database[44] that are of interest for catalytic applications[3] (AFX, CHA, FER, MFI and MOR) and performed a 100 ps NVT DFT MD simulation using the crystallographic unit cell parameters reported in the IZA database (see Supplementary Note 9). The CHA topology was included as a control system, to ensure that the MLP is robust with respect to changes in the unit cell volume. None of the MLP simulations presented obvious instabilities and the error on the forces is not excessive, even for frameworks that do not share any secondary building unit with CHA, varying between 196 meV Å^{-1} for MOR and 258 meV Å^{-1} for MFI. The quality of the zeolite trajectories, monitored through the Si–O and Si–Si RDFs, remains reasonably good with only small long-range differences for MFI (Supplementary Fig. 19). Testing the proton hopping reactivity in a systematic way for more frameworks would require a further set of expensive ab initio US simulations and, therefore, is outside the scope of this work. The results obtained on the all-silica frameworks, nonetheless, still indicate that the MLP can capture to a large extent the chemistry of Si–O–Si bonds and, therefore, we expect that not many additional DFT simulations would be needed to retrain it and extend its accuracy to new zeolite frameworks, for instance building on the transferable MLP for siliceous frameworks by Erlebach et al.[45] towards aluminum-containing zeolites of catalytic interest.

Discussion

Proton-transfer reactions are of primordial importance within zeolite catalysis. Thus far, it was unclear in how far NQEs affect the barriers and rates of proton hopping processes at realistic operating conditions, as their explicit inclusion through PIMD was prohibitively expensive if the underlying classical PES is evaluated at a DFT level of theory. Herein, we showed that a reactive MLP can be trained based on underlying high-temperature (873 K) US simulations at the DFT level, that provides kinetic results with a similar accuracy as the underlying DFT data. However, thanks to the enormous computational speedup gained by describing the PES based on the MLP compared to the original DFT energy and force calculations, the MLP can be used to perform virtually any type of simulation that relies on the classical PES of the considered reaction(s) over a broad range of temperatures. The proposed methodology thus not only succeeds in reproducing the underlying DFT simulations but comes with a series of advantages that were so far unreachable due to the prohibitively excessive computational cost.

First, the convergence of the free energy surfaces obtained from enhanced-sampling techniques can be improved by using many more umbrellas and by simulating for a longer time. Secondly, PIMD can be employed to explicitly account for the quantum nature of the nuclei in the system. While the inclusion of NQEs through MLPs has already been proposed in the literature[22,46,47], the application of PIMD/MLP to an activated event in heterogenous catalysis was still unexplored. We remark that for simulations at cryogenic temperatures the number of beads required to achieve converged results could become very large even for the MLP. This problem can be mitigated by coupling the MLP simulations with path integral acceleration techniques[20]. Thirdly, it also becomes possible to go beyond the TST approximation and explicitly include barrier recrossing via the reactive flux formalism, thereby obtaining the true kinetic constant of the reaction. Because of the thousands of short MD trajectories that have to be initialized atop the transition state, this type of calculation was so far too expensive to be performed at a DFT level of theory. The more efficient methodology for describing the forces and energies may also open the window to use methods like transition path sampling within the field of zeolite catalysis, which were thus far not truly accessible due to the large number of paths that needs to be sampled at the DFT level[48]. Finally, KIEs can be explicitly computed if the PIMD simulations are performed with different nuclear masses, as this does not affect the underlying PES learned by the MLP.

Our results show that the expected Arrhenius activation energy for the hopping process, considering all six hoppings and the coverages of the four oxygen sites, is 67.1 kJ mol^{-1} in the absence of NQEs, whereas including the quantum nature of the proton brings the activation energy down to 55.8 kJ mol^{-1}. When quantitively comparing this activation energy to experimental results, it is important to note that this study makes use of the revPBE-D3 level of theory, which is known to underestimate the activation energies of chemical reactions[16,49,50]. In this sense, our barriers will present a lower boundary for the chemically accurate activation energy. Because of the large improvement in data efficiency of newer MLP architectures (Section S10 of the Supplementary Information), we believe that training an accurate model based on a more expensive albeit more reliable exchange-correlation functionals should become feasible. The computed activation energy remains relatively higher than the experimentally available ones. The most likely source of discrepancy lies in the perfect crystalline nature of the adopted zeolite model. The presence of residual water molecules, defects (EFAL species, for instance) and an heterogeneous aluminum distribution are basically unavoidable at the macroscale and all these factors are known to potentially affect the behavior of protons in zeolites[17,19,51]. According to the simulations, a primary KIE of about 3 is expected at 273 K but no experimental evidence is available thus far to corroborate this result.

This proof-of-concept study presents a general scheme to obtain MLP models that can simulate proton hoppings and activated processes in zeolite catalysis with improved realism. The proposed methodology is, in principle, extendible to additional reactions and reactive environments, making it a valuable tool for studying a wide range of catalytic phenomena[52].

Methods

Umbrella sampling simulations

The hopping of the H-CHA BAS between the oxygens in the first coordination sphere of the Al defect was studied by means of umbrella sampling simulations[53,54]. In this approach, quadratic bias potentials (the 'umbrellas') are placed along a certain collective variable (q) which should smoothly vary between reactants and products. The bias has the form $V_i(q)=1/2K_i(q-q_{0,i})^2$, where K_i is the force constant of the ith umbrella and $q_{0,i}$ its center. An MD simulation is then performed within each umbrella. To study the proton hopping, the chosen collective variable is a difference of coordination numbers (CNs) between the BAS and the two oxygens involved in the hopping:

$$q=\mathrm{CN}(\mathrm{O}_i;\mathrm{H})-\mathrm{CN}(\mathrm{O}_j;\mathrm{H})=\frac{1-\left(\frac{r_{\mathrm{O}_i\mathrm{H}}}{r_0}\right)^N}{1-\left(\frac{r_{\mathrm{O}_i\mathrm{H}}}{r_0}\right)^{2N}}-\frac{1-\left(\frac{r_{\mathrm{O}_j\mathrm{H}}}{r_0}\right)^N}{1-\left(\frac{r_{\mathrm{O}_j\mathrm{H}}}{r_0}\right)^{2N}} \qquad (2)$$

The specific values of the N and r_0 parameters were adapted based on the reaction conditions, more information can be found in Supplementary Note 5.2. The bias potential was applied using PLUMED[55,56] and the final statistical analysis of the data was performed with our in-house developed ThermoLIB library[57]. For some of the hoppings, additional wall potentials were required to prevent undesired side reactions; further details are reported in Supplementary Note 3.1.4

DFT molecular dynamics

To perform the DFT MD simulations, the CP2K software package (version 7.1)[58,59] was employed to compute energies and forces at a revPBE-D3/TZVP[60–62] level of theory. Because of the mixed plane waves −atom-centered orbitals approach[63] used by CP2K, the plane waves energy cutoff was set to 350 Ry and GTH pseudopotentials[64] were used to smooth the electron density in the proximity of the nuclei. A significant dependency of the forces on the plane waves cutoff was found, but this was shown to have a negligible impact on the final FESs when much higher settings are used (Supplementary Note 3.1.5). The time step for the integration of the equations of motion was set to 0.5 fs. After equilibration of the unit cell (Supplementary Note 3.3.1), production runs were performed in the NVT ensemble using a Nosé-Hoover thermostat with a chain consisting of five beads[65,66] to control the temperature and a time constant of 334 fs (100 cm^{-1}).

MLP training and usage

A SchNet MLP was trained with the SchNetPack package on the DFT energies and forces which were extracted every 5 fs from the DFT US simulations at 873 K[28,29]. First, the energies and forces were unbiased by subtracting the bias potential applied in the US simulations with PLUMED[55,56]. The unbiased DFT datapoints were randomly divided in a training and validation set with a 80:20 ratio. Subsequently, the MLP was trained with a cutoff of 6 Å, 128 features, 50 gaussians and 6 interaction blocks. The resulting MAE on the validation set is 41.9 meV/Å. More details on the training are provided in Supplementary Note 5.1. Classical unbiased and US simulations with the trained MLP were performed with our in-house code YAFF[67] using a time step of 0.5 fs and a Nosé-Hoover thermostat with three beads for temperature control[65,66]. PIMD simulations were performed with the i-PI driver[68] using a time step of 0.25 fs and a PILE thermostat[69] with a time constant of 100 fs for temperature control. Because of the harmonic repulsion between the beads, some of them might explore regions of the phase space that are not necessarily well-sampled in classical DFT US. Therefore, we also performed an extra DFT PIMD US simulation for the 2–3 hopping (Supplementary Note 7.2) and the resulting FES shows an excellent agreement with the MLP one. It is important to remark that this agreement is very likely not generalizable to other systems or reactions and should always be tested appropriately[70]. In both the classical and PIMD US simulations, PLUMED was used to apply the bias.

Kinetic rate constant calculation

The plain activation free energy obtained from a FES is largely dependent on the choice of collective variable[31,32]. To remove such dependency, it is necessary to move towards a more general macroscopic property of the process under study, namely the kinetic rate constant. In the Bennett-Chandler approach to transition state theory[34,35], the rate constant of a reaction can be written as[71]:

$$k_{\mathrm{BC}}(t) = \langle \dot{q}(0)\theta(q(t) - q^*)\rangle_{q(0)=q^*} \frac{e^{-\beta F(q^*)}}{\int_{-\infty}^{q^*} e^{-\beta F(q)} \mathrm{d}q}, \quad (3)$$

where the first term is the ensemble average of the time derivatives of q for trajectories that, starting atop the transition state ($q(0) = q^*$), end up in the product basin (as imposed by the Heavyside function $\theta(q(t) - q^*)$). With the MLP, it is possible to explicitly compute the first term by performing a large number of unbiased MD simulations (5000 in this case) starting on the transition state and monitor how many of them effectively end up in the product basin[27,33]. The true rate constant is, in principle, given by $k_{\mathrm{BC}} = \lim_{t \to +\infty} k_{\mathrm{BC}}(t)$. Luckily, its value quickly reaches a plateau and 50 fs of simulation were sufficient to obtain well-converged results (Supplementary Note 6.2). The rate constant calculated in this manner is referred to as the Bennett-Chandler one (k_{BC}).

In general, this approach is too expensive, especially for the DFT case, so that only the approximate transition state theory constant ($k_{\mathrm{TST}} = \lim_{t \to 0^+} k_{\mathrm{BC}}(t)$) can be computed from the US trajectories, thereby avoiding the need for additional simulations. While k_{TST} represents an upper limit of the true kinetic constant, assuming a recrossing probability equal to zero, it can be used to compare the DFT and MLP results. Further details are reported in Supplementary Note 3.3. To calculate the quantum rate constants, taking NQEs into account, the approximate technique of RPMD was used (see Supplementary Note 7.4)[72]. Although this approximation can only capture short-time quantum effects, it has been shown to yield good quantum rates in comparison with other approximations[73] or quantum mechanical calculations[74].

As the kinetic constant values can span several orders of magnitude, we often make use of Eyring's equation to convert them into phenomenological barriers, which encode the same information while being – in our opinion – more tangible than a reaction rate.

$$\Delta F^{\ddagger}_{\mathrm{phen}} = -\frac{1}{\beta}\ln(\beta h k) \quad (4)$$

Data availability

The complete training set, examples of input files, processing scripts and the trained MLP have been deposited in the Zenodo database (https://zenodo.org/record/7267913#.Y2U8tHbMK3A). Any additional data is available from the authors upon request. An extended discussion of the results can be found in the Supplementary Information. Source data are provided with this paper.

Code availability

CP2K (https://github.com/cp2k/cp2k), PLUMED (https://github.com/plumed/plumed2), SchNetPack (https://github.com/atomistic-machine-learning/schnetpack) and YAFF (https://github.com/molmod/yaff) are all open source and freely available at the provided links. ThermoLIB is available upon request at https://molmod.ugent.be/software/thermolib.

References

1. Weissermel, K. & Arpe, H.-J. *Industrial Organic Chemistry* (Wiley-VCH Verlag GmbH, Weinheim, Germany, 2003).
2. Ennaert, T. et al. Potential and challenges of zeolite chemistry in the catalytic conversion of biomass. *Chem. Soc. Rev.* **45**, 584–611 (2016).
3. Van Speybroeck, V. et al. Advances in theory and their application within the field of zeolite chemistry. *Chem. Soc. Rev.* **44**, 7044–7111 (2015).
4. Baba, T., Inoue, Y., Shoji, H., Uematsu, T. & Ono, Y. Temperature-dependent lineshape of 1H magic-angle spinning nuclear magnetic resonance spectra of acidic hydroxyl groups in zeolites. *Microporous Mater.* **3**, 647–655 (1995).
5. Sarv, P., Tuherm, T., Lippmaa, E., Keskinen, K. & Root, A. Mobility of the acidic proton in Brønsted sites of H-Y, H-mordenite, and H-ZSM-5 zeolites, studied by high-temperature 1H MAS NMR. *J. Phys. Chem.* **99**, 13763–13768 (1995).
6. Baba, T., Komatsu, N. & Ono, Y. Mobility of the acidic protons in H-ZSM-5 as studied by variable temperature ^{1}H MAS NMR. *J. Phys. Chem. B* **102**, 804–808 (1998).
7. Franke, M. E. & Simon, U. Proton mobility in H-ZSM5 studied by impedance spectroscopy. *Solid State Ion.* **118**, 311–316 (1999).
8. Osuga, R., Yokoi, T., Doitomi, K., Hirao, H. & Kondo, J. N. Infrared investigation of dynamic behavior of Brønsted acid sites on zeolites at high temperatures. *J. Phys. Chem. C.* **121**, 25411–25420 (2017).
9. Fermann, J. T., Blanco, C. & Auerbach, S. Modeling proton mobility in acidic zeolite clusters. I. Convergence of transition state parameters from quantum chemistry. *J. Chem. Phys.* **112**, 6779–6786 (2000).
10. Ryder, J. A., Chakraborty, A. K. & Bell, A. T. Density functional theory study of proton mobility in zeolites: Proton migration and hydrogen exchange in ZSM-5. *J. Phys. Chem. B* **104**, 6998–7011 (2000).
11. Sierka, M. & Sauer, J. Finding transition structures in extended systems: a strategy based on a combined quantum mechanics-empirical valence bond approach. *J. Chem. Phys.* **112**, 6983–6996 (2000).
12. Sierka, M. & Sauer, J. Proton mobility in chabazite, faujasite, and ZSM-5 zeolite catalysts. Comparison based on ab initio calculations. *J. Phys. Chem. B* **105**, 1603–1613 (2001).
13. Tuma, C. & Sauer, J. A hybrid MP2/planewave-DFT scheme for large chemical systems: proton jumps in zeolites. *Chem. Phys. Lett.* **387**, 388–394 (2004).
14. Wang, Y. et al. Density functional theory study of proton hopping in MCM-22 zeolite. *Chem. Phys. Lett.* **388**, 363–366 (2004).
15. Franke, M. E., Sierka, M., Simon, U. & Sauer, J. Translational proton motion in zeolite H-ZSM-5. Energy barriers and jump rates from DFT calculations. *Phys. Chem. Chem. Phys.* **4**, 5207–5216 (2002).
16. Bučko, T., Gešvandtnerová, M. & Rocca, D. Ab initio calculations of free energy of activation at multiple electronic structure levels made affordable: an effective combination of perturbation theory and machine learning. *J. Chem. Theory Comput.* **16**, 6049–6060 (2020).
17. Kester, P. M., Crum, J. T., Li, S., Schneider, W. F. & Gounder, R. Effects of Brønsted acid site proximity in chabazite zeolites on OH infrared spectra and protolytic propane cracking kinetics. *J. Catal.* **395**, 210–226 (2021).
18. Huo, H., Peng, L. & Grey, C. P. Low temperature 1H MAS NMR spectroscopy studies of proton motion in zeolite HZSM-5. *J. Phys. Chem. C.* **113**, 8211–8219 (2009).
19. Li, G. & Pidko, E. A. The nature and catalytic function of cation sites in zeolites: a computational perspective. *ChemCatChem* **11**, 134–156 (2019).
20. Markland, T. E. & Ceriotti, M. Nuclear quantum effects enter the mainstream. *Nat. Rev. Chem.* **2**, 0109 (2018).
21. Li, C. & Voth, G. A. Using machine learning to greatly accelerate path integral ab initio molecular dynamics. *J. Chem. Theory Comput.* **18**, 599–604 (2022).

22. Rossi, M. Progress and challenges in ab initio simulations of quantum nuclei in weakly bonded systems. *J. Chem. Phys.* **154**, 170902 (2021).
23. Sauceda, H. E., Vassilev-Galindo, V., Chmiela, S., Müller, K. R. & Tkatchenko, A. Dynamical strengthening of covalent and non-covalent molecular interactions by nuclear quantum effects at finite temperature. *Nat. Commun.* **12**, 442 (2021).
24. Ceriotti, M. et al. Nuclear quantum effects in water and aqueous systems: experiment, theory, and current challenges. *Chem. Rev.* **116**, 7529–7550 (2016).
25. Li, X. Z., Walker, B. & Michaelides, A. Quantum nature of the hydrogen bond. *Proc. Natl Acad. Sci. USA* **108**, 6369–6373 (2011).
26. Ceriotti, M., Cuny, J., Parrinello, M. & Manolopoulos, D. E. Nuclear quantum effects and hydrogen bond fluctuations in water. *Proc. Natl Acad. Sci. USA* **110**, 15591–15596 (2013).
27. Peters, B. Reactive flux. In *Reaction Rate Theory and Rare Events Simulations,* 335–362 (Elsevier Inc., Amsterdam, 2017).
28. Schütt, K. T., Sauceda, H. E., Kindermans, P. J., Tkatchenko, A. & Müller, K. R. SchNet - a deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
29. Schütt, K. T. et al. SchNetPack: a deep learning toolbox for atomistic systems. *J. Chem. Theory Comput.* **15**, 448–455 (2019).
30. Batzner, S. et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).
31. Bailleul, S. et al. Ab initio enhanced sampling kinetic study on MTO ethene methylation reaction. *J. Catal.* **388**, 38–51 (2020).
32. Bučko, T., Chibani, S., Paul, J. F., Cantrel, L. & Badawi, M. Dissociative iodomethane adsorption on Ag-MOR and the formation of AgI clusters: an ab initio molecular dynamics study. *Phys. Chem. Chem. Phys.* **19**, 27530–27543 (2017).
33. Anderson, J. B. Statistical theories of chemical reactions. Distributions in the transition region. *J. Chem. Phys.* **58**, 4684–4692 (1973).
34. Bennett, C. H. Molecular dynamics and transition state theory: the simulation of infrequent events. In *Algorithms for Chemical Computations,* 63–97 (American Chemical Society, 1977).
35. Chandler, D. Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *J. Chem. Phys.* **68**, 2959–2970 (1978).
36. Hele, T. J. H. & Althorpe, S. C. Derivation of a true (t → 0+) quantum transition-state theory. I. Uniqueness and equivalence to ring-polymer molecular dynamics transition-state-theory. *J. Chem. Phys.* **138**, 084108 (2013).
37. Dale, H. J. A., Leach, A. G. & Lloyd-Jones, G. C. Heavy-atom kinetic isotope effects: primary interest or zero point? *J. Am. Chem. Soc.* **143**, 21079–21099 (2021).
38. Schroeder, C. et al. Hydrogen bond formation of Brønsted acid sites in zeolites. *Chem. Mater.* **32**, 1564–1574 (2020).
39. Vjunov, A. et al. Following solid-acid-catalyzed reactions by MAS NMR spectroscopy in liquid phase - Zeolite-catalyzed conversion of cyclohexanol in water. *Angew. Chem. Int. Ed.* **53**, 479–482 (2014).
40. Tomkins, P., Ranocchiari, M. & Van Bokhoven, J. A. direct conversion of methane to methanol under mild conditions over Cu-zeolites and beyond. *Acc. Chem. Res.* **50**, 418–425 (2017).
41. Bigeleisen, J. & Mayer, M. G. Calculation of equilibrium constants for isotopic exchange reactions. *J. Chem. Phys.* **15**, 261–267 (1947).
42. Bigeleisen, J. & Wolfsberg, M. Theoretical and experimental aspects of isotope effects in chemical kinetics. In *Advances in Chemical Physics* (ed. Prigogine, I.) 15–76 (John Wiley & Sons, Ltd, 1957).
43. Fermann, J. T. & Auerbach, S. Modeling proton mobility in acidic zeolite clusters. II. Room temperature tunneling effects from semiclassical rate theory. *J. Chem. Phys.* **112**, 6787–6794 (2000).
44. IZA Database of Zeolite Structures (accessed 28 March 2022). http://www.iza-structure.org/databases/.
45. Erlebach, A., Nachtigall, P. & Grajciar, L. Accurate large-scale simulations of siliceous zeolites by neural network potentials. *npj Comput. Mater.* **8**, 174 (2022).
46. Cendagorta, J. R., Shen, H., Bačić, Z. & Tuckerman, M. E. Enhanced sampling path integral methods using neural network potential energy surfaces with application to diffusion in hydrogen hydrates. *Adv. Theory Simul.* **4**, 2000258 (2021).
47. Hellström, M., Ceriotti, M. & Behler, J. Nuclear quantum effects in sodium hydroxide solutions from neural network molecular dynamics simulations. *J. Phys. Chem. B* **122**, 10158–10171 (2018).
48. Bolhuis, P. G., Chandler, D., Dellago, C. & Geissler, P. L. Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **53**, 291–318 (2002).
49. Sauer, J. Ab initio calculations for molecule-surface interactions with chemical accuracy. *Acc. Chem. Res.* **52**, 3502–3510 (2019).
50. Goncalves, T. J., Plessow, P. N. & Studt, F. On the accuracy of density functional theory in zeolite catalysis. *ChemCatChem* **11**, 4368–4376 (2019).
51. Liu, P. & Mei, D. Identifying free energy landscapes of proton-transfer processes between Brønsted acid sites and water clusters inside the zeolite pores. *J. Phys. Chem. C.* **124**, 22568–22576 (2020).
52. Ma, S. & Liu, Z. P. Machine learning potential era of zeolite simulation. *Chem. Sci.* **13**, 5055–5068 (2022).
53. Torrie, G. M. & Valleau, J. P. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.* **28**, 578–581 (1974).
54. Torrie, G. M. & Valleau, J. P. Monte Carlo study of a phase-separating liquid mixture by umbrella sampling. *J. Chem. Phys.* **66**, 1402–1408 (1977).
55. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: new feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613 (2014).
56. The PLUMED consortium. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **16**, 667–673 (2019).
57. ThermoLIB | Center for Molecular Modeling (accessed 19 November 2021) https://molmod.ugent.be/software/thermolib.
58. Vandevondele, J. et al. Quickstep: fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **167**, 103–128 (2005).
59. Kühne, T. D. et al. CP2K: An electronic structure and molecular dynamics software package -Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **152**, 194103 (2020).
60. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
61. Yang, K., Zheng, J., Zhao, Y. & Truhlar, D. G. Tests of the RPBE, revPBE, τ -HCTHhyb, ωb97X-D, and MOHLYP density functional approximations and 29 others against representative databases for diverse bond energies and barrier heights in catalysis. *J. Chem. Phys.* **132**, 1–10 (2010).
62. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010).
63. Lippert, G., Hutter, J. & Parrinello, M. The Gaussian and augmented-plane-wave density functional method for ab initio molecular dynamics simulations. *Theor. Chem. Acc.* **103**, 124–140 (1999).
64. Goedecker, S., Teter, M. & Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B - Condens. Matter Mater. Phys.* **54**, 1703–1710 (1996).
65. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268 (1984).
66. Martyna, G. J., Klein, M. L. & Tuckerman, M. Nosé-Hoover chains: the canonical ensemble via continuous dynamics. *J. Chem. Phys.* **97**, 2635–2643 (1992).

67. Verstraelen, T., Vanduyfhuys, L., Vandenbrande, S. & Rogge, S. M. J. Yaff, Yet Another Force Field. http://molmod.github.io/yaff/index.html.
68. Kapil, V. et al. i-PI 2.0: A universal force engine for advanced molecular simulations. *Comput. Phys. Commun.* **236**, 214–223 (2019).
69. Ceriotti, M., Parrinello, M., Markland, T. E. & Manolopoulos, D. E. Efficient stochastic thermostatting of path integral molecular dynamics. *J. Chem. Phys.* **133**, 124104 (2010).
70. Li, C., Paesani, F. & Voth, G. A. Static and dynamic correlations in water: comparison of classical ab initio molecular dynamics at elevated temperature with path integral simulations at ambient temperature. *J. Chem. Theory Comput.* **18**, 2124–2131 (2022).
71. Frenkel, D. & Smit, B. Rare events. In *Understanding Molecular Simulation* 431–464 (Academic Press, 2002).
72. Craig, I. R. & Manolopoulos, D. E. A refined ring polymer molecular dynamics theory of chemical reaction rates. *J. Chem. Phys.* **123**, 034102 (2005).
73. Voth, G. A., Chandler, D. & Miller, W. H. Rigorous formulation of quantum transition state theory and its dynamical corrections. *J. Chem. Phys.* **91**, 7749–7760 (1989).
74. Suleimanov, Y. V., Javier Aoiz, F. & Guo, H. Chemical reaction rate coefficients from ring polymer molecular dynamics: Theory and practical applications. *J. Phys. Chem. A* **120**, 8488–8502 (2016).
75. Bocus, M., Neale, S. E., Cnudde, P. & Van Speybroeck, V. Dynamic evolution of catalytic active sites within zeolite catalysis. In *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering* (Elsevier Inc., Amsterdam, 2021).

Acknowledgements

This work was supported by the Fund for Scientific Research Flanders (FWO, BioFact Excellence of Science project G0H0918N, ID EOS: 30902231, and project G024019N), the Flanders Industry Innovation Moonshot (ARCLATH II, No. HBC.2021.0254) and the Research Board of Ghent University (BOF). The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, FWO, and the Flemish Government – department EWI.

Author contributions

M.B. and V.V.S. initiated the discussion and designed the scope of the project. M.B. performed the classical DFT simulations. A.L. provided technical support for the PIMD simulations and performed the PIMD DFT simulations. Under the supervision of T.V., R.G. trained the MLP and performed the related simulations, with support from M.C.C. M.B., R.G. and A.L. analyzed the results, which were discussed among all authors. M.B. and V.V.S. wrote the manuscript, with contributions from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-023-36666-y.

Correspondence and requests for materials should be addressed to Veronique Van Speybroeck.

Peer review information *Nature Communications* thanks Donghai Mei and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at http://www.nature.com/reprints

A

List of Publications

Updated June 2023

Publications in international peer–reviewed journals

1. **Insight into the effects of confined hydrocarbon species on the lifetime of methanol conversion catalysts**
   Ines Lezcano-Gonzalez, Emma Campbell, Alexander E. J. Hoffman, Massimo Bocus, Igor V. Sazanovich, Mike Towrie, Miren Agote-Aran, Emma K. Gibson, Alex Greenaway, Kristof De Wispelaere, Veronique Van Speybroeck* and Andrew M. Beale*
   Nature Materials, **2020**, *19*, 1081–1087
   `https://doi.org/10.1038/s41563-020-0800-y`

2. **Shape-selective C-H activation of aromatics to biarylic compounds using molecular palladium in zeolites**
   Jannick Vercammen, Massimo Bocus, Samuel E. Neale, Aram Bugaev, Patrick Tomkins, Julianna Hajek, Sam Van Minnebruggen, Alexander Soldatov, Andraž Krajnc, Gregor Mali, Veronique Van Speybroeck* and Dirk E. De Vos*
   Nature Catalysis, **2020**, *3*, 1002–1009
   `https://doi.org/10.1038/s41929-020-00533-6`

3. **Mechanistic characterization of zeolite-catalyzed aromatic electrophilic substitution at realistic operating conditions**
   Massimo Bocus, Louis Vanduyfhuys, Frank De Proft, Bert M. Weckhuysen and Veronique Van Speybroeck*
   JACS Au, **2022**, *2*, 502–514
   `https://doi.org/10.1021/jacsau.1c00544`

4. **Insights into the mechanism and reactivity of zeolite catalyzed alkylphenol dealkylation**
   Massimo Bocus and Veronique Van Speybroeck*
   ACS Catalysis, **2022**, *12*, 14227–14242
   `https://doi.org/10.1021/acscatal.2c03844`

5. **Quantum free energy profiles for molecular proton transfers**
   Aran Lamaire, Maarten Cools-Ceuppens, Massimo Bocus, Toon Verstraelen and Veronique Van Speybroeck*
   Journal of Chemical Theory and Computations, **2023**, *19*, 18–24
   `https://doi.org/10.1021/acs.jctc.2c00874`

6. **Nuclear quantum effects on zeolite proton hopping kinetics explored with machine learning potentials and path integral molecular dynamics**
   Massimo Bocus[‡], Ruben Goeminne[‡], Aran Lamaire, Maarten Cools-Ceuppens, Toon Verstraelen and Veronique Van Speybroeck*
   Nature Communications, **2023**, *14*, 1008
   `https://doi.org/10.1038/s41467-023-36666-y`

7. **Undercoordinated confined water in Brønsted acidic zeolites speeds up the *O*-activated demethylation of guaiacol in hot-pressurized water**
   Elias Van den Broeck[‡], Massimo Bocus[‡], Xian Wu[‡], Matthias Bal[‡], Jeroen Bomon, Louis Vanduyfhuys, Bert F. Sels*, Bert U. W. Maes* and Veronique Van Speybroeck*
   Under revision

8. **Operando modeling of zeolite catalyzed reactions using first principle molecular dynamics simulations**

Veronique Van Speybroeck*, Massimo Bocus, Pieter Cnudde and Louis Vanduyfhuys
Under revision

B publications

1. **Dynamic evolution of catalytic active sites within zeolite catalysis**
   Massimo Bocus, Samuel E. Neale, Pieter Cnudde and Veronique Van Speybroeck
   in *Comprehensive Inorganic Chemistry III (third edition)*, **2023**, 165–200, Elsevier
   `https://doi.org/10.1016/B978-0-12-823144-9.00012-1`

Conference contributions

Oral presentations

1. **Theoretical insights into mechanism and reactivity of zeolite-catalyzed alkylphenols dealkylation**
   Massimo Bocus, Yuhe Liao, Bert F. Sels and Veronique Van Speybroeck
   NCCC XXI, Noordwijkerhout, The Netherlands, **2020**

2. **Operando Raman spectroscopy to pinpoint the onset of deactivation in zeolites**
   Alexander E. J. Hoffmann, Ines Lezcano-Gonzalez, Emma Campbell, Massimo Bocus, Igor V. Sazanovich, Mike Towrie, Andrew M. Beale and Veronique Van Speybroeck
   NCCC XXII, Online, **2021**

3. **Zeolite-catalyzed benzene ethylation: New mechanistic insights from an old-school reaction class**
   Massimo Bocus, Liesbeth De Bruecker, Louis Vanduyfhuys and Veronique Van Speybroeck
   ACS Spring Symposium, Online, **2021**

4. **The influence of nuclear quantum effects on proton hopping kinetics in the H-SSZ-13 zeolite through ab initio derived machine learning potentials**
   Massimo Bocus, Ruben Goeminne, Aran Lamaire, Maarten Cools-Ceuppens, Toon Verstraelen and Veronique Van Speybroeck
   NCCC XXIII, Noordwijkerhout, The Netherlands, **2022**

5. **Towards a realistic modeling of complex zeolite-catalyzed reactions: the case of guaiacol demethylation**
   Massimo Bocus, Elias Van den Broeck, Mathias Bal, Xian Wu, Jeroen Bomon, Louis Vanduyfhuys, Bert U. W. Maes, Bert F. Sels and Veronique Van Speybroeck
   ICTAC, Lyon, France, **2022**

Poster presentations

1. **First-principles mechanistic study of alkylphenols dealkylation catalyzed by acidic zeolites**
   Massimo Bocus, Julianna Hajek, Yuhe Liao, Bert F. Sels and Veronique Van Speybroeck
   NCCC XX, Noordwijkerhout, The Netherlands, **2019**

2. **Ab initio modeling of the zeolite-catalyzed conversion of alkylphenols into phenol and olefins**
   Massimo Bocus, Julianna Hajek, Yuhe Liao, Bert F. Sels and Veronique Van Speybroeck
   EuropaCat 2019, Aachen, Germany, **2019**

3. **Operando Raman spectroscopy to pinpoint the onset of deactivation in zeolites**
   Alexander E. J. Hoffmann, Ines Lezcano-Gonzalez, Emma Campbell, Massimo Bocus, Igor V. Sazanovich, Mike Towrie, Andrew M. Beale and Veronique Van Speybroeck
   FEZA, Online, **2021**

4. **Mechanistic Insights on the Water-Zeolite Synergy in the Conversion of Biomass-Derived Guaiacol**
   Massimo Bocus, Elias Van den Broeck, Mathias Bal, Xian Wu, Louis Vanduyfhuys, Bert U. W. Maes, Bert F. Sels and Veronique Van Speybroeck
   Water in Zeolites Workshop, Liblice Castle, Czech Republic, **2021**

B

Acknowledgments

This research was performed within the Excellence of Science (EOS) project BioFact (EOS ID 30902231), founded by the Research Foundation Flanders (FWO), and the FWO Research Projects G024019N and G0D0518N.

The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, FWO and the Flemish Government – department EWI.

Bibliography

[1] Wang, L., Sofer, Z. & Pumera, M. Will any crap we put into graphene increase its electrocatalytic effect? *ACS Nano* **14**, 21–25 (2020).

[2] Dirac, P. A. M. Quantum mechanics of many-electron systems. *Proceedings of the Royal Society of London A* **123**, 714–733 (1929).

[3] Ciccotti, G., Dellago, C., Ferrario, M., Hernández, E. & Tuckerman, M. Molecular simulations: past, present, and future (a Topical Issue in EPJB). *The European Physical Journal B* **95**, 3 (2022).

[4] Chang, C., Deringer, V. L., Katti, K. S., Van Speybroeck, V. & Wolverton, C. M. Simulations in the era of exascale computing. *Nature Reviews Materials* (2023).

[5] The Nobel Prize in Chemistry 1918. `https://www.nobelprize.org/prizes/chemistry/1918/summary/` (Accessed: 2023-02-22).

[6] The Nobel Prize in Chemistry 1931. `https://www.nobelprize.org/prizes/chemistry/1931/summary/` (Accessed: 2023-02-22).

[7] The Nobel Prize in Chemistry 2021. `https://www.nobelprize.org/prizes/chemistry/2021/summary/` (Accessed: 2022-11-28).

[8] McQuarrie, D. A. & Simon, J. D. *Physical chemistry: a molecular approach* (University Science Books, 1997).

[9] Bond, G. C., Keane, M. A., Kral, H. & Lercher, J. A. Compensation phenomena in heterogeneous catalysis: general principles and a possible explanation. *Catalysis Reviews* **42**, 323–383 (2000).

[10] Cubillas, P., Anderson, M. W. *et al.* *Zeolites and Catalysis: Synthesis, Reactions and Applications* (Wiley-VCH Verlag GmbH & Co. KGaA, 2010).

[11] Van Speybroeck, V. *et al.* Advances in theory and their application within the field of zeolite chemistry. *Chemical Society Reviews* **44**, 7044–7111 (2015).

[12] IZA database of zeolite structures. `http://www.iza-structure.org/databases/` (Accessed: 2022-12-05).

[13] Pophale, R., Cheeseman, P. A. & Deem, M. W. A database of new zeolite-like materials. *Physical Chemistry Chemical Physics* **13**, 12407–12412 (2011).

[14] Lesthaeghe, D., Van Speybroeck, V. & Waroquier, M. Theoretical evaluation of zeolite confinement effects on the reactivity of bulky intermediates. *Physical Chemistry Chemical Physics* **11**, 5222–5226 (2009).

[15] Smit, B. & Maesen, T. L. Towards a molecular understanding of shape selectivity. *Nature* **451**, 671–678 (2008).

[16] Weissermel, K. & Arpe, H.-J. *Industrial Organic Chemistry, Fourth Edition* (WILEY-VCH Verlag GmbH & Co. KGaA, 2003).

[17] Galabov, B., Nalbantova, D., Schleyer, P. v. R. & Schaefer III, H. F. Electrophilic aromatic substitution: New insights into an old class of reactions. *Accounts of Chemical Research* **49**, 1191–1199 (2016).

[18] Degnan Jr, T. F., Smith, C. M. & Venkat, C. R. Alkylation of aromatics with ethylene and propylene: recent developments in commercial processes. *Applied Catalysis A: General* **221**, 283–294 (2001).

[19] Vogt, E. T., Whiting, G. T., Chowdhury, A. D. & Weckhuysen, B. M. Zeolites and zeotypes for oil and gas conversion. In *Advances in Catalysis*, vol. 58, 143–314 (Elsevier, 2015).

[20] Venuto, P. B., Hamilton, L. A., Landis, P. S. & Wise, J. J. Organic reactions catalyzed by crystalline aluminosilicates: I. Alkylation reactions. *Journal of Catalysis* **5**, 81–98 (1966).

[21] Ennaert, T. *et al.* Potential and challenges of zeolite chemistry in the catalytic conversion of biomass. *Chemical Society Reviews* **45**, 584–611 (2016).

[22] Verboekend, D., Liao, Y., Schutyser, W. & Sels, B. F. Alkylphenols to phenol and olefins by zeolite catalysis: a pathway to valorize raw and fossilized lignocellulose. *Green Chemistry* **18**, 297–306 (2016).

[23] Liao, Y. *et al.* Propylphenol to phenol and propylene over acidic zeolites: role of shape selectivity and presence of steam. *ACS Catalysis* **8**, 7861–7878 (2018).

[24] Liao, Y., d'Halluin, M., Makshina, E., Verboekend, D. & Sels, B. F. Shape selectivity vapor-phase conversion of lignin-derived 4-ethylphenol to phenol and ethylene over acidic aluminosilicates: Impact of acid properties and pore constraint. *Applied Catalysis B: Environmental* **234**, 117–129 (2018).

[25] Liao, Y., Zhong, R., d'Halluin, M., Verboekend, D. & Sels, B. F. Aromatics production from lignocellulosic biomass: shape selective dealkylation of lignin-derived phenolics over hierarchical ZSM-5. *ACS Sustainable Chemistry & Engineering* **8**, 8713–8722 (2020).

[26] Van Speybroeck, V. *et al.* First principle chemical kinetics in zeolites: the methanol-to-olefin process as a case study. *Chemical Society Reviews* **43**, 7326–7357 (2014).

[27] Chowdhury, A. D. *et al.* Electrophilic aromatic substitution over zeolites generates Wheland-type reaction intermediates. *Nature Catalysis* **1**, 23–31 (2018).

[28] Vos, A. M., Schoonheydt, R. A., De Proft, F. & Geerlings, P. Reactivity descriptors and rate constants for acid zeolite catalyzed ethylation and isopropylation of benzene. *The Journal of Physical Chemistry B* **107**, 2001–2008 (2003).

[29] Arstad, B., Kolboe, S. & Swang, O. Theoretical investigation of arene alkylation by ethene and propene over acidic zeolites. *The Journal of Physical Chemistry B* **108**, 2300–2308 (2004).

[30] Namuangruk, S., Pantu, P. & Limtrakul, J. Alkylation of benzene with ethylene over faujasite zeolite investigated by the ONIOM method. *Journal of Catalysis* **225**, 523–530 (2004).

[31] Hansen, N., Brüggemann, T., Bell, A. T. & Keil, F. J. Theoretical investigation of benzene alkylation with ethene over H-ZSM-5. *The Journal of Physical Chemistry C* **112**, 15402–15411 (2008).

[32] Hansen, N., Kerber, T., Sauer, J., Bell, A. T. & Keil, F. J. Quantum chemical modeling of benzene ethylation over H-ZSM-5 approaching chemical accuracy: a hybrid MP2: DFT study. *Journal of the American Chemical Society* **132**, 11525–11538 (2010).

[33] Wang, D., Wang, C.-M., Yang, G., Du, Y.-J. & Yang, W.-M. First-principles kinetic study on benzene alkylation with ethanol vs. ethylene in H-ZSM-5. *Journal of Catalysis* **374**, 1–11 (2019).

[34] Acharya, D. *et al.* Stepwise or Concerted Mechanisms of Benzene Ethylation Catalyzed by Zeolites? Theoretical Analysis of Reaction Pathways. *Catalysis Letters* **151**, 3048–3056 (2021).

[35] Kolboe, S., Svelle, S. & Arstad, B. Theoretical study of ethylbenzenium ions: The mechanism for splitting off ethene, and the formation of a π complex of ethene and the benzenium ion. *The Journal of Physical Chemistry A* **113**, 917–923 (2009).

[36] Zakzeski, J., Bruijnincx, P. C., Jongerius, A. L. & Weckhuysen, B. M. The catalytic valorization of lignin for the production of renewable chemicals. *Chemical Reviews* **110**, 3552–3599 (2010).

[37] Calvo-Flores, F. G. & Dobado, J. A. Lignin as renewable raw material. *ChemSusChem* **3**, 1227–1235 (2010).

[38] Graglia, M., Kanna, N. & Esposito, D. Lignin refinery: Towards the preparation of renewable aromatic building blocks. *ChemBioEng Reviews* **2**, 377–392 (2015).

[39] Ha, J.-M. *et al.* Recent progress in the thermal and catalytic conversion of lignin. *Renewable and Sustainable Energy Reviews* **111**, 422–441 (2019).

[40] Renders, T., Van den Bossche, G., Vangeel, T., Van Aelst, K. & Sels, B. Reductive catalytic fractionation: state of the art of the lignin-first biorefinery. *Current Opinion in Biotechnology* **56**, 193–201 (2019).

[41] Liao, Y. *et al.* A sustainable wood biorefinery for low–carbon footprint chemicals production. *Science* **367**, 1385–1390 (2020).

[42] Schutyser, W. *et al.* Chemicals from lignin: an interplay of lignocellulose fractionation, depolymerisation, and upgrading. *Chemical Society Reviews* **47**, 852–908 (2018).

[43] Sun, Z., Fridrich, B., De Santi, A., Elangovan, S. & Barta, K. Bright side of lignin depolymerization: toward new platform chemicals. *Chemical Reviews* **118**, 614–678 (2018).

[44] Yang, L., Zhou, W., Seshan, K. & Li, Y. Green and efficient synthesis route of catechol from guaiacol. *Journal of Molecular Catalysis A: Chemical* **368**, 61–65 (2013).

[45] Wu, X., Fu, J. & Lu, X. Kinetics and mechanism of hydrothermal decomposition of lignin model compounds. *Industrial & Engineering Chemistry Research* **52**, 5016–5022 (2013).

[46] Bomon, J. *et al.* Brønsted acid catalyzed tandem defunctionalization of biorenewable ferulic acid and derivates into bio-catechol. *Angewandte Chemie International Edition* **59**, 3063–3068 (2020).

[47] Bomon, J. *et al.* Efficient demethylation of aromatic methyl ethers with HCl in water. *Green Chemistry* **23**, 1995–2009 (2021).

[48] Wu, X. *et al.* Lignin-First Monomers to Catechol: Rational Cleavage of C-O and C-C Bonds over Zeolites. *ChemSusChem* **15**, e202102248 (2022).

[49] Liu, Y. *et al.* Enhancing the catalytic activity of hydronium ions through constrained environments. *Nature Communications* **8**, 14113 (2017).

[50] Shi, H., Eckstein, S., Vjunov, A., Camaioni, D. M. & Lercher, J. A. Tailoring nanoscopic confines to maximize catalytic activity of hydronium ions. *Nature Communications* **8**, 15442 (2017).

[51] Pfriem, N. *et al.* Role of the ionic environment in enhancing the activity of reacting molecules in zeolite pores. *Science* **372**, 952–957 (2021).

[52] Vogt, C. & Weckhuysen, B. M. The concept of active site in heterogeneous catalysis. *Nature Reviews Chemistry* **6**, 89–111 (2022).

[53] Chizallet, C. Toward the atomic scale simulation of intricate acidic aluminosilicate catalysts. *ACS Catalysis* **10**, 5579–5601 (2020).

[54] Hu, Z.-P., Han, J., Wei, Y. & Liu, Z. Dynamic evolution of zeolite framework and metal-zeolite interface. *ACS Catalysis* **12**, 5060–5076 (2022).

[55] Bocus, M., Neale, S. E., Cnudde, P. & Van Speybroeck, V. Dynamic evolution of catalytic active sites within zeolite catalysis. In *Comprehensive Inorganic Chemistry III (third edition)*, 165–200 (Elsevier, 2023).

[56] Heard, C. J. *et al.* Zeolite (in) stability under aqueous or steaming conditions. *Advanced Materials* **32**, 2003264 (2020).

[57] Stanciakova, K. & Weckhuysen, B. M. Water–active site interactions in zeolites and their relevance in catalysis. *Trends in Chemistry* **3**, 456–468 (2021).

[58] Resasco, D. E., Crossley, S. P., Wang, B. & White, J. L. Interaction of water with zeolites: a review. *Catalysis Reviews* **63**, 302–362 (2021).

[59] Wang, M. *et al.* Genesis and stability of hydronium ions in zeolite channels. *Journal of the American Chemical Society* **141**, 3444–3455 (2019).

[60] Liu, P. & Mei, D. Identifying Free Energy Landscapes of Proton-Transfer Processes between Brønsted Acid Sites and Water Clusters Inside the Zeolite Pores. *The Journal of Physical Chemistry C* **124**, 22568–22576 (2020).

[61] Grifoni, E. *et al.* Confinement effects and acid strength in zeolites. *Nature Communications* **12**, 1–9 (2021).

[62] Heard, C. J. *et al.* Fast room temperature lability of aluminosilicate zeolites. *Nature Communications* **10**, 4690 (2019).

[63] Pugh, S. M., Wright, P. A., Law, D. J., Thompson, N. & Ashbrook, S. E. Facile, room-temperature 17O enrichment of zeolite frameworks revealed by solid-state NMR spectroscopy. *Journal of the American Chemical Society* **142**, 900–906 (2020).

[64] Lisboa, O., Sánchez, M. & Ruette, F. Modeling extra framework aluminum (EFAL) formation in the zeolite ZSM-5 using parametric quantum and DFT methods. *Journal of Molecular Catalysis A: Chemical* **294**, 93–101 (2008).

[65] Malola, S., Svelle, S., Bleken, F. L. & Swang, O. Detailed reaction paths for zeolite dealumination and desilication from density functional calculations. *Angewandte Chemie International Edition* **51**, 652–655 (2012).

[66] Silaghi, M.-C., Chizallet, C. & Raybaud, P. Challenges on molecular aspects of dealumination and desilication of zeolites. *Microporous and Mesoporous Materials* **191**, 82–96 (2014).

[67] Nielsen, M. *et al.* Kinetics of zeolite dealumination: insights from H-SSZ-13. *ACS Catalysis* **5**, 7131–7139 (2015).

[68] Silaghi, M.-C. *et al.* Regioselectivity of Al–O bond hydrolysis during zeolites dealumination unified by brønsted–evans–polanyi relationship. *ACS Catalysis* **5**, 11–15 (2015).

[69] Silaghi, M.-C., Chizallet, C., Sauer, J. & Raybaud, P. Dealumination mechanisms of zeolites and extra-framework aluminum confinement. *Journal of Catalysis* **339**, 242–255 (2016).

[70] Stanciakova, K., Ensing, B., Goltl, F., Bulo, R. E. & Weckhuysen, B. M. Cooperative role of water molecules during the initial stage of water-induced zeolite dealumination. *ACS Catalysis* **9**, 5119–5135 (2019).

[71] Nielsen, M. *et al.* Collective action of water molecules in zeolite dealumination. *Catalysis Science & Technology* **9**, 3721–3725 (2019).

[72] Liu, P., Liu, Q. & Mei, D. Dealumination of the H-BEA Zeolite via the S N2 Mechanism: A Theoretical Investigation. *The Journal of Physical Chemistry C* **125**, 24613–24621 (2021).

[73] Palčić, A. & Valtchev, V. Analysis and control of acid sites in zeolites. *Applied Catalysis A: General* **606**, 117795 (2020).

[74] Ravi, M., Sushkevich, V. L. & van Bokhoven, J. A. Towards a better understanding of Lewis acidic aluminium in zeolites. *Nature materials* **19**, 1047–1056 (2020).

[75] Liu, C., Li, G., Hensen, E. J. & Pidko, E. A. Nature and catalytic role of extraframework aluminum in faujasite zeolite: a theoretical perspective. *ACS Catalysis* **5**, 7024–7033 (2015).

[76] Wang, Z., Wang, L., Jiang, Y., Hunger, M. & Huang, J. Cooperativity of Brønsted and Lewis acid sites on zeolite for glycerol dehydration. *ACS Catalysis* **4**, 1144–1147 (2014).

[77] Chu, Y., Yi, X., Li, C., Sun, X. & Zheng, A. Brønsted/Lewis acid sites synergistically promote the initial C–C bond formation in the MTO reaction. *Chemical Science* **9**, 6470–6479 (2018).

[78] Bailleul, S. *et al.* A supramolecular view on the cooperative role of Brønsted and Lewis acid sites in zeolites for methanol conversion. *Journal of the American Chemical Society* **141**, 14823–14842 (2019).

[79] Liu, C., Li, G., Hensen, E. J. & Pidko, E. A. Relationship between acidity and catalytic reactivity of faujasite zeolite: A periodic DFT study. *Journal of Catalysis* **344**, 570–577 (2016).

[80] Mezari, B., Magusin, P. C., Almutairi, S. M., Pidko, E. A. & Hensen, E. J. Nature of enhanced Brønsted acidity induced by extraframework aluminum in an ultrastabilized faujasite zeolite: an in situ NMR study. *The Journal of Physical Chemistry C* **125**, 9050–9059 (2021).

[81] Chen, K. *et al.* Brønsted–Brønsted synergies between framework and noncrystalline protons in zeolite H-ZSM-5. *ACS Catalysis* **9**, 6124–6136 (2019).

[82] Chen, K. *et al.* Structure and catalytic characterization of a second framework Al (IV) site in zeolite catalysts revealed by NMR at 35.2 T. *Journal of the American Chemical Society* **142**, 7514–7523 (2020).

[83] Vjunov, A. *et al.* Tracking the chemical transformations at the Brønsted acid site upon water-induced deprotonation in a zeolite pore. *Chemistry of Materials* **29**, 9030–9042 (2017).

[84] Rey, J., Raybaud, P. & Chizallet, C. Ab initio simulation of the acid sites at the external surface of zeolite beta. *ChemCatChem* **9**, 2176–2185 (2017).

[85] Treps, L., Gomez, A., de Bruin, T. & Chizallet, C. Environment, stability and acidity of external surface sites of silicalite-1 and ZSM-5 micro and nano slabs, sheets, and crystals. *ACS Catalysis* **10**, 3297–3312 (2020).

[86] Martin, A. J., Mitchell, S., Mondelli, C., Jaydev, S. & Perez-Ramirez, J. Unifying views on catalyst deactivation. *Nature Catalysis* **5**, 854–866 (2022).

[87] Lezcano-Gonzalez, I. *et al.* Insight into the effects of confined hydrocarbon species on the lifetime of methanol conversion catalysts. *Nature Materials* **19**, 1081–1087 (2020).

[88] Yarulina, I., Chowdhury, A. D., Meirer, F., Weckhuysen, B. M. & Gascon, J. Recent trends and fundamental insights in the methanol-to-hydrocarbons process. *Nature Catalysis* **1**, 398–411 (2018).

[89] Plessow, P. N., Smith, A., Tischer, S. & Studt, F. Identification of the reaction sequence of the MTO initiation mechanism using ab initio-based kinetics. *Journal of the American Chemical Society* **141**, 5908–5915 (2019).

[90] Fečík, M., Plessow, P. N. & Studt, F. Theoretical investigation of the side-chain mechanism of the MTO process over H-SSZ-13 using DFT and ab initio calculations. *Catalysis Science & Technology* **11**, 3826–3833 (2021).

[91] Schmidt, J. E. *et al.* Coke formation in a zeolite crystal during the methanol-to-hydrocarbons reaction as studied with atom probe tomography. *Angewandte Chemie International Edition* **55**, 11173–11177 (2016).

[92] Mores, D. *et al.* Space-and time-resolved in-situ spectroscopy on the coke formation in molecular sieves: methanol-to-olefin conversion

over H-ZSM-5 and H-SAPO-34. *Chemistry–A European Journal* **14**, 11320–11327 (2008).

[93] Mores, D., Kornatowski, J., Olsbye, U. & Weckhuysen, B. M. Coke formation during the methanol-to-olefin conversion: in situ microspectroscopy on individual H-ZSM-5 crystals with different Brønsted acidity. *Chemistry–A European Journal* **17**, 2874–2884 (2011).

[94] Perdew, J. P. & Schmidt, K. Jacob's ladder of density functional approximations for the exchange-correlation energy. In *AIP Conference Proceedings*, vol. 577, 1–20 (American Institute of Physics, 2001).

[95] Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Physical Review Letters* **77**, 3865–3868 (1996).

[96] Goncalves, T. J., Plessow, P. N. & Studt, F. On the accuracy of density functional theory in zeolite catalysis. *ChemCatChem* **11**, 4368–4376 (2019).

[97] Plessow, P. N. & Studt, F. How accurately do approximate density functionals predict trends in acidic zeolite catalysis? *The Journal of Physical Chemistry Letters* **11**, 4305–4310 (2020).

[98] Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A* **38**, 3098–3100 (1988).

[99] Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* **37**, 785–789 (1988).

[100] Vosko, S. H., Wilk, L. & Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics* **58**, 1200–1211 (1980).

[101] Stephens, P. J., Devlin, F. J., Chabalowski, C. F. & Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *The Journal of Physical Chemistry* **98**, 11623–11627 (1994).

[102] Deng, C., Zhang, J.-X. & Lin, Z. Theoretical studies on Pd (ii)-catalyzed meta-selective C–H bond arylation of arenes. *ACS Catalysis* **8**, 2498–2507 (2018).

[103] Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **132**, 154104 (2010).

[104] Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry* **32**, 1456–1465 (2011).

[105] Kresse, G. & Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Physical Review B* **49**, 14251–14269 (1994).

[106] Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science* **6**, 15–50 (1996).

[107] Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B* **54**, 11169–11186 (1996).

[108] VandeVondele, J. *et al.* Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Computer Physics Communications* **167**, 103–128 (2005).

[109] Kühne, T. D. *et al.* CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *The Journal of Chemical Physics* **152**, 194103 (2020).

[110] Blöchl, P. E. Projector augmented-wave method. *Physical Review B* **50**, 17953–17979 (1994).

[111] Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B* **59**, 1758–1775 (1999).

[112] Lippert, G., Hutter, J. & Parrinello, M. A hybrid Gaussian and plane wave density functional scheme. *Molecular Physics* **92**, 477–488 (1997).

[113] Lippert, G., Hutter, J. & Parrinello, M. The Gaussian and augmented-plane-wave density functional method for ab initio molecular dynamics simulations. *Theoretical Chemistry Accounts* **103**, 124–140 (1999).

[114] Goedecker, S., Teter, M. & Hutter, J. Separable dual-space Gaussian pseudopotentials. *Physical Review B* **54**, 1703–1710 (1996).

[115] Bates, S. P. & Van Santen, R. A. The molecular basis of zeolite catalysis: a review of theoretical simulations. *Advances in Catalysis* **42**, 1–114 (1998).

[116] Van Santen, R. A. & Kramer, G. J. Reactivity theory of zeolitic Brønsted acidic sites. *Chemical Reviews* **95**, 637–660 (1995).

[117] Jones, A. J. & Iglesia, E. The strength of Brønsted acid sites in microporous aluminosilicates. *Acs Catalysis* **5**, 5741–5755 (2015).

[118] Chen, W., Yi, X., Liu, Z., Tang, X. & Zheng, A. Carbocation chemistry confined in zeolites: spectroscopic and theoretical characterizations. *Chemical Society Reviews* **51**, 4337–4385 (2022).

[119] Gong, X. *et al.* First-generation organic reaction intermediates in zeolite chemistry and catalysis. *Chemical Reviews* **122**, 14275–14345 (2022).

[120] Shah, R., Payne, M. C., Lee, M.-H. & Gale, J. D. Understanding the catalytic behavior of zeolites: a first-principles study of the adsorption of methanol. *Science* **271**, 1395–1397 (1996).

[121] Nusterer, E., Blöchl, P. E. & Schwarz, K. Structure and dynamics of methanol in a zeolite. *Angewandte Chemie International Edition* **35**, 175–177 (1996).

[122] Verstraelen, T., Van Speybroeck, V. & Waroquier, M. ZEOBUILDER: a GUI toolkit for the construction of complex molecular structures on the nanoscale with building blocks. *Journal of Chemical Information and Modeling* **48**, 1530–1541 (2008).

[123] Tuma, C. & Sauer, J. A hybrid MP2/planewave-DFT scheme for large chemical systems: proton jumps in zeolites. *Chemical Physics Letters* **387**, 388–394 (2004).

[124] Tuma, C. & Sauer, J. Treating dispersion effects in extended systems by hybrid MP2: DFT calculations—protonation of isobutene in zeolite ferrierite. *Physical Chemistry Chemical Physics* **8**, 3955–3965 (2006).

[125] Møller, C. & Plesset, M. S. Note on an approximation treatment for many-electron systems. *Physical Reviews* **46**, 618–622 (1934).

[126] Hampel, C., Peterson, K. A. & Werner, H.-J. A comparison of the efficiency and accuracy of the quadratic configuration interaction (QCISD), coupled cluster (CCSD), and Brueckner coupled cluster (BCCD) methods. *Chemical Physics Letters* **190**, 1–12 (1992).

[127] Neese, F., Hansen, A. & Liakos, D. G. Efficient and accurate approximations to the local coupled cluster singles doubles method using a truncated pair natural orbital basis. *The Journal of Chemical Physics* **131**, 064103 (2009).

[128] Riplinger, C. & Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *The Journal of Chemical Physics* **138**, 034106 (2013).

[129] Sauer, J. Ab initio calculations for molecule–surface interactions with chemical accuracy. *Accounts of Chemical Research* **52**, 3502–3510 (2019).

[130] Piccini, G., Alessio, M. & Sauer, J. Ab initio calculation of rate constants for molecule–surface reactions with chemical accuracy. *Angewandte Chemie International Edition* **128**, 5321–5323 (2016).

[131] Guisnet, M. & Gilson, J.-P. *Zeolites for Cleaner Technologies* (Imperial College Press, distributed by World Scientific Publishing co., 2002).

[132] Vermeiren, W. & Gilson, J.-P. Impact of zeolites on the petroleum and petrochemical industry. *Topics in Catalysis* **52**, 1131–1161 (2009).

[133] Vogt, E. T. & Weckhuysen, B. M. Fluid catalytic cracking: recent developments on the grand old lady of zeolite catalysis. *Chemical Society Reviews* **44**, 7342–7370 (2015).

[134] Bučko, T. & Hafner, J. The role of spatial constraints and entropy in the adsorption and transformation of hydrocarbons catalyzed by zeolites. *Journal of Catalysis* **329**, 32–48 (2015).

[135] Boronat, M., Viruela, P. & Corma, A. Theoretical study of the mechanism of zeolite-catalyzed isomerization reactions of linear butenes. *The Journal of Physical Chemistry A* **102**, 982–989 (1998).

[136] Kazansky, V. B. Adsorbed carbocations as transition states in heterogeneous acid catalyzed transformations of hydrocarbons. *Catalysis Today* **51**, 419–434 (1999).

[137] Rozanska, X., van Santen, R. A., Demuth, T., Hutschka, F. & Hafner, J. A periodic DFT study of isobutene chemisorption in proton-exchanged zeolites: dependence of reactivity on the zeolite framework structure. *The Journal of Physical Chemistry B* **107**, 1309–1315 (2003).

[138] Boronat, M., Viruela, P. & Corma, A. Reaction intermediates in acid catalysis by zeolites: prediction of the relative tendency to form alkoxides or carbocations as a function of hydrocarbon nature and active site structure. *Journal of the American Chemical Society* **126**, 3300–3309 (2004).

[139] Nieminen, V., Sierka, M., Murzin, D. Y. & Sauer, J. Stabilities of C3–C5 alkoxide species inside H-FER zeolite: a hybrid QM/MM study. *Journal of Catalysis* **231**, 393–404 (2005).

[140] Tuma, C. & Sauer, J. Protonated isobutene in zeolites: tert-butyl cation or alkoxide? *Angewandte Chemie International Edition* **44**, 4769–4771 (2005).

[141] Fang, H. *et al.* Theoretical investigation of the effects of the zeolite framework on the stability of carbenium ions. *The Journal of Physical Chemistry C* **115**, 7429–7439 (2011).

[142] Cohen, A. J., Mori-Sánchez, P. & Yang, W. Insights into current limitations of density functional theory. *Science* **321**, 792–794 (2008).

[143] Ren, Q., Rybicki, M. & Sauer, J. Interaction of C3–C5 Alkenes with Zeolitic Brønsted Sites: π-Complexes, Alkoxides, and Carbenium Ions in H-FER. *The Journal of Physical Chemistry C* **124**, 10067–10078 (2020).

[144] Cnudde, P., De Wispelaere, K., Van der Mynsbrugge, J., Waroquier, M. & Van Speybroeck, V. Effect of temperature and branching on the nature and stability of alkene cracking intermediates in H-ZSM-5. *Journal of Catalysis* **345**, 53–69 (2017).

[145] Cnudde, P. *et al.* How chain length and branching influence the alkene cracking reactivity on H-ZSM-5. *ACS catalysis* **8**, 9579–9595 (2018).

[146] Tuma, C., Kerber, T. & Sauer, J. The tert-Butyl Cation in H-Zeolites: Deprotonation to Isobutene and Conversion into Surface Alkoxides. *Angewandte Chemie International Edition* **49**, 4678–4680 (2010).

[147] Ristanović, Z., Kubarev, A. V., Hofkens, J., Roeffaers, M. B. & Weckhuysen, B. M. Single molecule nanospectroscopy visualizes proton-transfer processes within a zeolite crystal. *Journal of the American Chemical Society* **138**, 13586–13596 (2016).

[148] Ristanović, Z. *et al.* Reversible and site-dependent proton-transfer in zeolites uncovered at the single-molecule level. *Journal of the American Chemical Society* **140**, 14195–14205 (2018).

[149] Stepanov, A. G. & Zamaraev, K. I. 13C solid state NMR evidence for the existence of isobutyl carbenium ion in the reaction of isobutyl alcohol dehydration in H-ZSM-5 zeolite. *Catalysis letters* **19**, 153–158 (1993).

[150] Haw, J. F., Nicholas, J. B., Xu, T., Beck, L. W. & Ferguson, D. B. Physical organic chemistry of solid acids: lessons from in situ NMR

and theoretical chemistry. *Accounts of Chemical Research* **29**, 259–267 (1996).

[151] Dai, W. *et al.* Identification of tert-Butyl Cations in Zeolite H-ZSM-5: Evidence from NMR Spectroscopy and DFT Calculations. *Angewandte Chemie International Edition* **54**, 8783–8786 (2015).

[152] Sarazen, M. L. & Iglesia, E. Stability of bound species during alkene reactions on solid acids. *Proceedings of the National Academy of Sciences* **114**, E3900–E3908 (2017).

[153] Zhao, R., Haller, G. L. & Lercher, J. A. Alkene adsorption and cracking in acidic zeolites–A gradual process of understanding. *Microporous and Mesoporous Materials* 112390 (2022).

[154] Rupp, M. Machine learning for quantum mechanics in a nutshell. *International Journal of Quantum Chemistry* **115**, 1058–1073 (2015).

[155] De, S., Bartók, A. P., Csányi, G. & Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Physical Chemistry Chemical Physics* **18**, 13754–13769 (2016).

[156] Chehaibou, B., Badawi, M., Bučko, T., Bazhirov, T. & Rocca, D. Computing RPA adsorption enthalpies by machine learning thermodynamic perturbation theory. *Journal of Chemical Theory and Computation* **15**, 6333–6342 (2019).

[157] Bučko, T., Gesvandtnerova, M. & Rocca, D. Ab initio calculations of free energy of activation at multiple electronic structure levels made affordable: An effective combination of perturbation theory and machine learning. *Journal of Chemical Theory and Computation* **16**, 6049–6060 (2020).

[158] Gešvandtnerová, M., Rocca, D. & Bučko, T. Methanol carbonylation over acid mordenite: Insights from ab initio molecular dynamics and machine learning thermodynamic perturbation theory. *Journal of Catalysis* **396**, 166–178 (2021).

[159] Herzog, B. *et al.* Assessing the Accuracy of Machine Learning Thermodynamic Perturbation Theory: Density Functional Theory and Beyond. *Journal of Chemical Theory and Computation* **18**, 1382–1394 (2022).

[160] Carter, E., Ciccotti, G., Hynes, J. T. & Kapral, R. Constrained reaction coordinate dynamics for the simulation of rare events. *Chemical Physics Letters* **156**, 472–477 (1989).

[161] De Wispelaere, K., Plessow, P. N. & Studt, F. Toward Computing Accurate Free Energies in Heterogeneous Catalysis: a Case Study for Adsorbed Isobutene in H-ZSM-5. *ACS Physical Chemistry Au* **2**, 399–406 (2022).

[162] Del Ben, M., Hutter, J. & VandeVondele, J. Second-order Møller–Plesset perturbation theory in the condensed phase: An efficient and massively parallel Gaussian and plane waves approach. *Journal of Chemical Theory and Computation* **8**, 4177–4188 (2012).

[163] Del Ben, M., Hutter, J. & VandeVondele, J. Electron correlation in the condensed phase from a resolution of identity approach based on the Gaussian and plane waves scheme. *Journal of Chemical Theory and Computation* **9**, 2654–2671 (2013).

[164] Stein, F., Hutter, J. & Rybkin, V. V. Double-hybrid DFT functionals for the condensed phase: Gaussian and plane waves implementation and evaluation. *Molecules* **25**, 5174 (2020).

[165] Stein, F. & Hutter, J. Double-hybrid density functionals for the condensed phase: Gradients, stress tensor, and auxiliary-density matrix method acceleration. *The Journal of Chemical Physics* **156**, 074107 (2022).

[166] Eyring, H. The activated complex in chemical reactions. *The Journal of Chemical Physics* **3**, 107–115 (1935).

[167] Head, J. D. Computation of vibrational frequencies for adsorbates on surfaces. *International Journal of Quantum Chemistry* **65**, 827–838 (1997).

[168] Li, H. & Jensen, J. H. Partial Hessian vibrational analysis: the localization of the molecular vibrational energy and entropy. *Theoretical Chemistry Accounts* **107**, 211–219 (2002).

[169] De Moor, B. A. *et al.* Normal mode analysis in zeolites: toward an efficient calculation of adsorption entropies. *Journal of Chemical Theory and Computation* **7**, 1090–1101 (2011).

[170] Piccini, G. & Sauer, J. Effect of anharmonicity on adsorption thermodynamics. *Journal of Chemical Theory and Computation* **10**, 2479–2487 (2014).

[171] De Wispelaere, K., Vanduyfhuys, L. & Van Speybroeck, V. Entropy Contributions to Transition State Modeling. In Catlow, C. R. A., Van Speybroeck, V. & Van Santen, R. A. (eds.) *Modelling and Simulation*

in the Science of Micro- and Meso-Porous Materials, 189–228 (Elsevier, 2018).

[172] Collinge, G. *et al.* Effect of collective dynamics and anharmonicity on entropy in heterogenous catalysis: building the case for advanced molecular simulations. *ACS Catalysis* **10**, 9236–9260 (2020).

[173] De Moor, B. A., Reyniers, M.-F. & Marin, G. B. Physisorption and chemisorption of alkanes and alkenes in H-FAU: a combined ab initio–statistical thermodynamics study. *Physical Chemistry Chemical Physics* **11**, 2939–2958 (2009).

[174] De Moor, B. A., Reyniers, M.-F., Gobin, O. C., Lercher, J. A. & Marin, G. B. Adsorption of C2–C8 n-Alkanes in Zeolites. *Journal of Physical Chemistry C* **115**, 1204–1219 (2011).

[175] Zhi, Y. *et al.* Dehydration Pathways of 1-Propanol on HZSM-5 in the Presence and Absence of Water. *Journal of the American Chemical Society* **137**, 15781–15794 (2015).

[176] Herrmann, S. & Iglesia, E. Elementary steps in acetone condensation reactions catalyzed by aluminosilicates with diverse void structures. *Journal of Catalysis* **346**, 134–153 (2017).

[177] DeLuca, M., Kravchenko, P., Hoffman, A. & Hibbitts, D. Mechanism and kinetics of methylating C6–C12 methylbenzenes with methanol and dimethyl ether in H-MFI zeolites. *ACS Catalysis* **9**, 6444–6460 (2019).

[178] Jónsson, H., Mills, G. & Jacobsen, K. W. Nudged elastic band method for finding minimum energy paths of transitions. In Berne, B. J., Ciccotti, G., & Coker, D. F. (eds.) *Classical and Quantum Dynamics in Condensed Phase Simulations*, 385–404 (World Scientific, 1998).

[179] Henkelman, G., Uberuaga, B. P. & Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of Chemical Physics* **113**, 9901–9904 (2000).

[180] Heyden, A., Bell, A. T. & Keil, F. J. Efficient methods for finding transition states in chemical reactions: Comparison of improved dimer method and partitioned rational function optimization method. *The Journal of Chemical Physics* **123**, 224101 (2005).

[181] Larsen, A. H. *et al.* The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **29**, 273002 (2017).

[182] Pulay, P. Convergence acceleration of iterative sequences. The case of SCF iteration. *Chemical Physics Letters* **73**, 393–398 (1980).

[183] Bocus, M. & Van Speybroeck, V. Insights into the Mechanism and Reactivity of Zeolite-Catalyzed Alkylphenol Dealkylation. *ACS Catalysis* **12**, 14227–14242 (2022).

[184] The Nobel Prize in Chemistry 2010. `https://www.nobelprize.org/prizes/chemistry/2010/summary/` (Accessed: 2023-02-14).

[185] Vercammen, J. *et al.* Shape-selective C–H activation of aromatics to biarylic compounds using molecular palladium in zeolites. *Nature Catalysis* **3**, 1002–1009 (2020).

[186] Davies, D. L., Macgregor, S. A. & McMullin, C. L. Computational studies of carboxylate-assisted C–H activation and functionalization at group 8–10 transition metal centers. *Chemical Reviews* **117**, 8649–8709 (2017).

[187] Chen, B., Hou, X.-L., Li, Y.-X. & Wu, Y.-D. Mechanistic understanding of the unexpected meta selectivity in copper-catalyzed anilide C–H bond arylation. *Journal of the American Chemical Society* **133**, 7668–7671 (2011).

[188] Harvey, J. N., Aschi, M., Schwarz, H. & Koch, W. The singlet and triplet states of phenyl cation. A hybrid approach for locating minimum energy crossing points between non-interacting potential energy surfaces. *Theoretical Chemistry Accounts* **99**, 95–99 (1998).

[189] Car, R. & Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Physical Review Letters* **55**, 2471 (1985).

[190] Marx, D. & Hutter, J. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods* (Cambridge University Press, 2009).

[191] Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* **52**, 255–268 (1984).

[192] Martyna, G. J., Klein, M. L. & Tuckerman, M. Nosé–Hoover chains: The canonical ensemble via continuous dynamics. *The Journal of Chemical Physics* **97**, 2635–2643 (1992).

[193] Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics* **101**, 4177–4189 (1994).

[194] Frenkel, D. & Berend, S. *Understanding Molecular Simulations (Second Edition)* (Academic Press, 2002).

[195] Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **185**, 604–613 (2014).

[196] consortium, T. P. Promoting transparency and reproducibility in enhanced molecular simulations. *Nature Methods* **16**, 670–673 (2019).

[197] Valsson, O., Tiwary, P. & Parrinello, M. Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint. *Annual review of physical chemistry* **67**, 159–184 (2016).

[198] Sidky, H., Chen, W. & Ferguson, A. L. Machine learning for collective variable discovery and enhanced sampling in biomolecular simulation. *Molecular Physics* **118**, e1737742 (2020).

[199] Laio, A. & Parrinello, M. Escaping free-energy minima. *Proceedings of the National Academy of Sciences* **99**, 12562–12566 (2002).

[200] Bussi, G. & Laio, A. Using metadynamics to explore complex free-energy landscapes. *Nature Reviews Physics* **2**, 200–212 (2020).

[201] Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Physical Review Letters* **100**, 020603 (2008).

[202] Raiteri, P., Laio, A., Gervasio, F. L., Micheletti, C. & Parrinello, M. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *The Journal of Physical Chemistry B* **110**, 3533–3539 (2006).

[203] Torrie, G. M. & Valleau, J. P. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chemical Physics Letters* **28**, 578–581 (1974).

[204] Torrie, G. & Valleau, J. Monte Carlo study of a phase-separating liquid mixture by umbrella sampling. *The Journal of Chemical Physics* **66**, 1402–1408 (1977).

[205] Shirts, M. R. & Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics* **129**, 124105 (2008).

[206] Ensing, B., Laio, A., Parrinello, M. & Klein, M. L. A recipe for the computation of the free energy barrier and the lowest free energy path of concerted reactions. *The Journal of Physical Chemistry B* **109**, 6676–6687 (2005).

[207] Liu, P., Liu, Q., Liu, W., Peng, S. & Mei, D. Mechanistic insights into positional and skeletal isomerization of cyclohexene in the H-BEA zeolite. *Physical Chemistry Chemical Physics* **24**, 18043–18054 (2022).

[208] Chen, W. *et al.* Charge-separation driven mechanism via acylium ion intermediate migration during catalytic carbonylation in mordenite zeolite. *Nature Communications* **13**, 7106 (2022).

[209] Bal, K. M., Fukuhara, S., Shibuta, Y. & Neyts, E. C. Free energy barriers from biased molecular dynamics simulations. *The Journal of Chemical Physics* **153**, 114118 (2020).

[210] Bailleul, S. *et al.* Ab initio enhanced sampling kinetic study on MTO ethene methylation reaction. *Journal of Catalysis* **388**, 38–51 (2020).

[211] Dietschreit, J. C., Diestler, D. J. & Ochsenfeld, C. How to obtain reaction free energies from free-energy profiles. *The Journal of Chemical Physics* **156**, 114105 (2022).

[212] Dietschreit, J. C., Diestler, D. J., Hulm, A., Ochsenfeld, C. & Gómez-Bombarelli, R. From free-energy profiles to activation free energies. *The Journal of Chemical Physics* **157**, 084113 (2022).

[213] Anderson, J. B. Statistical theories of chemical reactions. Distributions in the transition region. *The Journal of Chemical Physics* **58**, 4684–4692 (1973).

[214] Bennett, C. H. Molecular dynamics and transition state theory: the simulation of infrequent events. In *Algorithms for Chemical Computations*, 63–97 (Americal Chemical Society, 1977).

[215] Chandler, D. Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *The Journal of Chemical Physics* **68**, 2959–2970 (1978).

[216] ThermoLIB | Center for Molecular Modeling. `https://molmod.ugent.be/software/thermolib` (Accessed: 2023-02-15).

[217] Frenkel, D. & Berend, S. Chapter 16 - Rare events. In *Understanding Molecular Simulations (Second Edition)*, 431–464 (Academic Press, 2002).

[218] Baron, P. Chapter 13 - Reactive flux. In *Reaction Rate Theory and Rare Events Simulations*, 335–362 (Elsevier, 2017).

[219] Bučko, T., Chibani, S., Paul, J.-F., Cantrel, L. & Badawi, M. Dissociative iodomethane adsorption on Ag-MOR and the formation of AgI clusters: an ab initio molecular dynamics study. *Physical Chemistry Chemical Physics* **19**, 27530–27543 (2017).

[220] Bolhuis, P. G., Chandler, D., Dellago, C. & Geissler, P. L. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annual Rreview of Physical Chemistry* **53**, 291–318 (2002).

[221] Chen, K. *et al.* Trace water amounts can increase benzene H/D exchange rates in an acidic zeolite. *Journal of Catalysis* **351**, 130–135 (2017).

[222] De Wispelaere, K. *et al.* Insight into the Effect of Water on the Methanol-to-Olefins Conversion in H-SAPO-34 from Molecular Simulations and in Situ Microspectroscopy. *ACS Catalysis* **6**, 1991–2002 (2016).

[223] Bailleul, S., Rogge, S. M., Vanduyfhuys, L. & Van Speybroeck, V. Insight into the Role of Water on the Methylation of Hexamethylbenzene in H-SAPO-34 from First Principle Molecular Dynamics Simulations. *ChemCatChem* **11**, 3993–4010 (2019).

[224] Bates, J. S., Bukowski, B. C., Greeley, J. & Gounder, R. Structure and solvation of confined water and water–ethanol clusters within microporous Brønsted acids and their effects on ethanol dehydration catalysis. *Chemical Science* **11**, 7102–7122 (2020).

[225] Liu, P., Hao, W. & Mei, D. Understanding the Effects of Water Molecules on Cyclohexanol Dehydration over Zeolitic Acid Sites. *The Journal of Physical Chemistry C* **125**, 15283–15291 (2021).

[226] Eckstein, S. *et al.* Influence of hydronium ions in zeolites on sorption. *Angewandte Chemie International Edition* **58**, 3450–3455 (2019).

[227] Bocus, M., Vanduyfhuys, L., De Proft, F., Weckhuysen, B. M. & Van Speybroeck, V. Mechanistic Characterization of Zeolite-Catalyzed Aromatic Electrophilic Substitution at Realistic Operating Conditions. *JACS Au* **2**, 502–514 (2022).

[228] Pérez de Alba Ortíz, A., Tiwari, A., Puthenkalathil, R. C. & Ensing, B. Advances in enhanced sampling along adaptive paths of collective variables. *The Journal of Chemical Physics* **149**, 072320 (2018).

[229] Grifoni, E., Piccini, G. & Parrinello, M. Microscopic description of acid–base equilibrium. *Proceedings of the National Academy of Sciences* **116**, 4054–4057 (2019).

[230] Bregante, D. T. *et al.* The shape of water in zeolites and its impact on epoxidation catalysis. *Nature Catalysis* **4**, 797–808 (2021).

[231] Berger, F., Rybicki, M. & Sauer, J. Molecular Dynamics with Chemical Accuracy – Alkane Adsorption in Acidic Zeolites. *ACS Catalysis* **13**, 2011–2024 (2023).

[232] ChatGPT: Optimizing Language Models for Dialogue. `https://openai.com/blog/chatgpt/` (Accessed: 2022-12-06).

[233] Unke, O. T. *et al.* Machine learning force fields. *Chemical Reviews* **121**, 10142–10186 (2021).

[234] Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. Schnet–a deep learning architecture for molecules and materials. *The Journal of Chemical Physics* **148**, 241722 (2018).

[235] Schütt, K. T. *et al.* SchNetPack: a deep learning toolbox for atomistic systems. *Journal of Chemical Theory and Computation* **15**, 448–455 (2019).

[236] Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical Review Letters* **98**, 146401 (2007).

[237] Bocus, M. *et al.* Nuclear quantum effects on zeolite proton hopping kinetics explored with machine learning potentials and path integral molecular dynamics. *Nature Communications* **14**, 1008 (2023).

[238] Markland, T. E. & Ceriotti, M. Nuclear quantum effects enter the mainstream. *Nature Reviews Chemistry* **2**, 0109 (2018).

[239] Liu, Y. & Li, J. An accurate potential energy surface and ring polymer molecular dynamics study of the $Cl + CH_4 \rightarrow HCl + CH_3$ reaction. *Physical Chemistry Chemical Physics* **22**, 344–353 (2020).

[240] Bigeleisen, J. & Mayer, M. G. Calculation of equilibrium constants for isotopic exchange reactions. *The Journal of Chemical Physics* **15**, 261–267 (1947).

[241] Bigeleisen, J. & Wolfsberg, M. Theoretical and experimental aspects of isotope effects in chemical kinetics. In *Advances in Chemical Physics*, 15–76 (John Wiley & Sons, Ltd, 1957).

[242] Ma, S. & Liu, Z.-P. Machine learning potential era of zeolite simulation. *Chemical Science* **13**, 5055–5068 (2022).

[243] Erlebach, A., Nachtigall, P. & Grajciar, L. Accurate large-scale simulations of siliceous zeolites by neural network potentials. *npj Computational Materials* **8**, 174 (2022).

[244] Sours, T. G. & Kulkarni, A. R. Predicting Structural Properties of Pure Silica Zeolites Using Deep Neural Network Potentials. *The Journal of Physical Chemistry C* **127**, 1455–1463 (2023).

[245] Batzner, S. *et al.* E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications* **13**, 2453 (2022).

[246] Vandenhaute, S., Cools-Ceuppens, M., DeKeyser, S., Verstraelen, T. & Van Speybroeck, V. Machine learning potentials for metal-organic frameworks using an incremental learning approach. *npj Computational Materials* **9**, 19 (2023).

[247] PsiFlow – interatomic potentials using online learning. `https://github.com/svandenhaute/psiflow` (Accessed: 2023-02-21).