Theoretical Studies on the Adsorption of Phosphoric Acid on γ-Al2O3 at the Solid Liquid Interface

Adrian Hühn

To cite this version:

Adrian Hühn. Theoretical Studies on the Adsorption of Phosphoric Acid on γ-Al2O3 at the Solid Liquid Interface. Catalysis. Université de Lyon, 2021. English. NNT : 2021LYSEN072 . tel-03934687

HAL Id: tel-03934687

https://theses.hal.science/tel-03934687

Submitted on 11 Jan 2023

Numéro National de Thèse : 2021LYSEN072

THESE DE DOCTORAT DE L'UNIVERSITE DE LYON

opérée par
l'Ecole Normale Supérieure de Lyon

Ecole Doctorale N°206
École Doctorale de Chimie (Chimie, Procédés, Environnement)

Discipline : Chimie

Soutenue publiquement le 25/11/2021, par :
Adrian HÜHN

Theoretical Studies on the Adsorption of Phosphoric Acid on γ-Al_2O_3 at the Solid-Liquid Interface

Études théoriques sur l'adsorption de l'acide phosphorique sur γ-Al_2O_3 à l'interface solide-liquide

---

Devant le jury composé de :

MILET, Anne	Professeure des Universités	Université Grenoble Alpes	Rapporteure
COSTA, Dominique	Directrice de Recherche	Chimie ParisTech PSL	Rapporteure
FLEURAT-LESSARD, Paul	Professeur des Universités	Université Bourgogne Franche Comté	Examinateur
GERVAIS, Christel	Professeure des Universités	Sorbonne Université	Examinatrice
CORRAL VALERO, Manuel	Ingénieur Recherche	IFP Energies nouvelles	Examinateur
MICHEL, Carine	Chargée de Recherche CNRS	ENS de Lyon	Directrice de thèse
RAYBAUD, Pascal	Ingénieur Recherche	IFP Energies nouvelles	Co-encadrant

Contents

List of Figures

List of Tables

Acknowledgements

First, I want to thank my thesis supervisors Carine Michel, Pascal Raybaud and Manuel Corral Valero for all the support and advice. Without your dedicated involvement this work would have never been possible. Thank you, Carine, for your tireless encouragement and your engaged focus on everyone's strengths. I am grateful to encounter your mindset. Thank you, Pascal, for sharing your experience and incredibly detailed knowledge with me. You knew every detail to enrich a discussion and I always enjoyed your enthusiasm. Thank you, Manuel, for calm guidance and your great patience with me on a personal level. You helped me a lot to communicate technical details in an understandable manner and allowed me to substantially improve my discussions. I honestly humbles me having had the chance to work with you. Having three supervisors can certainly be complicated, so I consider myself very lucky to have caught this great team.

To the jury of my defence: I want to thank Anne Milet and Dominique Costa, who agreed to evaluate this thesis, as well as Paul Fleurat-Lessart and Christel Gervais for taking their time to participate as jury members. As expected for an independent jury, we do not know each other allowing you to examine my work without any conflict of interest. Because of this, I appreciate the time you are going to spend for this and hope I can sustain the initial interest you had in my work.

I also am greatly indebted to Dorothea Wisser, who carried out the complex NMR experiments which were crucial in this work, and to Tao Jiang who introduced me to machine learning methods and explored various different options for NMR predictions training on the basis of DFT calculations. For all the productive discussions in the past three years I am grateful to Jérôme Rey, Daniel Walden, Thomas Pigeon, Teddy Roy, Stephan Steinmann, David Loffreda and certainly many other people.

Not least I want to thank my friends and colleagues I made here, including my reliable friends Amit Sahu and Sharmin Sharna who gave me great company and personal support over the last three years, and especially my colleagues Sarah Blanck and Laureline Treps who helped me on every request with official paperwork in the French language. I must also thank Malika, my wife and companion, who supported and encouraged me and who carefully read this work for additional corrections.

Finally, I also have to thank the ROAD4CAT industrial chair, part of the project IDEXLYON, as well as ENS Lyon for funding my position and this work.

Abstract

Phosphate adsorption on oxides is of significant interest in heterogeneous catalysis. The CoMoS/γ-Al_2O_3 hydrodesulfuration catalyst which is used at the scale of petroleum refinement, uses phosphate additives for preparation. Its preparation follows the stages of wet incipient impregnation, drying and activation by thermal treatment. Understanding the chemistry involved in these processes requires models of the oxide-water and oxide-air interfaces at different temperatures, which are also of interest in different fields. Oxide surfaces like γ-Al_2O_3 are by themselves very complex, which renders the characterization of surface phosphate species even more intricate.

Yet by today, improved NMR spectroscopic methods complemented with proven computational models allowed us now to mold an atomistic model of phosphate speciation. The present work starts with the creation of a systematic database by exploring more than 1000 geometries at different adsorption modes, adsorption sites, degrees of coverage and an optimized hydrogen bond environment for two polyphosphate species. For the most stable results magnetic shielding was calculated at DFT level. Good agreement was achieved with ^{31}P NMR experiments at drying conditions. Liquid conditions were simulated using metadynamics and showed a different set of most stable species. The explored reaction energies and kinetic barriers consistently show that elevated temperatures are needed to form the experimentally observed species. A ^{31}P NMR prediction model based on machine learning of chemical shifts from the local structure was developed to access chemical shifts in dynamic simulations.

Resumé

L'adsorption des phosphates sur les oxydes présente un intérêt significatif pour les catalyseurs hétérogènes. Le catalyseur d'hydrodésulfuration CoMoS/γ-Al_2O_3 qui est utilisé à l'échelle du raffinage du pétrole, utilise des additifs phosphatés pour sa préparation. Sa préparation suit les étapes d'imprégnation humide, de séchage et d'activation par traitement thermique. La compréhension de la chimie impliquée dans ces processus nécessite des modèles des interfaces oxyde-eau et oxyde-air à différentes températures, qui présentent également un intérêt dans différents domaines. Les surfaces d'oxyde comme la γ-Al_2O_3 sont en elles-mêmes très complexes, ce qui rend la caractérisation des espèces de phosphate de surface encore plus compliquée.

Pourtant, aujourd'hui, les méthodes améliorées de spectroscopie RMN complétées par des modèles de calcul éprouvés nous ont permis d'élaborer un modèle atomique de spéciation des phosphates. Le présent travail commence par la création d'une base de données systématique en explorant plus de 1000 géométries à différents modes d'adsorption, sites d'adsorption, degrés de couverture et un environnement de liaison hydrogène optimisé pour deux espèces de polyphosphates. Pour obtenir les résultats les plus stables, le blindage magnétique a été calculé au niveau DFT. Un bon accord a été obtenu avec les expériences de ^{31}P RMN dans des conditions de séchage. Les conditions liquides ont été simulées en utilisant la métadynamique et ont montré un ensemble différent d'espèces les plus stables. Les énergies de réaction et les barrières cinétiques explorées montrent systématiquement que des températures élevées sont nécessaires pour former les espèces observées expérimentalement. Un modèle de prédiction ^{31}P RMN basé sur l'apprentissage automatique des déplacements chimiques à partir de la structure locale a été développé pour accéder aux déplacements chimiques dans les simulations dynamiques.

Introduction

Catalysts consist of active sites at which the reactions take place that would not occur otherwise due to considerable energy barriers. For a catalyst to be economically viable, it has to be ensured that the amount of active sites is maximized. In heterogeneous catalysts, this is typically done by dispersing the active phase on a support with a high specific surface area. The preparation of these catalysts is delicate and demands well-controlled processing steps. It starts from a suitable support material which is impregnated with a solution containing, for example, metal precursors and additives. The impregnated support is then dried to remove the solvent and activated by thermal treatment, whereby catalytically active sites are formed.

The advantage of heterogeneous catalysts is that the support can be a cheap mineral or oxide like alumina or silicates which carries small amounts of a highly dispersed active phase. However, it also means that several components are involved in the course of a reaction – at least the support, active phase and reactant. On top of this, several additives are commonly used during preparation. At the mesoscale, material transport within the support influences the preparation as well as the performance of the final catalyst. Much of the knowledge about the preparation of heterogeneous catalysts is based on empirical knowledge and it is difficult to rationalize the effect of each component without accurate models. For example, several organic and inorganic compounds, e. g. phosphoric acid, are known to improve the final catalyst, but it is yet unclear what the effect of those additives is during preparation and how they can impact the catalytic performance. Moreover, organic additives are typically removed during the drying or activation steps of preparation, while phosphates would remain as part of the catalyst. This raises more questions about the actual role of phosphates in every preparation step and the final catalyst.

An important example for such a catalyst is the γ-Al_2O_3-supported cobalt molybdenum sulfide catalyst (CoMoS/γ-Al_2O_3) that is used in petroleum refinement for hydrodesulfurization (HDS) and hydrodenitrogenation (HDN) treatments.[1–3] At this scale, it is critical to optimize the efficiency of the catalyst as well as the use of materials which also includes ingredients for the catalyst. More aspects such as aging determine the overall life span of the catalyst. An optimum phosphate concentration at the oxide support must be targeted to avoid the formation of undesirable and inactive phases during activation.[2] Precise control over the phosphorus content and its dispersion on the catalytic support will not only impact the catalyst performance but can also minimize the consumption of phosphates.

Van Cleve et al. have demonstrated that phosphonate deposition increases the hydrothermal stability of γ-Al_2O_3, hindering the phase transition to boehmite and thus allowing the use of this support in aqueous phase reforming processes.[3] Phosphates may be added to the impregnation solution in order to modify the interplay between various metal ions in the solution and their behavior inside the alumina pores.[1] Here, the authors argue that phosphorus influences the interaction of the metallic precursors with the support and impacts the dispersion, size and morphology of the active phase.[4] More recently, van Haandel et al. reported that the activity of hydrodesulfurization CoMoS catalysts was improved by adding phosphate since it reduced the formation of the undesired Co_9S_8 phase.[5] This is again showing that phosphorus impacts the synthesis of the active phase.

Van Veen et al. suggest that an undesirable aluminium phosphate phase appears at high phosphate loading, when only using phosphates and molybdates for the catalyst preparation.[6,7] However, those are added together with other additives and metal precursors during catalyst preparation, leading to a much more complex multicomponent mixture. Combined spectroscopic studies (Raman, IR, UV/vis and Raman/NMR) show that the phosphate–γ-Al_2O_3 interaction is most critical.[8,9] Phosphoric acid is usually co-impregnated with other salts during catalyst preparation and Bergwerff et al. show that the speciation of these co-impregnation solutions can be extremely complex.[1,8] In some cases, the chemical equilibrium can shift towards the formation of precursor species of aluminium phosphates at the surface of alumina.[10–12] Moreover, it was shown that phosphate interacts more strongly with the alumina surface than molybdenum species by various characterization techniques such as spatially resolved Raman, UV/vis and NIR spectroscopy as well as by Magnetic Resonance Imaging combined to Streamline Raman Imaging.[8,9]

According to the results of *in situ* EXAFS experiments by Nicosia et al.,[13] phosphates increase the saturation of active edges in CoMoS/γ-Al_2O_3 catalysts. The authors rationalize their findings with support deactivation by phosphates. Several characterization studies have already been devoted to the intricate connection between the speciation of phosphate on the surface of the support as function of concentration, and its beneficial or deleterious effects in catalysis. X-ray absorption spectroscopy revealed that the presence of phosphate modifies the local environment of Mo at the metal-support interface.[13] However, no direct insight into the phosphate local environment could be deduced from those studies. Although IR-based experiments on dry samples and samples in solution demonstrate an interaction between phosphate anions and γ-Al_2O_3 carriers, it remains difficult to unambiguously assign the adsorption mode of phosphate anions.[14–16] At the H_2O/corundum interface, it was proposed that phosphate binds to corundum in a bidentate adsorption mode (two phosphate oxygen bind to the surface).[17] These studies rely on the comparison against the experimental spectra of known reference samples or against DFT studies, but this assignment remains difficult to achieve in the case of adsorbed phosphate anions, since the vibrational frequencies cannot be deconvoluted unambiguously.

Computational models of the atomic scale of the materials can be of great help to disentangle the findings of these studies. However, in order to do this, these models must be developed in close connection with experiments. It is in fact straightforward to construct a few model geometries and find patterns within them that match some of the experimental observations. The problem with this approach is that it might be biased towards the experimental observation, and this is ultimately not the same as an open-ended systematic search. In contrast, it is much more complicated to demonstrate that the model is plausible from a theoretical and experimental perspective. Even more important, this approach paves the way for predictions of experimental results and suggestions for new experiments. In the future, it might allow for the generation of new catalyst designs by computational modelling. For this purpose, a catalyst model needs to include all relevant components, and it must be shown that it is likely to form at synthesis conditions, but – most importantly – it needs to be consistent with all the experimental observations. At the same time, the model design needs to be feasible for testing with the available computational resources. For the ambitious goal – to fully understand the heterogeneous catalyst – one needs to know *all* interactions between the individual components. This work intends to elaborate on one of these interactions, that between phosphates and γ-Al_2O_3 during impregnation and at drying conditions.

The first challenge in this endeavor is to construct a model that accurately reflects the experimental conditions. In order to avoid missing any potentially relevant structures and to keep the research open-ended, the initial search was focussed on finding a computational protocol that is feasible with the available resources, minimizes bias and is testable. For phosphate adsorption on an oxide surface, the number of adsorption sites largely depends on the complexity of the alumina surface. Moreover, phosphate itself is a complex compound that due to its hygroscopic nature has to be discussed in context of water content, and which can polymerize at drying conditions.

Chapter 3 gives a comprehensive insight into how more than 1000 geometries were explored by density functional theory (DFT) calculations. Clearly, the real surface species cannot be represented by a single reference adsorption state. However, it will be shown in this chapter, how a relatively small number of structures can be chosen unambiguously. The resulting geometries were selected by their free enthalpies. The second challenge is to verify those model structures with experimental results.

In chapter 4, chemical shifts were calculated for these selected geometries and compared with ^{31}P-MAS NMR experiments, that were done in collaboration with researchers at Centre de RMN à Très Hauts Champs (Université de Lyon) and IFP Energies nouvelles. This work is published in ref. [18]. It is also illustrated here, how well the most stable systems agree with experimental NMR spectra. Overall, excluding all but the most favorable systems is a straightforward method of information reduction and allows to extract conclusions by direct comparison.

It was also used later in chapter 5 in order to choose the adsorption mode for biased molecular dynamics calculations. The questions in this chapter are, how does phosphate behave under impregnation conditions? And which adsorption states are accessible by kinetics? Biased molecular dynamics (metadynamics) were used in order to test the adsorption and desorption mechanism as well as several more complex phosphorylation reactions between phosphates, water and the most important surface sites. It is not feasible to evaluate all geometries produced with the initial exploration, thus in chapters 3 to 5, trends are only identified within a hand full of structures. However, different evaluation methods permit the use of hundreds or all of the test structures and extract data in histograms and averaged structure parameters (bond lengths, angles, etc). Such methods, today often referred to as data mining and machine learning, allow to search for complex patterns in the data. The difference between these two methods is quite subtle as they are both used to give a rationale for certain trends and to obtain predictions. This aims at a third challenge, which is to identify general trends in the produced data. In chapter 6 we tested several chemical descriptors and machine learning methods to construct NMR chemical shift prediction schemes. This prediction scheme is built using the geometries from chapter 3 by identifying the complex structure-chemical shift relationships. Chemical shift estimates are of comparable accuracy as DFT but orders of magnitude faster, reducing the calculation time from several hours to fractions of a second. This would allow ^{31}P NMR predictions for larger adsorption systems and trajectories from molecular dynamics. Beyond the mere construction of a faster prediction system based on geometries, the presented machine learning methods can also give insight into the leading components that influence the chemical shift.

In its broader context, this work is based on ongoing scientific efforts that aim at an improved understanding of alumina surfaces and phosphate chemistry. Much work has been done in the area of solid state NMR spectroscopy, while experimental method development allows for more and more complex experiments. For computational modelling, a variety of bulk and surface models of transition alumina are available nowadays, some of them well established.[19–24] While several adsorption studies of phosphates and phosphonates on oxide surfaces exist, the mechanistic insights are still limited. Mechanistic studies of phosphates are much better covered in molecular biology and one of the ideas tested in this thesis is if their methods can be transferred to mineral surfaces. The following chapter gives an overview of what has been done in the area of phosphate chemistry and applications in general as well as adsorption in particular. That includes the currently available methods for sensitive ^{31}P NMR experiments and the current knowledge about reaction kinetics of several reactions involving phosphates.

1. State of the Art

This chapter will first present the role of phosphates and transition alumina in different research areas as well as applications in several industrial sectors. The following sections will give an overview of many fields where the different phosphate interactions are important, followed by a review of the phosphate chemistry relevant in this work and the currently available spectroscopic methods. In particular, recent improvements in surface enhanced NMR techniques permitted complex ^{31}P quantum correlation experiments and the generation of two-dimensional ^{27}Al$-^{31}$P NMR spectra.[18] The last sections review the current state of the available γ-Al_2O_3 bulk and surface models.

1.1. Importance of phosphate-mineral interactions

Phosphate interactions with very different environments play a critical role in natural and human-made materials but our understanding of these systems is still limited. These materials include catalyst supports and minerals, which are important in the chemical industry and agriculture, as well as macromolecules in biological systems. Most phosphorus-containing chemicals are derived from phosphates, in which phosphorus binds to four oxygen atoms. The simplest and most common phosphates are based on the orthophosphate anion PO_4^{3-} and its protonated derivative HPO_4^{2-}, $H_2PO_4^{-}$ and H_3PO_4. Orthophosphates are common in minerals and esters, typically with all protons substituted by cations (Ca^{2+} in apatite, Al^{3+} in berlinite) or organic moieties. Phosphorus is rather abundant (0.1 wt% of the Earth's crust, mostly in sediments of phosphate rock),[25,26] and is involved in countless applications and natural processes. Its use in fertilizers, agrochemicals, detergents, corrosion inhibitors, food additives and pharmaceuticals gives it a critical role in modern civilization. Due to its range and scale of usage, phosphorus is listed as a critical raw material by the European Commission.[27]

The synthesis of phosphorus chemicals is still wasteful and has a high potential for optimization towards more energy-efficient processes and minimization of by-products.[28] Moreover, phosphorus-based agrochemicals and detergents pose a yet incalculable risk to surface waters.[29,30] In particular, the assessment of the environmental impact depends on our understanding of adsorption and degradation processes on soil minerals. Agriculture is certainly the most ubiquitous sector where the interaction of soluble phosphates with aluminosilica in soil has an essential contribution to crop yields and fertilizer consumption. Torres et al. studied the bioavailability of phosphate, that is the availability of monomers (orthophosphate) in aqueous solution.[31] They found different rates for polyphosphate hydrolysis and adsorption depending on the soil and the involved polyphosphates. These bioavailable phosphates are free inorganic phosphates in groundwater and water bodies like lakes, rivers and oceans. Eutrophication, due to fertilizers from mined phosphates that accumulate in those water bodies, also becomes an emerging concern.[28] Its symptoms are in some cases treated with alumina compounds. Phosphate management of lakes is already done for more than 50 years by precipitation with $Al(OH)_3$.[32] Its success, however, depends a lot on the presence of other ions and aging characteristics which demands a well tuned, holistic strategy.[33–35] In this context it would certainly be of help to understand the adsorption kinetics of phosphates on $Al(OH)_3$. Another recently published study emphasized the importance of free phosphates in order to produce accurate Earth system models for climate change prediction.[36]

Phosphate compounds are also one of many standard compounds needed in large scale industrial processes. For heterogeneous catalysts based on alumina substrate, phosphoric acid is generally added during the impregnation step. It is reported that phosphoric acid increases the final catalytic performance in hydrodenitrogenation (HDN) as well as hydrodesulfurization (HDS) treatments.[4] This behaviour is rationalized as a surface effect with the commonly used γ-Al_2O_3 substrate as well as inducing changes to metalpolyanions in the impregnation solution. There is a general agreement in the literature that phosphates and phosphonates strongly interact with γ-Al_2O_3 surfaces but it was also reported that phosphate is incorporated in dissolved metal precursors.[1,3,8]

The presence of phosphorus in biomolecules like ATP, biopolymers like DNA or biominerals such as tooth enamel, underlines its existential biological role.[37,38] Nucleic acids like RNA and DNA are phosphate condensation products, as well as adenosine polyphosphate like ATP and ADP for which there still remain many open questions regarding their reactivity, formation and contributions to biochemical processes.[39,40] A more fundamental question that may also involve mineral surfaces and phosphates is, which chemical processes may have lead to the first proto-biochemical structures on Earth.[41,42] According to Georgelin et al. a key piece for those early proto-biochemical phosphate condensation reactions could be reactions on alternating wet and dry silicate surfaces as produced at ocean shores by tides.[41] Kee et al. propose a reaction with a phosphite precursor in presence of Ca^{2+} ions.[42]

All these examples show research in very different contexts but the problems that were faced in them are similar throughout the areas of interest. Phosphates can strongly interact with mineral surfaces or even be a part of a mineral, while they simultaneously interact with several other components in the system, in solution or at dry surfaces. One integral part of understanding any of those systems is to understand the role of phosphates as adsorbate species. Based on accurate phosphate adsorption models, one can add further components depending on the system of interest. This strategy should enable one to eventually build comprehensive models and understand these complicated systems in industry or nature in much more detail.

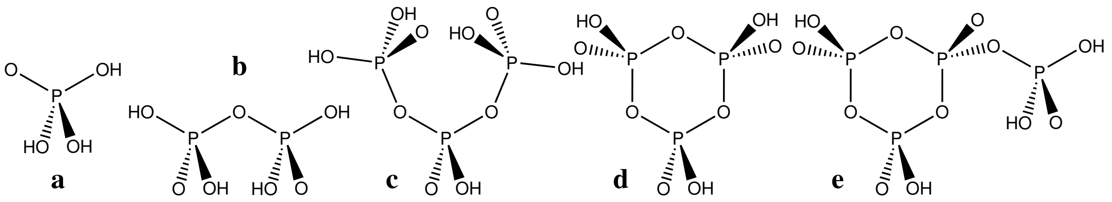


Figure 1.1.: Condensed phosphoric acid molecules: a) orthophosphoric acid, H_3PO_4; b) pyrophosphoric acid, $H_4P_2O_7$; c) triphosphoric acid, $H_5P_3O_{10}$; d) trimetaphosphoric acid, $H_3P_3O_9$; e) isotetrametaphosphoric acid, $H_4P_4O_{12}$

1.1.1. The condensed phosphates

Phosphates can assemble in polymer chains, rings and even form cross-links between those chains by condensation reactions, where for each link, one H_2O is removed. Pyrophosphate $H_4P_2O_7$ (fig. 1.1b) is the first polyphosphate that consists of more than one phosphate group, and besides orthophosphate (fig. 1.1a) the only polyphosphate with an individual retained name. Longer phosphate chains are, for example, triphosphate $H_5P_3O_{10}$ (fig. 1.1c) or tetraphosphate $H_6P_4O_{13}$. The chemical formula of such linear chains is $H_{n+2}P_nO_{3n+1}$ which is equivalent to $(HPO_3)_n(H_2O)$. HPO_3 is a metaphosphate or phosphoryl group (in this work it will always be referred to as metaphosphate).[43] Ring-shaped phosphate polymers are called polymetaphosphates based on the fact that the molar weights of their ions, $[PO_3]_n^{n-}$ are exact multiples of the metaphosphate group. The smallest stable polymetaphosphate consists of three condensed phosphates and forms a six-membered ring (fig. 1.1d). Their chemistry is similar to long polyphosphates, where the number of terminal groups is negligible, but long polyphosphates are not water soluble. Overall, they are less prone to hydrolysis which occurs preferably at the chain ends.[31,44] Moreover, long polyphosphate chains slowly decompose at 60°C, thereby also forming trimetaphosphate. Thilo theorizes that therefore the polyphosphate chains are folded or in a helical structure.[43] Even higher degrees of condensation exist in form of cross-linked polyphosphates where ring or chain phosphates are phosphorylated in a third direction.

Those are characterized by a local chemical environment where, instead of one or two P−O−P bridging bonds, three bridging bonds start from the same phosphorus atom. Consequently, these bonds are higher in energy and significantly less stable. They may occur in polyphosphate glasses at high temperatures and in absence of water. One small molecular representative of these phosphates is the phosphorus oxide which forms P_4O_{10} molecules, and the isotetrametaphosphate (fig. 1.1e) that is formed during hydrolysis of the former.[43] There is considerable experimental research done in the field of phosphate condensation, also including several accurate measurements of thermodynamic and kinetic data that were carried out over the last decades.[31,44–47]

1.1.2. Computational studies of phosphate condensation

From a computational perspective, phosphate condensation studies can involve a wide variety of condensation products. Early attempts (25 years ago) to computationally reproduce the reaction energy for pyrophosphate hydrolysis only showed qualitative agreement with experiments.[48] While accurate quantum mechanical methods were used (Møller–Plesset perturbation theory, MP2), the models only represented static gas phase molecules without entropy contributions. Nevertheless, these models still gave a first insight with structural information like accurate molecule geometries including intramolecular hydrogen bonds. A much more recent computational study by Tripathi et al. disentangles the hydrolysis reaction mechanism of guanosine triphosphate involving explicit solvent molecules in a mixed quantum mechanics/molecular mechanics (QM/MM) metadynamics approach.[49] Compared to the early models, their study includes the statistical dynamic properties and resolves free energy differences as well as barriers. Similar studies by Prasad et al. and Glaves et al. address the same hydrolysis reaction in different biochemical environments and discuss the mechanistic details, like concerted or stepwise reactions.[50,51] Adsorption studies with inorganic material interfaces are, in comparison to biochemical and organic chemical studies, still at their infancy, also because complex modelling techniques such as QM/MM are not always available due to the lack of parameters for heavy elements in force field parts. On alumina or oxidized aluminium surfaces, phosphate or phosphonate adsorption is commonly modelled as part of a range of compounds; for example, in context of lubricant or corrosion inhibition studies.[52–54] Hence, phosphate adsorption is embedded in context with several different adsorbing groups, but mechanistic details of adsorption were not in the scope of these studies. To my best knowledge, no studies exist that simultaneously include oxide adsorption and condensation phenomena of phosphates. Phosphate condensation reactions are in general phosphorylation reactions, but these reactions could not only occur *at* the alumina surface but also *with* the surface itself. The following section will give a comprehensive explanation of the challenges involved in phosphorylation reactions.

1.1.3. The phosphorylation reaction

In a simplified context, phosphates may be considered as constant structural PO_4 units which is justified in perspective of the strong P−O bonds involved. However, several elementary reactions involve P−O bond breakage and formation. Among those are polyphosphate condensation and hydrolysis as well as formation of phosphate esters, both of which are critical reactions in biochemistry. In the context of surface speciation, phosphorylation reactions have not yet been investigated. However, they might be crucial in order to explain the formation of some surface species as during some of the preparation steps, the reaction conditions are likely to allow these reactions.

Phosphorylation reactions have some noticeable similarities with nucleophilic substitution reactions known in organic chemistry. A metaphosphate group (PO_3^-) is moved from one moiety to the other. With an attacking group Y and a leaving group X, this can be written as:

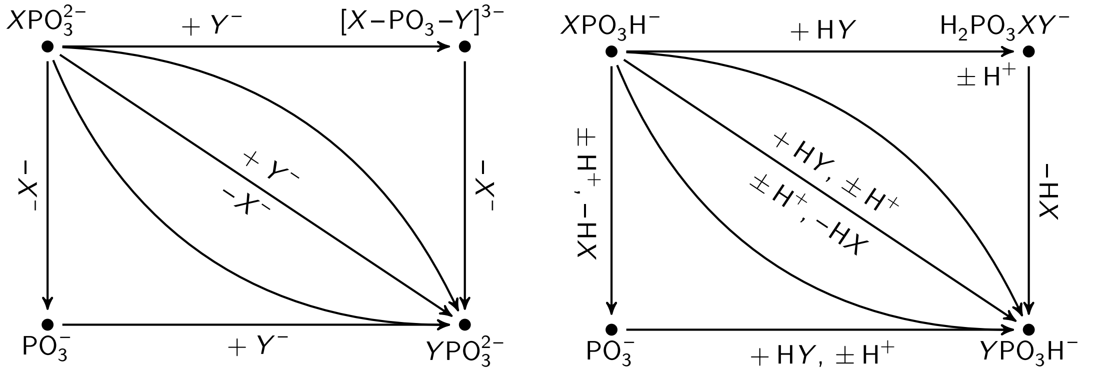
$$X\mathrm{PO}_3^{2-} + Y^- \longrightarrow X^- + Y\mathrm{PO}_3^{2-} \tag{1.1}$$

For simpler equations, X and Y are assumed to be anionic moieties with one negative charge, and the metaphosphate moiety to be deprotonated. These conditions occur in strong alkaline environments. In neutral or acidic environments, the reaction could be:

$$X\mathrm{PO_3H^-} + \mathrm{H}Y \longrightarrow \mathrm{H}X + Y\mathrm{PO_3H^-} \qquad (1.2)$$

with X and Y as well as the metaphosphate protonated. By substituting $X = \mathrm{OH^-}$ and $Y = \mathrm{HPO_4^{2-}}$ in eq. 1.1 or 1.2, the equations describe pyrophosphate condensation; swapping X and Y results in pyrophosphate hydrolysis. Instead, one can also substitute Y with a surface oxygen site, which would lead to the phosphorylation of that site. This would allow to add phosphates in those positions where a substitution of oxide ions would be otherwise unlikely. The protonation states in eq. 1.1 and 1.2 are a necessary simplification to express reaction equations, since the actual protonation states are generally not known. In fact, proton transfer reactions are critical components for understanding phosphorylations.[40]

Most importantly, another reaction intermediate is possible. While nucleophilic substitution reactions are well understood in terms of eliminations (S_N1) or concerted reactions (S_N2), phosphorylations also allow for initial additions. In fact, stable penta- and even hexavalent oxo-phosphate complexes are well known.[55,56] This leaves many more possible reaction pathways than, for example, in case of nucleophilic substitutions. The overall reaction is either limited by the P-O bond breakage/formation or by proton transfer. Fig. 1.2 gives an overview of the range of possible reaction types. The two reaction schemes summarize the range of reactions and intermediates for P–O bond formation and cleavage as a starting point to systematically analyze the whole phosphorylation reaction as a function of two bonds. Ref. [40] and [51] use a third axis to extend fig. 1.2 for proton transfer. However, by including protonation states of $\mathrm{PO_3^-}$, X- and Y- separately, one needs to define up to three more axes in general, which corresponds to up to five



(a) Alkaline conditions **(b)** Neutral/acidic conditions

Figure 1.2.: Phosphorylation reaction between sites X and Y, where vertical and horizontal directions refer to bond cleavage and formation with each site, respectively. Either concerted reaction or two types of intermediates can be expected: the addition complex with pentacoordinated phosphorus (top right) or elimination with a metaphosphate intermediate (bottom left). In strong alkaline environments (a), the reaction is only driven by formation and cleavage of phosphorus bonds. In neutral or acidic conditions (b), the reaction also depends on proton transfer reactions that have to be considered in every reaction path.

dimensions for searching transition states and intermediates. According to ref. [51], it is also important to ask for the number of involved H_2O molecules in the proton transport, which adds yet another dimension to the reaction mechanism.

The most fundamental reaction is the hydrolysis of pyrophosphoric acid ($H_4P_2O_7$) to form two orthophosphoric acid molecules (H_3PO_4) in aqueous solution. Several experimental studies have produced thermodynamic and kinetic data for this hydrolysis reaction. The first deprotonation of orthophosphoric acid and the first two deprotonations of pyrophosphate occur rapidly, hence the hydrolysis of $H_2P_2O_7^{2-}$ to $2\,H_2PO_4^-$ is a reasonable assumption. At 25°C, the reaction free enthalpy and free enthalpy of activation of this hydrolysis are -32 and $119\,\mathrm{kJ\,mol^{-1}}$, respectively. To my best knowledge, the only mechanistic studies of this particular reaction were conducted in the 1990s, neglecting solvation and finite temperature effects.[48] Later studies based on molecular dynamics or

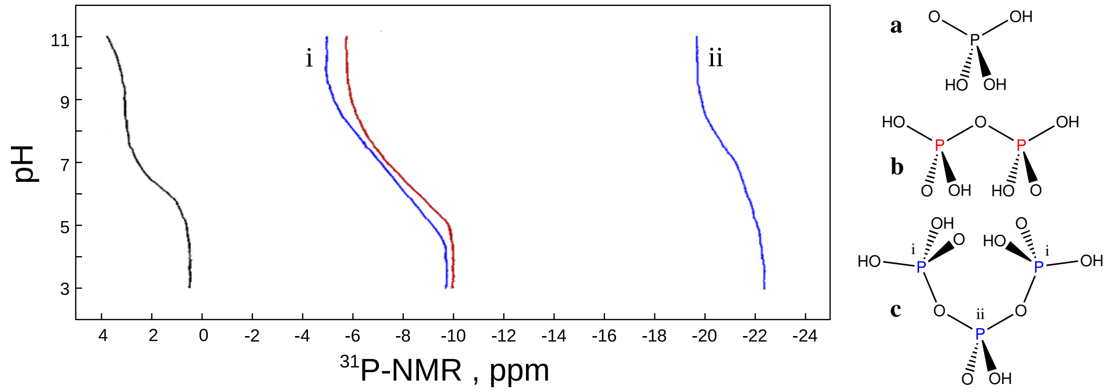


Figure 1.3.: Polyphosphate ^{31}P NMR chemical shift in aqueous solution depending on pH with respect to 85 % phosphoric acid. a: orthophosphoric acid, H_3PO_4 (black line); b: pyrophosphoric acid, $H_4P_2O_7$ (red line); c: triphosphoric acid, $H_5P_3O_{10}$ (two blue lines). Plot created with experimental data from Yoza et al.[57]

transition state search were focussed on specific, typically biochemical problems.[49–51] Thus, our understanding of the fundamental mechanism of phosphoric acid condensation in aqueous solutions is still limited. However, since thermodynamic data of the overall reaction is available, this condensation reaction is simulated as a performance benchmark in chapter 5.

1.1.4. Solid-state ^{31}P NMR of phosphates

Nuclear Magnetic Resonance (NMR) is one of the most powerful techniques to reach an atomic scale determination of phosphate interaction with alumina. It is an element-selective method which allows to draw direct conclusions about the chemical environment around the phosphorus atom. Yoza et al. produced ^{31}P NMR spectra for a broad range of phosphorus species in aqueous solution at a pH range from 3 to 11.[57] Fig. 1.3 shows their experimental results on the three smallest polyphosphates: orthophosphate, pyrophosphate and triphosphate. The signal of orthophosphate is at 0 ppm at high acidity and for increasing pH values the signal is shifted to slightly more positive values up to $3 - 5\,\mathrm{ppm}$. This is related to an increased shielding of the deprotonated species. The pyrophosphate and triphosphate signals are clearly shifted to negative values, and triphosphate produces two signals. Pyrophosphate consists of two groups connected by oxygen, which produces a signal at -10 ppm at pH=3. Triphosphate $H_5P_3O_{10}$ is the first polyphosphate with one chain group connecting its two terminal groups: $H_2O_3P{-}O{-}(HPO_2){-}O{-}PO_3H_2$. This leads to another signal for the central phosphate group at about -22 ppm (pH=3). The shift is related to a deshielding when substituting hydroxy groups in orthophosphate by one or two more electronegative phosphate groups. As for orthophosphate, signals by terminal and chain phosphate groups are gradually shifted to positive ppm values at higher pH as a consequence of deprotonation.

NMR spectroscopy on solids can be evaluated in similar ways as solution NMR. However, compared to solution NMR, the anisotropy of the chemical environment deteriorates the signal resolution of solid state NMR. The rapid molecular movement in solution produces an average isotropic environment where momentary anisotropy is “averaged out”. In contrast, the orientation of the anisotropic chemical environment is constant in a dry powder sample. As a consequence, the solid state NMR spectrum is an ensemble of random anisotropic directions instead of well resolved averages. Orientation averaging is, however, possible by rotating the probe i. e. by rapidly spinning the probe head. Using an angle of about 55° relative to the magnetic field vector, all directions in three-dimensional space are transformed into one another (c. f. fig. 1.4), provided that the spinning frequency is fast enough. This can best be understood by imaging a cube spinning on its vertex. The angle between any edge of the cube and the surface normal, or equivalently the cube’s body diagonal, describes this magic angle. Fig. 1.4 provides another graphical explanation of the angle and spinning operation. The spinning method is called *magic angle spinning* (MAS) and, only by including this spinning operation, well resolved solid state spectra are accessible. Bleam et al. conducted MAS ^{31}P NMR experiments on

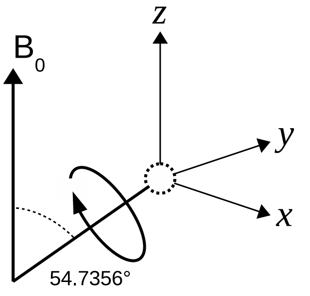
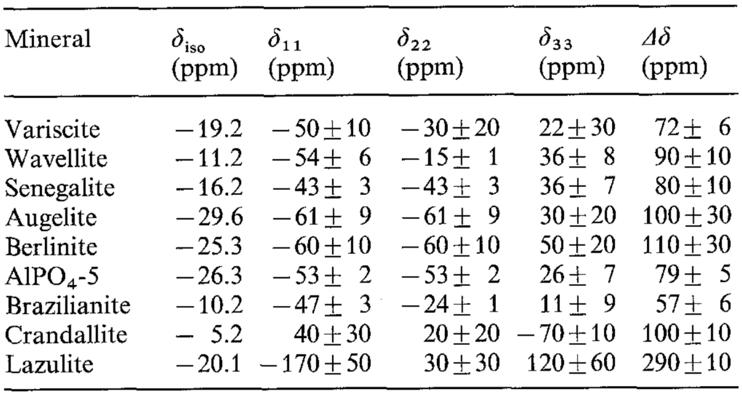


Figure 1.4.: Probe spinning geometry in MAS NMR spectroscopy. The probe spins around a defined axis with respect to the magnetic field B_0.

Table 1.1.: ^{31}P NMR isotropic chemical shifts (δ_{iso}) and principal components of chemical shift tensors (δ_{ii}, $i = 1, 2, 3$) and the anisotropy $\Delta\delta$ for several aluminium phosphate minerals. $\Delta\delta = \delta_{33} - \delta_{11}$, $\delta_{11} < \delta_{22} < \delta_{33}$. Table and caption (rephrased) from Bleam et al.[58]



Mineral	δ_{iso} (ppm)	δ_{11} (ppm)	δ_{22} (ppm)	δ_{33} (ppm)	$\Delta\delta$ (ppm)
Variscite	−19.2	−50±10	−30±20	22±30	72± 6
Wavellite	−11.2	−54± 6	−15± 1	36± 8	90±10
Senegalite	−16.2	−43± 3	−43± 3	36± 7	80±10
Augelite	−29.6	−61± 9	−61± 9	30±20	100±30
Berlinite	−25.3	−60±10	−60±10	50±20	110±30
$AlPO_4$-5	−26.3	−53± 2	−53± 2	26± 7	79± 5
Brazilianite	−10.2	−47± 3	−24± 1	11± 9	57± 6
Crandallite	− 5.2	40±30	20±20	−70±10	100±10
Lazulite	−20.1	−170±50	30±30	120±60	290±10

several aluminium phosphate minerals and estimated the anisotropy of the signal.[58] The principal components of the anisotropy tensor (tab. 1.1) vary over dramatic ranges of 50 to 100 ppm while the isotropic shift range of the presented minerals is only up to 20 ppm. The error estimates of isotropic chemical shifts (±0.1 ppm) are much lower than anisotropic components.

1.1.5. Surface-sensitive ^{27}Al and ^{31}P NMR

For signal enhancement, cross-polarization (CP) experiments are performed, where an abundant nucleus (usually ^{1}H) is polarized, and which is brought into spin-thermal contact via dipolar coupling. DeCanio et al. recorded ^{27}Al and ^{31}P NMR spectra, with and without CP, of phosphate on γ-Al_2O_3 after impregnation with an aqueous solution of phosphoric acid and then drying at 120°C.[12] A single broad peak is observed in the $\{^{1}H\}^{31}$P CP NMR spectra in fig. 1.5c, which is shifted to lower ppm with higher phosphate loading. The authors attribute this peak shift to an overlap of orthophosphate,

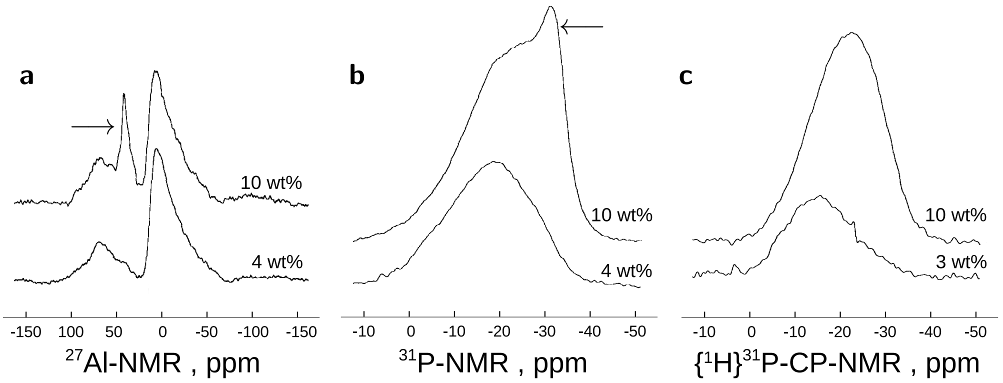


Figure 1.5.: Direct polarization ^{27}Al- and ^{31}P-MAS NMR and cross polarization NMR of dried phosphate species on γ-Al_2O_3. Plot created with experimental data from DeCanio et al.[12]

pyrophosphate and an increasing amount of other short-chained polyphosphate species with higher loading. At higher loading between 4 and 10 wt% phosphate, a distinct peak occurs at -32 ppm in the direct polarization ^{31}P spectra in fig. 1.5b. This indicates the formation of a water-free layer of aluminium phosphate, which does not contain protons and is thus cannot be observed in the $\{^1H\}^{31}P$ CP NMR spectra. In the direct polarization ^{27}Al spectra (see fig. 1.5a), the onset of formation of amorphous aluminium phosphate can be seen starting at 4 wt% phosphate by the appearance of a peak at 38 ppm. With a higher phosphate content, a narrower peak at 40 ppm, assigned to the crystalline aluminium phosphate, appears.

A similar trend in ^{31}P CP NMR spectra was observed by van Eck et al. studying the adsorption from an ammonium phosphate solution on γ-Al_2O_3 (230 m^2/g) dried at 110 °C.[59] Using $^{31}P\{^{27}Al\}$ REDOR and TRAPDOR experiments, the authors also suggest the formation of both polyphosphates and a layer of aluminium phosphate. Despite the insights gained by those studies, there are still fundamental questions concerning the structure of the adsorbed species. For example, DeCanio et al. demonstrate that phosphates adsorb predominantly in the form of monomeric species at phosphate loading below 1 wt%, which corresponds to a coverage of approximately 0.3 monolayers.[12] For higher phosphate loading it remains difficult to quantify at which point these species form oligomers and precursors of aluminium phosphate. Moreover, the local structure at the surface is still unknown. Yet, the dentation modes of phosphates adsorbed on alumina as well as changes to the surface could give very useful insights about the role of these species in catalysis.

Using surface signal enhancement techniques like Dynamic Nuclear Polarization Surface Enhanced NMR Spectroscopy (DNP SENS) can vastly increase the signal gain which improves the signal-to-noise ratio. DNP SENS has revolutionized solid-state NMR of surfaces as the improved gain enables complex correlation experiments, which would take too long to record with conventional methods. The idea is to add an exogenous diradical solution to the sample, thereby creating a large electronic polarization on the sample surface. This electronic polarization is then transferred to the nuclei by irradiating the sample with a high-power microwave source, which results in improved signal gain of up to a factor of 100.[60–63] In combination with CP, DNP SENS yields sufficient signal intensity to record two-dimensional ^{27}Al-^{31}P correlation spectra, as well as ^{31}P double quantum and triple quantum coherence experiments, to directly detect oligomers.[18] The determination of adsorption sites and modes at the surface of the alumina support is highly challenging due to a high degree of complexity of local sites. Computational modelling can be the key in order to complement experimental techniques.

1.1.6. Combined NMR and DFT studies

While the experimental work is critical to narrow down the requirements to the model, at this point none of the presented publications was able to produce a coherent and complete picture for phosphate-γ-Al_2O_3 interaction. Recently, by the combination of DNP NMR and DFT calculations, it was possible to unravel the nature of Brønsted acid sites at the surface of amorphous silica-alumina.[64] Lately, it was also possible to determine the location of hydroxyl groups at the edges of γ-Al_2O_3 crystallites thanks to the combination of ^1H NMR experiments and DFT calculations of chemical shifts.[65] Those DFT calculations rely on complex γ-Al_2O_3 models detailed in the coming sections.

The requirement for these studies is an accurate prediction of NMR chemical shifts from DFT electronic structure calculations. For periodic calculations with pseudopotentials, the most common technique is the gauge-including projector-augmented wave (GIPAW) method,[66,67] that allows accurate chemical shift estimates for all atoms in the system. Many theoretical studies, devoted to the calculations of phosphorus NMR chemical shifts inside various bulk materials, provide an accurate assignment of phosphate sites in various minerals.[68–71] Wei Li et al. showed by combining REAPDOR NMR characterization and GIPAW calculations that bridging bidentate phosphates would be adsorbed at the interface of water/boehmite (γ-AlOOH).[72] Nevertheless, since the authors used a rather simplified Al2 dinuclear cluster model for describing the boehmite surface, this may provoke questioning about the transferability of the results to other types of aluminum oxides such as γ-Al_2O_3. More recently, F. Tielens et al. investigated by GIPAW NMR calculations the ^{17}O and ^{31}P spectroscopic features of phosphate adsorption at the H_2O—TiO_2(anatase) interface and identified the bidentate adsorption mode.[73]

While GIPAW gives quantitative predictions of NMR chemical shifts, it does not provide a rationale for its origin. Dawson and Ashbrook produced a set of empirical equations involving average $P-O$ distances and $P-O-Al$ angles.[74] Multivariate regression of a test set of DFT calculations on an aluminium phosphate model cluster was used for function fitting. The authors later extended and applied their method based on experimental data and a wider range of aluminium phosphates.[69,75] It is at the moment the only known model to map structural parameters of aluminium phosphates to NMR spectroscopic results.

GIPAW calculations require calculations at DFT level, with high demands towards precision of geometry and electronic structure. Thus, these simulations can be relatively expensive, in particular if the number of relevant nuclei is small (in the simplest case of orthophosphate adsorption, one electronic structure calculation produces one chemical shift). Empirical models, like Dawson and Ashbrooks, can also be a useful tool for reducing this computational expense. An early attempt to predict ^{31}P NMR chemical shifts by machine learning was done by West.[76] In 1993, he tried to train neural networks using a large database of experimental chemical shifts for phosphorus compounds, but the accuracy was not satisfactory. Recently, Chaker et al. tested several machine learning methods to predict NMR iso-shifts of ^{17}O, ^{27}Al and ^{29}Si in aluminosilicate glasses.[77] Their results are instructive in how to produce an accurate machine learning model with predictable error estimation relative to GIPAW results.

1.2. Building a γ-Al_2O_3 model

1.2.1. Characteristics of γ-Al_2O_3

Transition alumina is a versatile group of materials with well-characterized pore and particle sizes as well as high specific surfaces.[78] Among them, γ-Al_2O_3 is economically the most important one since it is used as a support in large scale industrial catalysis. It has a disordered oxide structure, as characterized by X-ray diffraction (XRD) and spectroscopic methods, which is a big obstacle for atomistic modelling. Nevertheless, several surface models have been tested and established in the past decades. These models allowed for the first time to build computational models to describe surface speciation on γ-Al_2O_3. In this work, γ-Al_2O_3 model surfaces were used in order to establish a robust understanding of phosphate speciation at different experimental conditions.

As one of the most common supports in heterogeneous catalysts, γ-Al_2O_3 is a critical material in large scale industrial processes.[2,78] The properties of γ-Al_2O_3 are best understood in the context of its other polymorphs. Alumina are oxides of the composition Al_2O_3 that can be synthesized by calcination from aluminium hydroxide $Al(OH)_3$ and aluminium oxide hydroxide AlOOH. Several configurations exist for all these minerals: bayerite and gibbsite are both $Al(OH)_3$ and the two most important oxide hydroxides are diaspore (α-AlOOH) and boehmite (γ-AlOOH). The only stable alumina configuration is corundum (α-Al_2O_3), but several metastable transition alumina are formed from one another by temperature-controlled solid phase transformations.[78,79] Partial pressure of water also influences the reaction process; even grinding and crushing can have an effect on the reaction as it improves water release for a given temperature and it might be a different energy source for phase transformations.[78,80]

γ-Al_2O_3 is formed as the first phase of boehmite dehydration. The sequence can be illustrated like this:

$$\text{boehmite} \xrightarrow{450\,^{\circ}\text{C}} \gamma\text{-Al}_2\text{O}_3 \xrightarrow{750\,^{\circ}\text{C}} \delta\text{-Al}_2\text{O}_3 \xrightarrow{1000\,^{\circ}\text{C}} \theta\text{-Al}_2\text{O}_3 \xrightarrow{1200\,^{\circ}\text{C}} \text{corundum}$$

From γ-Al_2O_3, the phase transitions to δ-Al_2O_3, θ-Al_2O_3 and γ-Al_2O_3 define the stability order, with γ-Al_2O_3 as the least stable phase in accordance with Ostwald's rule.[81] More transition aluminas (χ-, κ- and η-Al_2O_3) can be obtained by direct calcination of bayerite and gibbsite, forming different phase transition pathways.[78,79] All transition aluminas form θ- and α-Al_2O_3 at 1000°C. γ-, δ- and θ-Al_2O_3 all share a similar distorted *fcc* crystal lattice of oxygen atoms. While θ-Al_2O_3 is fully characterized and known to crystallize in a β-Ga_2O_3-type lattice, γ-, and δ-Al_2O_3 are disordered systems with irregularly distributed aluminium atoms. This was shown by various techniques including XRD, NMR, IR and Raman spectra.[79] They nevertheless show a certain degree of crystallinity, most notably a face-centered cubic (*fcc*) lattice of oxygen ions. ^{27}Al NMR results show that in γ-Al_2O_3, $25 - 31\,\%$ of aluminium ions are in tetrahedral sites,[78] and this proportion is increased to $35\,\%$ in δ-Al_2O_3. For γ-Al_2O_3, new evaluations show that the number of tetrahedral sites might be slightly underestimated, and that the amount of tetrahedral sites is more similar to δ-$\mathrm{Al_2O_3}$.[82] Initially, γ-$\mathrm{Al_2O_3}$ was described as a distorted spinel-like structure and the high-temperature δ-$\mathrm{Al_2O_3}$ as a less disordered, more spinel-like structure.[83] This early observation is consistent with the distribution of aluminium in tetrahedral and octahedral sites, but the authors did not try to rationalize the pronounced anisotropy of γ-$\mathrm{Al_2O_3}$, which would not occur in a cubic spinel structure. After calcination, both polymorphs also contain pentacoordinated aluminium atoms.[79,84] On a mesoscopic level, it is known that γ-$\mathrm{Al_2O_3}$ is a porous compound that typically provides surface areas of 180 to $240\,\mathrm{m^2\,g^{-1}}$. The disorder in those systems substantially compromises the efforts to describe these materials on an atomistic level. Nevertheless, in the last decades, several research groups proposed a range of models for the description of γ-$\mathrm{Al_2O_3}$ and δ-$\mathrm{Al_2O_3}$. The following sections provide an overview of the available γ-$\mathrm{Al_2O_3}$ bulk models, as well as on how to produce surface models from them and the current understanding about this system.

1.2.2. Bulk model system

Disorder is a problem for any kind of simulation as the atomic structure must be fully defined first. In other words, we need to decide where to place the atoms to create an approximate representation. For the γ-$\mathrm{Al_2O_3}$ structure we only know two bulk properties with certainty: the oxygen atoms are ordered in a distorted *fcc* lattice and the aluminium atoms are distributed in approximately 1/3 tetrahedral and 2/3 octahedral sites. Beyond that, structural features similar to spinels can be imposed from XRD intensities. This description already dates back to six decades ago.[83] Kinetic considerations and other aspects, like the topotactic transformation from its precursor boehmite, indicate that γ-$\mathrm{Al_2O_3}$ should have a structural relation to boehmite as well. The increase of computational capacities nowadays also allows for a systematic search for structures within a large configuration space. Structures based on all these assumptions have been explored, and within the dedicated literature, spinel and nonspinel models are typically distinguished.

A spinel is a crystal with the chemical formula $MgAl_2O_4$. It is a cubic oxide in which Mg^{2+} and Al^{3+} are in defined tetrahedral and octahedral sites. Many mixed oxide compounds are known with the same crystal structure but different cations, and it is also known that spinels are to some degree disordered due to swapped ion positions. To transfer the spinel structure of $MgAl_2O_4$ to Al_2O_3, Mg^{2+} are exchanged with 2/3 Al^{3+}, which leaves 'vacancies' for every third stoichiometric unit as in $Al_{\frac{8}{3}}\square_{\frac{1}{3}}O_4$. The primitive spinel unit cell contains two stoichiometric units of AB_2X_4; this is tripled for the smallest defective spinel structure, hence it contains six AB_2X_4 (42 atoms) units or eight Al_2O_3 units (40 atoms, without two defects), respectively.[19] The defect sites could be on tetrahedral or octahedral sites but *ab initio* calculations showed that octahedral defects are energetically favored, and tetrahedral sites are maximized, making up 37.5 % of all aluminium sites. As a slightly different approach, hydrogen spinels (HAl_5O_8) were tested as well.[85] Wolverton and Hass compared both models and concluded that HAl_5O_8 is thermodynamically unfavorable compared to the hydrogen-free models and boehmite but it might exist as a metastable phase.[19]

Based on a structural relation to boehmite, Krokidis et al. constructed a 'dehydrated boehmite' model by removing OH and H from its layered structure and manually arranging a structural collapse followed by identifying sensible tetrahedral site transitions.[22] The resulting structure is the first nonspinel structure (see fig. 1.6). They tested structures with 0 to 50 % tetrahedral sites; among them, those with 25 % lead to the most stable configuration, followed by those with 31 % and 37 % tetrahedral sites. The structure with 25 % was considered as the best case. This unit cell has a different shape and monoclinic symmetry, but almost the same cell volume as the defect spinel models with the same number of atoms.[86] In an exhaustive comparison of spinel and nonspinel models, Paglia et al. generated more than 100k defect spinel models by alternating defect sites of super cells containing 32 Al_2O_3.[23] The authors then compared their results with more than 500k systematically generated nonspinel models and concluded that, on average, 40 % of the aluminium atoms occupy nonspinel sites. Moreover, they ruled out spinel models based on the fact that their 100k generated variations generally produced mismatched diffraction patterns. This was in qualitative agreement with Krokidis' model, however since the latter is not based on a spinel unit cell, the number of spinel vs nonspinel sites cannot be clearly partitioned.

Paglia et al. and Ferreira et al. compared theoretical models from each group (defect spinel,[19,20] hydrogen spinel,[21] Krokidis nonspinel[22] and Paglia nonspinel[23]) with neutron diffraction experiments and ^{27}Al NMR spectroscopic data.[86,87] The model fitting to neutron diffraction data showed that Paglia's nonspinel models were in the best agreement with Rietveld refinement simulations, followed by Krokidis' nonspinel, defect spinel and hydrogen spinel in this order, but with overall comparable accuracies. ^{27}Al-MAS NMR iso-shifts showed similar results, but in particular the absence of a correct prediction of pentacoordinated aluminium atoms indicated that surface aluminium atoms are likely to contribute significantly to the experiments. It must be underlined, that these comparative studies neglect surface effects.[86,87] Even though this is generally a reasonable assumption, γ-Al_2O_3 in particular is known to have large specific surfaces, increasing the number of contributing surface atoms. For example, Łodziana et al. showed in 2004 that certain hydroxylated alumina surface models of the related θ-Al_2O_3 exhibit a slightly "negative" surface energy, which means that the surfaces are thermodynamically more stable than the bulk structure.[88] This is clearly impossible for clean surface cuts, however, in principle adsorbates (here: H_2O) can stabilize the surfaces. It must be noted here that the negative surface energy reported here concerned only one of several surface models, and at the time the computational models neglected more accurate descriptions of long range dispersion interactions which are available today.[89]

Busca summarized that none of the four computational γ-Al_2O_3 models can be rationally excluded on the basis of thermodynamic stability, since γ-Al_2O_3 is the first transition alumina of boehmite, and neither the γ-Al_2O_3 nor the δ-Al_2O_3 phase are fully characterized. A very stable model intended for γ-Al_2O_3 might rather be a good model δ-Al_2O_3.[79] These considerations are in favor of Krokidis' model, since it is the only model involving a conceptual dehydration of boehmite and thus a construction that includes a structural relation to its precursor. Krokidis' model is also considered as one of the most popular ones,[79,82] while, for example, hydrogen-containing γ-Al_2O_3 models are overall rejected.[79] Prins recently reviewed the current knowledge of the γ-Al_2O_3 structure.[82] It was pointed out that new ^{27}Al NMR experiments and the previously available ^{17}O NMR data indicate that higher percentages ($35 - 37.5$) of tetrahedral sites should be expected; this indicates that the existing nonspinel models have too few aluminium atoms in tetrahedral sites. On the other hand, the author discussed findings of a decreased number of tetrahedral sites in γ-Al_2O_3 catalysts. As a consequence, the concerned nonspinel models would still be useful to produce interface models. While Prins seems to disregard Krokidis' model based on his new analysis of NMR data, he acknowledged its successful application starting from Digne's hydrated surface studies,[24] followed by many computational studies in heterogeneous catalysis.[18,90–92] While the author has a valid concern regarding the underestimated fraction of tetrahedral aluminium atoms, it does not exclude the Krokidis nonspinel per se. As mentioned above, Krokidis et al. discussed a range of structures with up to 50 % tetrahedral sites, with models composed of slightly more tetrahedral alumina as additional candidates.

Generally, Busca and Prins concluded independently that structural disorder and limits in experimental methods did not yet allow for an atomic level characterization of γ-Al_2O_3.[79,82] Busca emphasized the disorder which can be demonstrated by multiple experimental techniques and suggested to focus on θ-Al_2O_3 which is better characterized for surface studies. Prins, on the other hand, recommended to evaluate resulting transition alumina phases by using new high-temperature synthesis routes: Zhang et al. observed an Al_2O_3 thin film formation after oxidation of NiAl alloys at 850°C, which they identified as γ-Al_2O_3.[93] While Prins also raised concerns about the unexpected stability and possible impurities of Ni stabilizing these phases, he suggested further analysis of these results as the thin-films are much larger than γ-Al_2O_3 crystallites synthesized from boehmite.

Ultimately, our knowledge of γ-Al_2O_3 model systems is inconclusive, and their usability and shortcomings are still discussed. Recent improvements regarding the characterization of δ-Al_2O_3 are, however, promising. Kovarik et al. revealed that δ-Al_2O_3 is composed of at least 4 phases that can intergrow in two crystallographic directions.[94] The authors supported their proposal by DFT models and comparison with electron microscopy data. Their results demonstrated how involved computational modelling along with experimental investigations allows for a comprehensive characterization of complex irregular phases such as in this case δ-Al_2O_3 – the structurally most similar polymorph to γ-Al_2O_3. It should also be noted, that the δ-Al_2O_3 phases are not based on a defect spinel structure.

1.2.3. Surface models

While the bulk structure of γ-Al_2O_3 is of fundamental interest, its applications as a high surface compound (i. e. as a catalyst substrate), raise even more interest in its surfaces. It is established that γ-Al_2O_3 crystallites are products of a topotactic dehydration of boehmite and that the most exposed facet is (1 1 0) with about 70 % (cf. fig. 1.9).[78] To construct surfaces from the γ-Al_2O_3 bulk model (fig. 1.6), the unit cell is cleaved along one crystallographic direction, and repeated several times along the new surface normal direction in order to mimic lower bulk layers, see fig. 1.7. Its thickness and the available vacuum over the surface become parameters that determine how accurate the surface is represented. However, this process is not straightforward as it has to be decided along which surface direction the surface is cut and which atoms are exposed on the

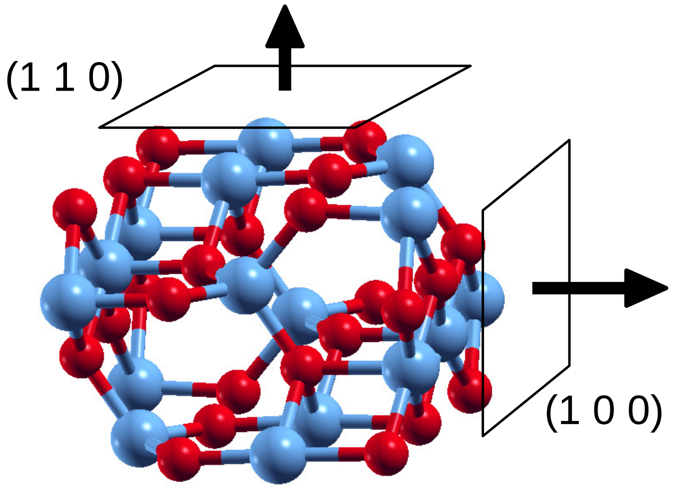


Figure 1.6.: Unit cell content of the γ-Al_2O_3 bulk model by Krokidis et al.[22]. Oxygen and aluminium atoms are colored in red and blue. Cuts along the horizontal lead to (1 1 0) model facets (illustrated as plane and normal vector), whereas cuts along the vertical produce (1 0 0) model facets. Note that the Miller indices are with respect to the *fcc* lattice.

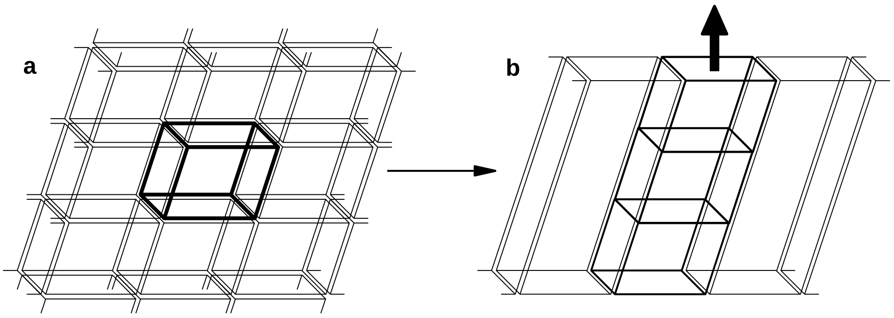


Figure 1.7.: Cutting a surface slab from a bulk unit cell (a). Several copies of the bulk unit cell produce a surface slab (b) with a thickness that depends on the bulk geometry.

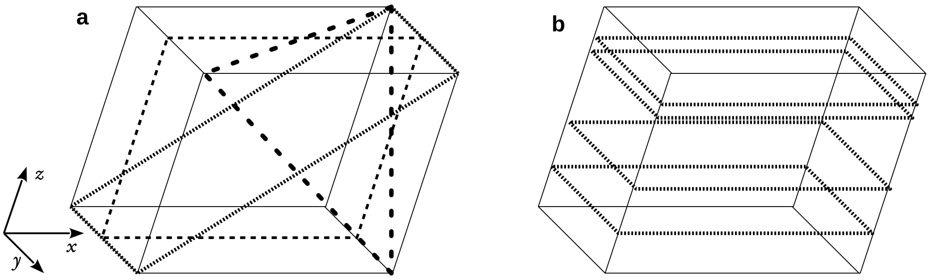


Figure 1.8.: Cutting a surface slab along arbitrary surface directions in (a): (0 1 0) with small dashed line, (1 0 1) with dotted line, (1 1 1) with rounded dashed line. After choosing the Miller indices, the slab can be cut along different layers (b), which lead to different surface terminations.

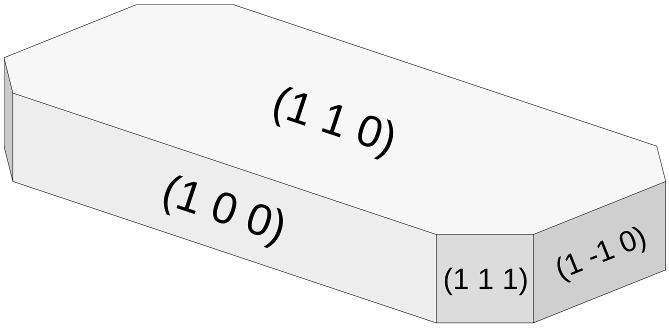


Figure 1.9.: Shape of a γ-Al_2O_3 platelet after topotactic transformation from boehmite. The (1 1 0) and (1 0 0) facets are most exposed. Miller indices are given with respect to the oxygen *fcc* lattice (often referred to as "spinel" symmetry).

final surface model. Fig. 1.8 illustrates the decisions one has to make. For example, the unit cell of Krokidis' model (ball and stick model in fig. 1.6), was used to build surface slabs for the (1 0 0) and (1 1 0) facet. In the (1 1 0) facet, the topmost layer always consists of four aluminium atoms and six oxygen atoms, however, the atoms in the next layer are differently arranged than the first. This alternative layer is topmost if a different layer is chosen as illustrated in fig. 1.8. Consequently, the Krokidis bulk model can be used to produce a range of different surface models. Most surface cuts along these directions would, however, lead to asymmetric or nonstoichiometric models, because top and bottom terminations are different, or the resulting slab cut has a different stoichiometry than the bulk model. With these considerations, Digne et al. in 2002 identified useful model surfaces from Krokidis' bulk model and tested them for surface hydration.[24,95] The surface hydration can be summarized as follows: H_2O adsorption occurs at the unsaturated surface aluminium atoms, followed by water splitting at the more acidic H_2O sites that distributes the protons to the most basic alumina oxygen. For comprehensive details, refer to chapter 3. The surface hydration is complete when the aluminium coordination sites are occupied. Typically half of the H_2O are dissociated. Wischert et al. studied low hydration states of the (1 1 0) facet at H_2O loads of at most 50 % H_2O molecules.[92] They discovered that at these conditions, the (1 1 0) model facet is subject to reconstruction by displacement of one aluminium atom to a neighboring tetrahedral site. It was recently shown that this aluminium atom is rather labile at aqueous conditions and might be subject to dissolution.[96] Since then, Digne's slab models were applied in several interface studies and good agreement with experimental results was attested.[53,65,91,97]

Prins criticized the commonly studied (1 1 0) facet introduced by Digne et al. because it is different from the spinel-like (1 1 0) surfaces, and does not produce considerable surface reconstruction due to its stoichiometric termination.[82] Pinto et al. observed this reconstruction on defect spinel models, however, surface hydration was neglected at the time .[98] Prins also noted that, since γ-Al_2O_3 is formed in a topotactic phase transition from boehmite, different surface terminations should be tested, as the cleaved model surface by Digne is not the only candidate. A recent study systematically explored these different surface terminations which should be expected by construction from boehmite dehydration.[99] While Digne's surface models are not the final answer, they remain useful to study possible surface chemistry at γ-Al_2O_3 facets.

1.2.4. Surface adsorption

The goal of this study is to obtain a set of candidate structures that can account for the adsorption modes obtained after impregnation in solution and drying. For that, one first needs to know the predominant surface facets available for adsorption. About 70 % of the facets are (1 1 0) and after that (1 0 0), as established in section 1.2.3.[78] For exploration, the drying state was considered, which circumvents difficult modelling of solvent effects. The initial assumption is that after drying, phosphates only bind to aluminium atoms via oxygen sites that are accessible at the surface. The fully hydrated surface models established by Digne et al. were used throughout this work.[24] In their publication, the hydration state of several surface terminations was tested at varying

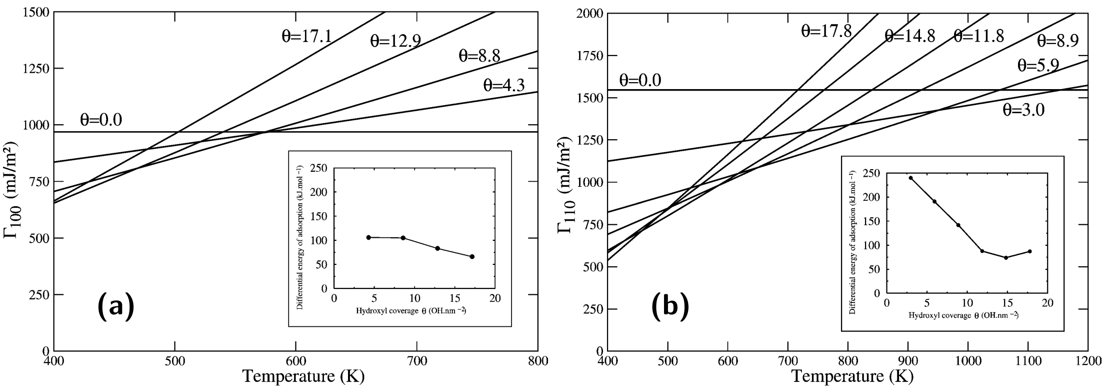


Figure 1.10.: Surface free enthalpy of γ-Al_2O_3(1 0 0) (a) and γ-Al_2O_3(1 1 0) (b) for different degrees of hydration with coverage values θ given OH nm^{-2}. The inset graphs show the differential free enthalpy of H_2O adsorption as a function of θ. This is the free enthalpy of each adsorbing H_2O molecule in the sequence of full hydration. Figures are from Digne et al. in ref.[95].

temperatures. According to ref. [95], the (1 0 0) and (1 1 0) facets are each hydrated at 120°C, which also shown in fig. 1.10. The more recently discovered surface reconstruction of (1 1 0) first described by Wischert et al.,[92] will be shown as crucial part of this study. The model adsorption of phosphates on the surface can be carried out with various strategies, but here the general idea was to substitute the surface hydroxyls in a reaction similar to

$$H_3PO_4 + {}^{\star}OH \longrightarrow {}^{\star}H_2PO_4 + H_2O$$

where $\star$ is the adsorption site of the hydroxy group. A similar computational adsorption study was done by Luschtinetz et al. involving phosphonates (H_2RPO_3) and several aluminium oxides and hydroxides.[52] The authors used the substitution approach with hydroxy groups on these surfaces. In particular, they explored a wide range of adsorption modes of two phosphonates on corundum, bayerite and boehmite surfaces by a systematic substitution of surface oxygen and hydroxide sites. The challenge there was similar to this adsorption study. However, adsorption on γ-Al_2O_3 is more complex because it has more diverse surface adsorption sites and complex hydrogen bond networks for each adsorption structure.

The work in this thesis also includes a detailed study of coverage effects and pyrophosphate adsorption, which to my best knowledge has not been attempted before by means of computational phosphate adsorption models. In chapter 3 the γ-Al_2O_3 surface structures and the exploration of candidate geometry will be discussed in detail. Moreover, chapter 4 presents NMR chemical shift simulations which validates the created adsorption models and gives unprecedented insight in the surface speciation of phosphates on γ-Al_2O_3.

2. Methods

The results produced for this thesis are exclusively computational models of γ-Al_2O_3 surfaces with the ambition to create a detailed understanding of phosphate adsorption on this material. Electronic structure calculations and construction principles of interface models are well established, however they rely on numerous approximations and assumptions that balance the accuracy of the results with computational performance. This chapter gives a short description of the applied methods and lists all important parameters used. A comprehensive discussion of the range of available methods as well as accuracy considerations are given in appendix A.1 (p. 143).

2.1. Computational setup

The presented models have have to be assessed from three distinct point of views: (i) electronic structure, (ii) structural aspects or geometry of atom positions, (iii) system sampling. While all electronic structure calculations were done on a similar level of density functional theory (DFT), there are several important differences in the technical details. The surface models were adjusted to answer different questions efficiently. Sampling has several meanings for the distinct challenges: for the database presented in chapter 3 and 4 it refers to a part of a global optimization to ensure that the identified geometry minimizes the energy; in contrast the sampling in chapter 5 and 6 is a central contribution in those results, where it contributes to the machine learning quality and free energy at the solid-liquid interface. The following sections summarize the employed approximations on these levels separately.

2.1.1. Electronic structure of periodic DFT

All calculations were conducted using at PBE-D3 level of theory,[89,100] which is a very common DFT exchange correlation potential based on the general gradient approximation that is combined with a force-field like term included to add long range dispersion interactions.

Small systems include single surface unit cells with unit cell parameters for the two γ-Al_2O_3 facets: (1 0 0) 8.358Å×5.547Å and (1 1 0) 8.358Å×8.035Å. Each system contains around 100 atoms. The larger systems are supercells constructed with two to three times larger unit cell vectors for about 16Å×16Å sized surface unit cells. Together with a 17Å thick layer of water molecules, they contain 800–1000 atoms. The small systems allow reaching higher accuracies for of electronic structure calculations than the latter. A high accuracy was important here, since the systems were also used to calculate NMR chemical shifts which are very sensitive the local structure. The larger systems are surface supercells build for molecular dynamics, where fast calculations are much more important.

For small systems with higher accuracy requirements, the electron structure was spanned in projector-augmented wave functions (PAW) with a kinetic energy cutoff at 500 eV and a density cutoff of 605 eV.[101] The PAW pseudopotentials are based on a norm conserving reference density, which allowed such a smooth density cutoff.[102] For efficient calculations, VASP (v. 5.3.5) was used.[103] Other parameters that can influence the electronic structure are smearing (set to 10 meV Gaussian-type smearing which is negligible at the end of a geometry optimization but stabilizes certain stating configurations) and two switches determining overall precision (PREC = accurate) and for optimized projector evaluation in real space (LREAL = automatic). The Brillouin zone was sampled by Γ-centered Monkhorst-Pack grids of 2×3×1 and 2×2×1 on the (1 0 0) and (1 1 0) model facets, reflecting the different surface cell vectors and the aperiodic surface normal vector. These parameters were thoroughly benchmarked and lead to well converged structures and energies.

The large systems were constructed as supercells with 16×16Å wide surface area, on which a 17Å thick layer of water was equilibrated using molecular dynamics. The Brillouin for these supercells is decreased such that calculations based only on the Γ-point have equivalent accuracy as the 2×3 and 2×2 grids in the small systems. The calculations were carried out with the QUICKSTEP module which is part of the CP2K software package.[104] QUICKSTEP uses a Gaussian *and* plane wave (GPW) method to represent Kohn-Sham orbitals and electron density. The difference to PAW calculations is explained in more details in appendix A.1.3 (p. 151). Norm-conserving GTH pseudopotentials were used together with molopt-DZVP basis sets.[104,105] The auxiliary density cutoff is set to 400 Ry (5442 eV). which is a bit higher than the software default of 280 Ry. The value has to be for higher accurate density transformation because the local basis used for orbitals in CP2K is significantly harder; see appendix A.1.3 (p. 151) for a detailed explanation.

2.1.2. DFT chemical shift calculations

NMR chemical shift calculations were carried out using the gauge-including projector-augmented plane wave method (GIPAW) by Yates, Pickard and Mauri.[66,106] The chemical shift tensor is defined by the ratio of induced magnetic field and external field:

$$\sigma(\mathbf{r}_\mathrm{N}) = \frac{\vec{B}_\mathrm{ind}(\mathbf{r}_\mathrm{N})}{\vec{B}_\mathrm{ext}} \tag{2.1}$$

After careful convergence tests, the Monkhorst-Pack k-point grids were increased to 3×5×1 and 3×3×1 for the surface models. The calculation involves numerical k-space derivatives, for which step sizes of DQ = 0.003 were used and symmetry operations were discarded where necessary (LNMR_SYM_RED=.TRUE). Furthermore, quadrupolar coupling constants (C_q) were estimated using quadrupole moments of 146.6 and -25.58 for ^{27}Al and ^{17}O. To ensure accurate electronic convergence SCF convergence was continued until consecutive values differed less than 10^{-8}eV (EDIFF).

2.2. Phase space sampling

Theoretical models for phosphoric acid adsorption on γ-Al_2O_3 model facets require as minimum a notion on how a phosphate group can adsorb to this interfaces. This notion can be translated into several adsorption candidates for which different sites are identified on the surface models. Depending on the number of candidates and our knowledge about the system, there are many different sampling methods for atomistic structures. Starting from manually testing individual adsorption mode with hand-made geometries is common, and typically followed by exploring all remaining adsorption modes if feasible, or at least several of them of the exploration space is too large. As alternative, one can test automated sampling methods. Popular sampling methods include ab-initio molecular dynamics (AIMD), Monte Carlo sampling or genetic algorithms,[107,108] which are more or less suitable for the actual challenge. AIMD is often used to sample gas phase or liquid phases, whereas Monte Carlo sampling and genetic algorithms can be used for global optimizations with little to no initial knowledge about the system. When including structural knowledge into the sampling, the situation is much more complicated, since the choice on which parts to sample (i. e. adsorption sites and reactions) and which to restrict (i. e. the surface model) have to be explicitly implemented. AIMD allows accurate sampling of water at the γ-Al_2O_3 model interfaces,[96] which also allows to test adsorption models, however at high computational cost.

In this work, the adsorption sites were sampled by manual structure generation based on a detailed analysis of surface patterns and a resulting combinatoric sample space. In-depth considerations are given in sections 3.2 and 3.3 (p. 58). For the solid-liquid interface, some of the most stable adsorption sites were simulated by means of AIMD. In this context each sampled structure as well their AIMD simulations represent one spot of the phase space. Even though AIMD in principle allows for dynamic change between adsorption sites, the time scales of these simulations are in a range of 10 ps to 50 ps which is just enough to sample liquid water but not to observe scarce transitions. In order to connect these spots, or in other words, to study the speciation mechanism at the impregnation conditions, biased AIMD calculations were carried out. The biasing approach applied here is metadynamics or metaD for short. This section also addresses the free energy and free enthalpy estimates, since in contrast to the electronic energy, which is only defined for a single geometry, it is directly related to the phase space by the degrees of freedom available to the system.

2.2.1. Geometry optimization

More than 1000 geometry optimizations were carried out within the structural exploration phase of this work. To focus on structure generation rather than correcting unsuccessful optimizations, the optimization algorithm had to be more reliable than any of the standard implementations in VASP. The fast inertial relaxation engine (FIRE)[109] was implemented in VASP by Henkelman et al. for their nudged-elastic-band (NEB)[110] based transition state search methods, but it can also be used separately. It is the only method that efficiently optimizes internal rotamers like free Al−OH and P−OH without user intervention, since it is only follows the force vector and adjusts its step size during the optimization. The quasi-Newton method in VASP (IBRION=1)[111] does not succeed for reasons explained in ref. [109]. The success of the optimizing algorithm also depends on the precision with which the force were determined. This depends the electronic optimization, which in this case done until 10^{-6} eV (EDIFF), while the forces were optimized until the largest force on all atoms was below $10\,\mathrm{meV\AA^{-1}}$.

2.2.2. Partial Hessian matrix and harmonic free energies

Frequency calculations were carried out on the best candidate geometry for each adsorption site. Since the data was acquired to compute vibrational component of the free energy computations, the parameters were optimized as well. For the description of the relevant vibrational surface modes, *partial* Hessian matrices were calculated by finite differences. The electronic optimization was done until 10^{-8} eV, displacements of 0.015Å were used for all atoms. To limit the number of atom displacements to the relevant area only surface atoms and adsorbates were included. Without further processing, this partial Hessian matrices contain frustrated translations due to the neglected subsurface atoms. These are difficult to identify by hand, but they can be removed by projection, where they were treated as free translations. The remaining partial Hessian matrix were diagonalized and the resulting vibrational modes were checked for imaginary modes. If imaginary modes occurred, the structures were reoptimized with tighter force criteria until the imaginary mode vanished.

The vibrational frequencies were furthermore used to calculate thermodynamic correction terms for the Gibbs free enthalpy G. The overall free enthalpy is defined as

$$G = E_{\mathrm{el}} + U + pV - TS \tag{2.2}$$

where contributions to the inner energy U and entropy S of free molecules are separated in translational, rotational and vibrational components:

$$U = U_{\mathrm{trans}} + U_{\mathrm{rot}} + U_{\mathrm{vib}} \tag{2.3}$$

$$S = S_{\mathrm{trans}} + S_{\mathrm{rot}} + S_{\mathrm{vib}} \tag{2.4}$$

Each component can be determined by constructing their partition function.[112] They whole expression for the free enthalpy is thus:

$$G = E_{\text{el}} + U_{\text{trans}} + U_{\text{rot}} + U_{\text{vib}} + pV - T\big(S_{\text{trans}} + S_{\text{rot}} + S_{\text{vib}}\big) \tag{2.5}$$

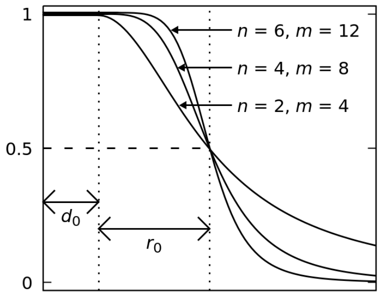
It is necessary to calculate all terms for gas-phase reference molecules like H_2O and H_3PO_4. For the surfaces, translational and rotational components vanish, as well as pV since surfaces do not have a partial pressure, leaving only vibrational contributions:

$$G_{\text{surf}} = E_{\text{el}} + U_{\text{vib}} - TS_{\text{vib}} \tag{2.6}$$

The lowest vibrational frequencies have the most significant contribution to the overall thermodynamic correction terms, which is the reason why it was important to remove the frustrated translations discussed above. In connection with the phase space, the calculation and diagonalization of a Hessian matrix for harmonic vibrational mode is nothing else than a simple analytical alternative to sampling.

2.2.3. Molecular dynamics

The explicit solvation of the surface and surface species requires to include dynamic, statistical effects of the solvent. For this, AIMD simulations were carried out at DFT level as described in section 2.1.1. Room temperature calculations where simulated with a slightly elevated simulation temperature of 330 K using the CSVR thermostat (100 fs time constant) in order to compensate for systematic errors known about PBE-D3.[113] In order to increase simulation throughput, the hydrogen atomic weight was increased to $3\,\text{g}\,\text{mol}^{-1}$ which allowed to carry out the simulations with a step size of 1 fs. These parameters were in all AIMD. Unbiased AIMD calculations were carried out for at least 10 ps to equilibrate the system before starting metadynamics simulations.



$$s(r) = \frac{1-\left(\frac{r-d_0}{r_0}\right)^n}{1-\left(\frac{r-d_0}{r_0}\right)^m} \quad \text{for} \quad r > d_0$$

Figure 2.1.: Rational switching function to define a coordination collective variable. Four parameters, d_0, r_0, n and m define its behavior.

2.2.4. Bias AIMD along collective variables

Biased AIMD simulations are a critical tool to study rare events. AIMD simulations are carried out for time scales in the range of picoseconds, where the probability of most chemical reactions is very low. Biases are used to nudge the system over reaction barriers. There are many possible methods to do so, but regardless of the actual method, one first needs to define a direction within phase space in which a bias should be applied. It is important to note that really *any* differentiable system quantity can be used to add biases to the system. When defined, a bias exerts an additional force at the atoms in the system. The simplest variable to bias for chemical bond formation and cleavage is the distance between the atoms of interest. For large systems, the difficult question is, which distance to consider. For example in case of P−O bond breakage in PO_4^{3-}, one can define any of the four bonds, whereas for bond formation with H_2O, oxygen atoms in several surrounding H_2O molecules could be chosen. This complexity cannot be added by means of a distance, which is only defined for a single pair of atoms, but coordination is a much more suitable concept. In general, the coordination c is a collective variable

(CV) which counts the number of bonds within the

$$\text{same group:}\quad c=\sum_{i=1}^{N}\sum_{j>i}^{N}s(r_{ij})\qquad\text{or different groups:}\quad c=\sum_{i=1}^{N}\sum_{j=1}^{M}s(r_{ij})\tag{2.7}$$

A bond can be defined as an atom pair with a distance r_{ij} below a user-defined cutoff distance. For a continuously differentiable definition, a smooth transition between bonded and non-bonded pairs, a switching function has to be established. A common switching function for the coordination CV is:

$$s(r)=\begin{cases}\frac{1-\left(\frac{r-d_0}{r_0}\right)^n}{1-\left(\frac{r-d_0}{r_0}\right)^m}\,, & \text{if}\quad r>d_0\\ 1\,, & \text{otherwise}\end{cases}\tag{2.8}$$

The function is plotted in fig. 2.1. The cutoff distance in this function is given d_0+r_0, where r_0 stretches the overall function and d_0 is a radial shift. The exponents n and m change the stiffness and particularly influence the decay at long distances. For convenience, they are typically chosen such that $m=2n$. Only this ensures that the sum d_0+r_0 can be used as a cutoff value: $s(d_0{+}r_0)=\frac{n}{m}=0.5$. The definition of parameters is an involved process which required months of testing, before identifying a set of parameters that performs well. The most flexible set of parameters identified only depended on r_0; the other parameters were fixed at $d_0=0, n=6, m=12$. For the coordination of P−O, r_0 is set to 2.2Å, for Al−O to 2.4Å, and for O−H to 1.4Å. This parameters are quite specific to the targeted element pairs to be split or formed but with local effects, like the coordination of octahedral and tetrahedral aluminium sites, or differences between μ_1-OH and μ_3-O, these parameters are not necessarily ideal in all situations.

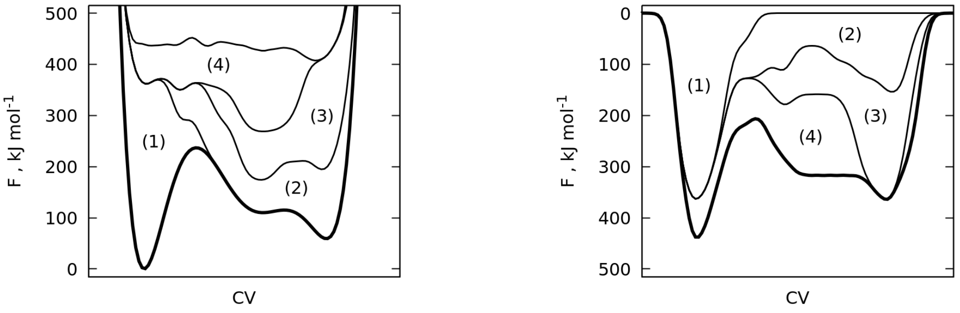
These functions are implemented in the PLUMED code,[114–116] which is compatible with CP2K. The functionality is available with the COORDINATION keyword. It should also be noted, that it is used as a general method to count the number of bonds within groups of atoms, which different from coordination numbers. Those describe the number of other atoms to the same atomic center, however the coordination number of a single atom can be described for one center, if this one is the only element in the group. For brevity, this section just covered the most basic CVs used the most within chapter 5. switching coordination function parametrized in this work to mathematically describe coordinations as a sum. It should be noted that *any* differentiable function can be used as a switching function switching function can be used and more complex quantities than the sum in eq. 2.7 are needed in some cases. Those are described in appendix A.4 (p. 162).

2.2.5. Metadynamics

There are several methods to add a bias to AIMD along a set of CV's. In this work, metadynamics (metaD)[117] and, more specifically, well-tempered metadynamics (WT-metaD)[118] were carried out to study phosphate adsorption at the solid-liquid interface. The idea in these methods is to deposit a potential $V(s,t)$ in form of gaussian functions (in jargon hills) along positions of a CV s in order to force the system to leave this particular area in the free energy surface (FES) $F(s)$:

$$V(s,t) = \sum_{\tau=0}^{t} W(\tau) \exp\left(-\frac{(s-s_\tau)^2}{2\sigma^2}\right) \tag{2.9}$$

The potential is given as a sum over τ gaussian during the simulation time t, where the σ parameter sets the gaussian width. The gaussian height in standard metadynamics is a constant. It can be shown that the potential V(s,t) converges with the FES shifted by a constant.[119] The process is illustrated in fig. 2.2. For WTmetaD, the gaussian height is decreased depending on the existing bias potential $V(s,t)$ at the point of the next



(a) filling a known FES with gaussians **(b)** recreating the same FES from gaussians

Figure 2.2.: Illustration of how a collective variable (CV) is biased by metadynamics. The free energy surface (FES) is based on an arbitrarily chosen polynomial ($125 - 45x + 35x^2 + 3x^3 - 7x^4 + 0.1x^5 + 0.3x^6$), whereas the biasing potential is a sum of gaussians fixed on 25 grid points along the graph.

gaussian deposition at time t:

$$W(t) = W_0 \exp\left(-\frac{V(s,t)}{\Delta T}\right) \tag{2.10}$$

where ΔT is another input parameter. The bias factor allows for an overall better convergence behavior, since the growth rate of $V(s,t)$ is limited, but even at full convergence, $V(s,t)$ does not fully compensate the FES:

$$V(s, t \to \infty) = -\frac{\Delta T}{\Delta T + T} F(s) + C \tag{2.11}$$

Instead, $V(s,t)$ would have to be scaled by a factor $\gamma = -(T + \Delta T)/T$. This factor γ is referred to as bias factor and most commonly reported in publications instead of values for ΔT. Overall, one needs to define a gaussian height W_0 and width σ, a deposition pace $1/t$ and a bias factor γ for a one-dimensional WTmetaD simulation. For further dimensions, i. e. to bias multi-dimensional FES projections along further CV's, one needs to define additional width parameters along the CV directions. In this work, for all presented WTmetaD simulations, a deposition pace of 100 ps^{-1} was used, which is every 10 simulation steps. The bias factor was set to $\gamma = \frac{\Delta E}{k_B T} = 72.9$ assuming $200\,\mathrm{kJ\,mol^{-1}}$ is the largest barrier needed to be overcome. This bias factor is smaller than the one used in previous work,[96] and therefore expected to allow slightly faster convergence. It also is the default method to determine the best bias factor in a new derivative method of WTmetaD.[120]

AIMD simulations are generally very time-consuming calculations, where WTmetaD is no exception. The discussion above considers the modification of the AIMD trajectory by constructing additional energy terms which generate forces in forthcoming simulation steps. The overall bias potential can only build up in areas of this single trajectory. The computation time that the trajectory, in this context referred to as *walker*, requires to cover the overall area defined by the CV's is the minimum calculation time. By adding more walkers that simultaneously contribute to the bias potential, one can parallelize the whole process. In fact, the original publication of the metadynamics (metaD) method describes this parallelization technique.[117] The most obvious advantage of this parallelization is reduced computation time at the cost of higher parallel workload. A more subtle advantage is that the parallel walkers, while ultimately sampling different areas of the FES, they can also limit initial oscillation effects from an alternating coverage of different areas in the FES, if the starting points are distributed accordingly. For example 100 walkers were used in ref. [120]. In case of AIMD, the limiting factor is the overall available resources, hence for this work, the number of walkers was always limited to 4.

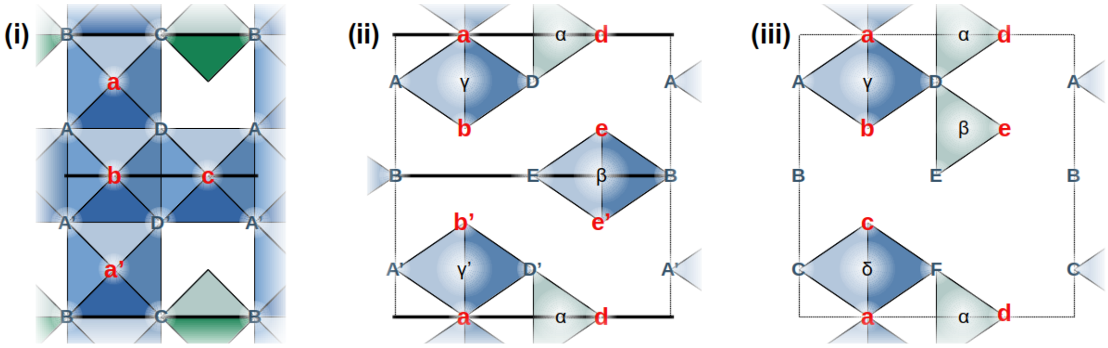
3. Structural exploration

The aim of the work presented in this chapter is a detailed exploration of all relevant adsorption patterns of H_3PO_4 on hydrated γ-Al_2O_3. This exploration also sets the foundation for all derived work, that will be discussed in chapters 4, 5 and 6. Compared to other adsorption studies,[17,52] in which all surface oxygen sites were μ_2-OH sites, the hydrogen networks of γ-Al_2O_3 surface models are more complex. They involve partially protonated μ_3-O sites and μ_1-OH as well as μ_1-H_2O sites with rotational degrees of freedom. This requires an extended search for the protonation state of every generated system, from reference to adsorption candidate. Overall, three reference structures were used as a starting point for the adsorption states of phosphates on γ-Al_2O_3: two Digne's original surface models,[95] here referred to as (1 0 0) and n(1 1 0), and Wischert's reconstruction, R(1 1 0).[92] These reference structures were reoptimized at the applied level of theory, including an initial, thorough exploration of the proton network. The most stable configurations are illustrated in fig. 3.1b. The sites relevant for chemisorption in these structures are oxygen atoms highlighted in red for sites that originate from surface hydration and blue for the topmost layer of alumina oxygen atoms. This distinction is arbitrary but a quite useful first hierarchical order in which phosphates may adsorb by substitution at these sites: first substitution at hydration sites, then eventually sites of the alumina network. The structures in fig. 3.1b reflect the chemical bonding situation accurately and even vaguely map the hydrogen bond network of each surface, however, but they are too complex to systematically address the exploration of different adsorption patterns. Fig. 3.1a is an abstract representation of the top view of the three surfaces that includes the coordination polyhedrons of the aluminium atoms and labels for the highlighted oxygen atoms; lowercase labels for hydration sites, uppercase labels for alumina network sites. Moreover, this surface representation includes symmetry aspects of Digne's (1 0 0) and n(1 1 0), see fig. 3.1i and ii. Except for the hydrogen network, these symmetries overall reduce th number of possible hydration states, which considerably reduces the number of adsorption modes. Thus also the labels were adjusted. Symmetry-equivalent sites were provided with the same letters and one marked with a prime. For example on fig. 3.1ii the site b and b' These labels are important throughout the whole thesis, because they are used to refer to individual adsorption sites where necessary.

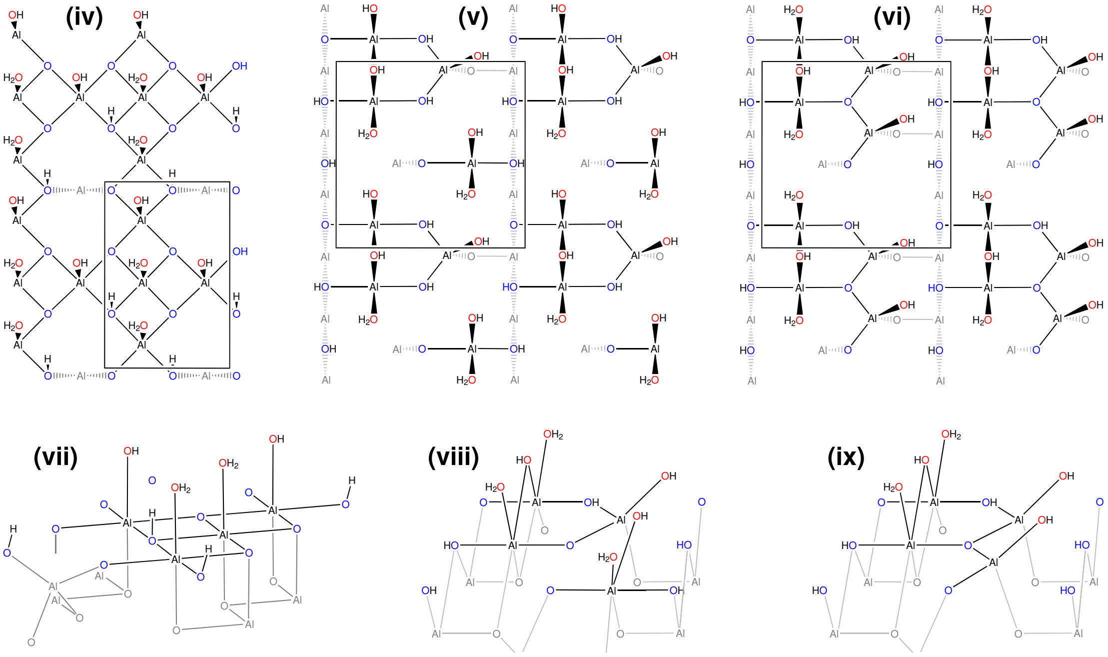
With the involved challenges in finding the relevant adsorption modes, the next three sections explain in detail, how all adsorption structures were sampled. First, some general considerations are laid out, including an explanation on how structural exploration is related to phosphate speciation at drying conditions and how the adsorption energy can be interpreted at the solid-liquid interface. This is followed by a detailed description and a notation for all the adsorption sites. After that, the most stable adsorption sites are analyzed and construction principles for them are identified.

3.1. Drying conditions and structural exploration

For the purpose of covering all the targeted adsorption patterns, one needs to choose the relevant environmental conditions. This choice is also strictly necessary in order to quantitatively compare different adsorption candidates by their stability. The preparation steps of a heterogeneous catalyst starting from the calcined γ-Al_2O_3 support are wet-incipient impregnation, drying and activation. Of those, we are mostly interested in impregnation (solid-liquid interface, 25°C, 1 atm) and drying conditions (solid-gas interface, 120°C and 2 kPa for H_2O – air moisture levels of 60 % at ambient conditions).



(a) Abstract structural representation involving local symmetry patterns and global symmetry features. Al atoms are in the center of their coordination octahedra (blue) and tetrahedra (green) and labelled with Greek letters when necessary. Oxygen sites from adsorbed water are labelled with red lowercase while alumina oxygen sites are labelled with blue upper case letters. Mirror planes are shown as thick black lines.



(b) Top view (iv, v, vi) and side view (vii, viii, ix) structures including the protonation states of all individual sites. Subsurface atoms are given in grey to complete the local coordination of the surface oxygen atoms.

Figure 3.1.: Three model reference surfaces of γ-Al_2O_3 (1 0 0)[24] (leftmost structures), n(1 1 0)[24] (centered structures) and R(1 1 0)[92] (rightmost structures). Oxygen atoms are highlighted in all structures to distinguish the sites of preadsorbed water (red) and alumina oxygen (blue). The structures in (a) provide the necessary level of abstraction from (b), and in particular the labelling that is critical to follow the discussion in this work in detail.

Starting from the impregnation conditions is impractical for structural exploration, since it would require to simulate liquid water at the surface. Sampling of liquids using molecular dynamics and comparing stabilities is computationally demanding and complicated. In contrast, at drying conditions the surface model only involves adsorbed water with comparably well-defined positions. The interaction with the gas phase over the surface can be neglected, and the translational and rotational degrees of freedom in the gas phase itself can be described with analytical terms within ideal gas and rigid rotor approximations (cf. eq 2.5, p. 45 and ref. [112]). Thus, drying conditions can be simulated by creating static models of surface adsorption, and comparing their energies. Only H_2O (g) has to be considered as a gas-phase reference state, as will be shown below. The adsorption enthalpy of H_2O was well described by Digne et al. using only (nowadays) inexpensive static calculations.[24] Since these are isobaric conditions, i. e. the partial pressure of H_2O vapor is considered constant, the relevant energetic quantity is the free enthalpy.

After impregnation with H_3PO_4, the stability of the adsorbed phosphate species at drying conditions depends on a competition for sites between H_2O and H_3PO_4 adsorbates. It seems trivial to describe this by means of adsorption enthalpies, but it brings out a subtle problem: while during impregnation, the adsorption of phosphate species can be described in chemical balance with phosphates in solution, no such reference exists in the dried state, since the solution had evaporated and phosphates or phosphoric acid are not volatile at drying conditions. Hence, for a thermodynamic description of phosphates at the dried interface, adsorption free enthalpies raise questions because no consistent phosphate reservoir such as H_3PO_4(aq) exists. Consequently, it is not possible to define an unambiguous adsorption free enthalpy, the only relevant quantities being free enthalpy differences between model adsorption candidates. As a pragmatic choice (i. e. simple reproducibility, knowability from the beginning of the study), the theoretical adsorption free enthalpy is estimated with respect to H_3PO_4(g) using the same parameters as for H_2O

(120°C and 2 kPa). An equally valid choice would be any related phosphate-containing reference, for example the most stable adsorption state.

Since the surfaces are fully hydrated, there are two cases of H_3PO_4 adsorption: physisorption or chemisorption after substitution of chemisorbed water, $^\star H_2O$. Moreover, a single H_3PO_4 can substitute more than one $^\star H_2O$ molecule, thereby forming higher dentation modes which will be explained in detail in section 3.3. The adsorption reaction for any such reaction of a single H_3PO_4 can be written in the following simplified way:

$$\mathrm{H_3PO_4} + x\,{}^\star\mathrm{H_2O} \longrightarrow {}^\star\mathrm{H_3PO_4} + x\,\mathrm{H_2O}\uparrow \tag{3.1}$$

where the "$\star$" indicates surface adsorbed species and the only variable x indicates the amount of substituted water. Physisorption is described with $x = 0$. Even without inspecting the reference surfaces in fig. 3.1b in full detail, it should be clear that most surface species are actually hydroxyl moieties; the protonation state of the adsorbed phosphate moreover defies chemical intuition. A more general reaction involving these aspects is:

$$\mathrm{H_3PO_4} + (x{-}y)\,{}^\star\mathrm{H_2O} + y\,{}^\star\mathrm{OH} \longrightarrow {}^\star\mathrm{H}_{3-z}\mathrm{PO_4} + (z{-}y)\,{}^\star\mathrm{H} + x\,\mathrm{H_2O}\uparrow \tag{3.2}$$

In this equation, x still defines the number of substituted H_2O, but the two additional variables y account for substitution of hydroxyl moieties and z for the protonation state of the adsorbed phosphate. The shifted proton is given as $^\star H$ which actually either forms $^\star OH$ or $^\star H_2O$. Only eq. 3.2 correctly describes the adsorption of orthophosphate species; however, it should be noted that x is the only variable which changes the material balance of the surface, and therefore eq. 3.1 is sufficient to determine the reaction free enthalpy. Variables y and z only describe proton transitions on the surface. The adsorption free enthalpy is calculated from the absolute free enthalpy (cf. eq. 2.6, p. 2.6) of the tested candidates $G[^\star H_3PO_4]$ with the reference surface $G[\star]$ (which includes preadsorbed $\star H_2O$, cf. fig. 3.1b), and the gas-phase references $\mathrm{G[H_3PO_4]}$ and $\mathrm{G[H_2O]}$:

$$\Delta \mathrm{G_{ads}} = \mathrm{G[{}^{\star}H_3PO_4]} + x\,\mathrm{G[H_2O\,(g)]} - \mathrm{G[\star]} - \mathrm{G[H_3PO_4(g)]} \tag{3.3}$$

The phosphate reservoir $\mathrm{G[H_3PO_4(g)]}$ is always constant but the surface reference, $\mathrm{G[\star]}$ depends on the facet. This is important when comparing two adsorption free enthalpies from different model facets, because systematic errors in both reference calculations can be different, limiting the accuracy of this comparison. Consequently, free enthalpy comparisons between the (1 0 0) and (1 1 0) facet have to be discussed with more caution. It also adds an important aspect to the comparison of the unreconstructed and reconstructed structural references of (1 1 0), since for this facet it is desirable to use the same $\mathrm{G[\star]}$ in order to avoid this problem altogether. Wischert's reconstruction was identified in a study about low water coverage at high temperatures, which lead to the transition of the β-aluminium atom (see fig. 3.1 ii and iii) from an octahedral to a tetrahedral site. This structural change in fact reduces the number of available adsorption sites by one, hence the number of preadsorbed H_2O on R(1 1 0) is reduced compared to n(1 1 0) as well. The reaction from one (1 1 0) structural reference [n] to another [R] is given by:

$$[\mathrm{n}] \rightleftharpoons [\mathrm{R}] + \mathrm{H_2O}\,, \qquad \Delta G = +29\,\mathrm{kJ\,mol^{-1}} \tag{3.4}$$

The reconstructed structural reference R(1 1 0) is with a free enthalpy of this reaction of 29 $\mathrm{kJ\,mol^{-1}}$ less stable than the unreconstructed n(1 1 0). To my best knowledge this parameter is reported by us for the first time, since Wischert discovered the reconstruction in a study of high-temperature models with incomplete hydration,[92,121,122] and other studies involving Wischert's R(1 1 0) did not compare it with Digne's n(1 1 0).[96] By combining the reaction eq. 3.1 and eq. 3.4, one can calculate free enthalpies for adsorption on both structural references with respect to the same surface reference:

$$\Delta G_{\mathrm{ads}}^{[\mathrm{R}]} = G[{}^{\star}H_3PO_4] + (x+1)\,G[H_2O\,(g)] - G[n] - G[H_3PO_4(g)] \tag{3.5}$$

where G[n] is the unreconstructed reference. The reconstruction reaction in eq. 3.4 adds $1\,H_2O$ to the equation. Overall, these considerations allow for an extensive exploration of the most exposed (1 1 0) facet even including the possibility of reconstruction while avoiding limits in the comparability.

For adsorption of more than one phosphate, as for orthophosphate coadsorption or pyrophosphate adsorption, eq. 3.2 needs to be applied for both phosphate atoms, however, this way we introduce a second phosphate from the arbitrarily chosen reference of H_3PO_4 in gas phase. For the sake of a good global comparison, average adsorption enthalpies will be given, as the effect of the phosphate reference remains constant this way. For the adsorption of two H_3PO_4, leading to a substitution of x and x' H_2O molecules, the *average* adsorption free enthalpy is defined as:

$$\Delta G_{\mathrm{ads}}^{2\,\mathrm{P}} = \frac{1}{2}\Big\{G[2\,{}^{\star}H_3PO_4] + (x+x')\,G[H_2O\,(g)] - G[\star]\Big\} - G[H_3PO_4(g)] \tag{3.6}$$

where 2 P indicates coadsorption. For pyrophosphate, the orthophosphate condensation reaction needs to be included:

$$2\,H_3PO_4 \longrightarrow H_4P_2O_7 + H_2O$$

Similar to the surface reconstruction in eq. 3.4, one additional H_2O molecule is released. The pyrophosphate adsorption enthalpy values are, like the other enthalpy values, reported with respect to one mole H_3PO_4:

$$\Delta G_{\mathrm{ads}}^{\mathrm{P}_2} = \frac{1}{2}\Big\{G\,[{}^{\star}H_4P_2O_7] + (x+x'+1)\,G[H_2O\,(g)] - G[\star]\Big\} - G[H_3PO_4(g)] \tag{3.7}$$

where P_2 is a short form for pyrophosphates. Finally, for coadsorption and pyrophosphate adsorption on R(1 1 0), the free number of H_2O molecules is again adjusted:

$$\Delta G^{2\,P}_{ads} = \frac{1}{2}\Big\{G[2\,{}^{\star}H_3PO_4] + (x{+}x'{+}1)\,G[H_2O\,(g)] - G[\star]\Big\} - G[H_3PO_4(g)] \tag{3.8}$$

$$\Delta G^{P_2}_{ads} = \frac{1}{2}\Big\{G\,[{}^{\star}H_4P_2O_7] + (x{+}x'{+}2)\,G[H_2O\,(g)] - G[\star]\Big\} - G[H_3PO_4(g)] \tag{3.9}$$

Eq. 3.3 and 3.5–3.9 describe the material balance based on x and x'. Structural considerations are described by y and z in eq. 3.2, i. e. the substituted surface groups and proton transfer from orthophosphoric acid to the surface. The latter is a general problem for any adsorption structure that starts from a reference structure with an involved hydrogen bond network. There is in fact no reason to assume that the adsorbate can be constructed without significantly impacting the hydrogen bonds. In principle, it has to be recreated for every adsorption. The next section will discuss how this can be done based on a study of the reference states in fig. 3.1b.

3.2. Hydrogen bond network

In order to determine accurate phosphate adsorption free enthalpies, the correct hydrogen bond network needs to be identified. Some of the adsorbing H_2O molecules are dissociated and produce hydroxide ions that form hydrogen bond networks which cover the γ-Al_2O_3 facets. The quest for the best hydrogen bond network is a global optimization problem and thus difficult to achieve with the "traditional" static DFT workflow, in which geometries are optimized by following the force gradient towards the next local minimum. Global optimizations are typically done either by combinatorics, i. e. exploring all possible protonation states and hydrogen bonds by autogenerated structures, or by random sampling (e. g. Monte Carlo simulations). These methods are computationally demanding and require in-depth knowledge about the model system in order to establish the configuration space (in this case, of the considered protonations and hydrogen bond networks).

As the first step, the original results from Digne and Wischert were adapted and re-optimized at the considered level of theory (PBE-D3). For all the reference surfaces, several starting configurations with different protonations were tested. The results are overall consistent with the original structures by Digne, but slightly different hydrogen bond networks for (1 1 0) were found to be most stable. The hydrogen bond networks drawn in fig.3.1b are, for each structural reference, the most stable cases at this level of theory.

In an attempt to rank the mobility of protons at the distinct sites, one should focus on protonation states for symmetry-equivalent sites like, for example, b/b', e/e', A/A' and D/D' in fig. 3.1a. If these sites have different protonation states, they must be more prone to protonation or deprotonation. In case of H_2O substitution by an adsorbate that increases or reduces the number of protons participating in the hydrogen bond network, these sites can accept an additional proton or donate a proton to more basic sites. For example, the site pairs b/b' and e/e' are μ_1-OH/μ_1-H_2O while A/A' is μ_3-O/μ_3-OH. In contrast, D/D' are each μ_2-OH. Hydrogen bonds can stabilize the deprotonated states; i. e. μ_1-OH or μ_3-O are either stabilized by hydrogen bonds or being protonated – or neither of them, leading to a slightly less stable structure. For a complete picture they need to be included. By listing the different oxygen μ_x bonding modes on n(1 1 0), we see that:

- $\frac{1}{2}$ of μ_1 sites are OH or H_2O;
- $\frac{2}{3}$ of μ_2 sites are OH and $\frac{1}{3}$ are O (stabilized by 2 hydrogen bonds);
- $\frac{2}{3}$ of μ_3 sites are OH and $\frac{1}{3}$ are O (stabilized by 1 hydrogen bond).

The similarity between μ_2 and μ_3 is surprising, since μ_2 should be more basic than μ_3. It seems that, while the local environment of the site matters, it is also important that the protons are regularly dispersed. With H_3PO_4 adsorption the number of protons in the system changes depending on the number of substituted H_2O. The capability of the sites to yield or remove protons depends on their proton affinities. Most information for this lies in the reference surfaces, but after the first adsorption tests of phosphate it was found that this adsorbate typically dissociates and transfers one or two protons to the surface (more for higher dentation). All these considerations lead to proton affinities that can be ranked in the following way:

$$\mu_1\text{-O} \gg \mu_2\text{-O} > \underbrace{\text{P}{-}\text{O} \approx \mu_3\text{-O} \approx \mu_1^{\text{Al}_O}\text{-OH}}_{\text{mobile}} > \mu_1^{\text{Al}_T}\text{-OH} \gg \mu_2\text{-OH}, \mu_3\text{-OH}, \mu_4\text{-O} \quad (3.10)$$

Here, $\mu_1^{\text{Al}_O}$ and $\mu_1^{\text{Al}_T}$ are oxygen atoms at octahedral and tetrahedral aluminium atoms. For most adsorption candidates, this allowed to limit the range of protonation states to only a 3-4 test calculations which were calculated by DFT. Still, in some cases the number of combinations was more than 30, thus the number of calculations was limited to about 15 cases for which the protons were well dispersed over the surface and layers. On average, six hydrogen bond networks were tested per considered site, which lead to around 1000 geometry optimizations for the whole phosphate adsorption database. The following section describes which sites were actually considered.

3.3. Surfaces and adsorption modes

To understand all tested phosphate surface species in detail, one has to combine all possible adsorption modes with the given surface models. The adsorption modes of chemisorbed molecules or ions can be understood by the number of bonds they form with the surface and by the actual surface sites they bind to. A simple adsorption

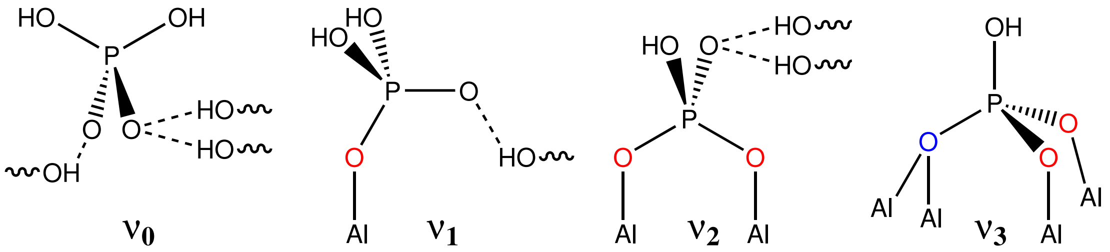


Figure 3.2.: Adsorption modes of orthophosphates on γ-Al_2O_3. Physisorption (ν_0) and several dentation patterns are possible: monodentates (ν_1), bidentates (ν_2) and tridentates (ν_3). All phosphate oxygen atoms are coordinated with at least one more atom other than phosphorus. The dentation sites are highlighted in red for substituted water adsorption sites, and blue for alumina oxygen sites.

species that only binds to the surface via a single atom is commonly described in a notation like μ_x, where x refers to the number of surface atoms it binds to. For more complex adsorption cases, like for phosphates binding via several oxygen atoms to the surface, there is no generally accepted notation available, but the bond patterns are described as dentations.[52,72] In this work, a short notation is introduced based on the Greek letter ν. For one orthophosphate (PO_4) unit, x in ν_x refers to the number of phosphate oxygen atoms binding to the surface. Notably, the variable x in the reaction eq. 3.2 and x in ν_x are equivalent, because by the definition of the reference surfaces, the phosphate adsorption always involves substitution of OH or H_2O, releasing an equivalent amount of water. Fig. 3.2 shows all the considered adsorption modes from physisorption (ν_0) to tridentates (ν_3). Except for physisorption, all adsorption modes are constructed by substitution of preadsorbed water sites (cf. red colored surface sites in fig. 3.1), which are mostly μ_1-OH and μ_1-H_2O. For the higher dentation modes (ν_2 and ν_3), the actual adsorption requires to identify matching adsorption site patterns. In case of ν_2, a minimum requirement is two neighboring adsorption sites with a matching distance, whereas ν_3 needs an arrangement of three adsorption sites close to a regular triangle. Based on the findings of Luschtinetz et al.,[52] the distance criterion was held on relatively

Table 3.1.: 51 adsorption site candidates on the three surface models. The site labels in superscript are given in fig. 3.1i, ii and iii. The full structure exploration included all possible adsorption combinations based on neighboring sites. The struck through structures are excluded later on, since two phosphate oxygen atoms bind to the same aluminium atom.

Facet	Monodentate	Bidentate	Tridentate
(1 0 0)	ν_1^a, ν_1^b, ν_1^c	ν_2^{ab}, $\nu_{2*}^{aa'}$, ν_2^{bc}, ν_{2*}^{cb}	ν_3^{abc}, ν_{3*}^{acb}, ~~$\nu_{3*}^{aa'B}$~~, ~~$\nu_{3*}^{aa'C}$~~, ~~ν_3^{abA}~~
n(1 1 0)	ν_1^a, ν_1^b, ν_1^d, ν_1^e	~~ν_2^{ab}~~, $\nu_2^{bb'}$, ν_2^{de}, ~~$\nu_2^{ee'}$~~	~~ν_3^{abA}~~, ~~ν_3^{abD}~~, $\nu_3^{bb'B}$, $\nu_3^{bb'E}$, ν_3^{deA}, ~~ν_3^{deD}~~, ~~$\nu_3^{ee'B}$~~, ~~$\nu_3^{ee'E}$~~
R(1 1 0)	$\nu_1^a, \nu_1^b, \nu_1^c, \nu_1^d, \nu_1^e$	~~ν_2^{ab}~~, ~~ν_2^{ac}~~, ν_2^{bc}, ν_2^{de}, ν_2^{dC}, ~~ν_2^{dF}~~, ν_2^{eB}, ~~ν_2^{eE}~~	~~ν_3^{abA}~~, ~~ν_3^{abD}~~, ~~ν_3^{acC}~~, ~~ν_3^{acF}~~, ν_3^{bcB}, ν_3^{bcE}, ~~ν_3^{deD}~~, ν_3^{deA}, ~~ν_3^{cBC}~~, ~~ν_3^{cEF}~~

*sites are not in direct view of fig. 3.1i due to periodic boundary conditions

strictly, ignoring "distant" neighbors with distances larger than 3Å. Since these patterns, in particular the triangular patterns are not easily attained, alumina oxygen sites (cf. blue colored surface sites in fig. 3.1) were included in the construction of several higher dentation modes. A known reaction mechanism to form such adsorption species is the phosphorylation, as discussed in section 1.1.3 (p. 18).

Now, the systematic exploration of all possible adsorption is a major challenge, because despite four distinct adsorption modes, the three structural references provide a considerable number of adsorption sites. In order to constrain the number of explored structures to the most relevant cases, selection criteria were identified and applied for next explored structures. Overall the three model surfaces are comprised of 15 water adsorption sites and 18 alumina oxygen sites. In order to keep track of every tested adsorption site, the ν_x notation is extended by the site with the labels given in fig. 3.1a. For example, on (1 0 0) the monodentate, bidentate and tridentate adsorption modes are labelled ν_1^a, ν_2^{ab} and ν_3^{abc}, respectively. Due to symmetry, the overall number of adsorption sites can be slightly reduced, which if reflected in the labels, e. g. b and b' on n(1 1 0) are equivalent.

What followed based on these considerations was the systematic listing of all possible sites. Even after excluding symmetry-equivalent cases, this left at least 51 chemisorbed candidate sites which are tabulated in tab. 3.1. Physisorption (ν_0) was handled separately by positional screening of the surfaces, leading up to 16 test structures (including hydrogen bond optimization, see section 3.2) for each physisorption result. Without further knowledge, it is difficult to reduce this number. It should be noted that for each site, sampling the hydrogen bond network are necessary, leading to hundreds of calculations, as discussed in section 3.2. Moreover, for the purpose of processing the results and extending this data base with relevant candidates, further selection criteria are needed. A full list of free enthalpies is given in tab. A.1 (appendix, p. 154). After analyzing the free enthalpies of all candidates on $(1\,0\,0)$ and $\mathrm{n}(1\,1\,0)$, the following criteria were derived:

1. Two phosphate oxygen atoms never bind to the same aluminium atom.

2. Substitution of μ_1-O is generally preferred over μ_2-O and μ_3-O.

3. Free phosphate oxygen atoms are either protonated or stabilized by hydrogen bonds.

The first criterion is consistent with one of the findings in ref. [52]. It simplified the structural exploration on $\mathrm{R}(1\,1\,0)$, where several of the candidates in tab. 3.1 were discarded. However, it should be noted that some of the unfavorable adsorption structures lead to a critical further insight. For example, $\mathrm{n}(1\,1\,0)\nu_3^{ee'B}$ leads to the reconstruction of $\mathrm{R}(1\,1\,0)\nu_3^{eB}$, which is one of the most stable adsorption modes discovered; it in fact showed that it was necessary to include models based on Wischert's reconstruction. $(1\,0\,0)\nu_3^{abA}$, while clearly violating the first criterion, is only $10\,\mathrm{kJ\,mol^{-1}}$ less stable than $(1\,0\,0)\nu_3^{abc}$. After all, while there are some exceptions, many adsorption candidates can be excluded only based on the first criterion. As shown in tab. 3.1, only eight ν_3 candidates remain out of originally 23. Overall it reduces the number of relevant sites to 30 from originally 51 candidates. The second criterion excludes monodentates on alumina oxygen sites, which in most cases were not considered due to steric limitations, but it would also exclude two water adsorption sites, ν_1^a on n(1 1 0) and R(1 1 0). As a consequence of the third criterion, the phosphate protonation state for most cases was HPO_4^{2-}, as at least one oxygen protrudes from the surface and hence cannot be stabilized in a hydrogen bond.

Based on the results for adsorption of a single orthophosphate on the three γ-Al_2O_3 model facets, two additional cases were studied: coadsorption of two orthophosphates and pyrophosphate adsorption. This was critical since it is well established by ^{31}P NMR that the phosphate species change with coverage.[12,59] While there is no tight control over coverage for the computational model, the actual modelled coverage can be calculated based on the surface unit cell. For one phosphate on the (1 0 0) facet, the coverage is $2.2\,\mathrm{P\,nm^{-2}}$, on (1 1 0) it is $1.5\,\mathrm{P\,nm^{-2}}$. These values double when adding a second phosphate to the surface model (4.3 and $3.0\,\mathrm{P\,nm^{-2}}$). The ν_x notation for coadsorption is concatenated in case of coadsorption like in $\nu_x{+}\nu_{x'}$. The new variable x' also leads to water desorption as given in reaction eq. 3.2, depending on the adsorption mode $\nu_{x'}$ of the additional phosphate. For polyphosphates like pyrophosphate, the ν_x notation was extended by additional digits, such as $\nu_{xx'}$. The superscript site information is comma-separated, e. g. $\nu_{22}^{bc,eB}$.

Even after elimination of several candidates by the first criterion for the adsorption of single orthophosphates, the number of adsorption combinations is overwhelming. The full scope of coadsorption includes all combinations of x and x' in $\nu_x{+}\nu_{x'}$. With values from 0 to 3, this gives 10 combinations from lowest to highest overall adsorptions, i. e. ν_{00} to ν_{33}. As shown in tab. 3.1, four adsorption modes lead to at least 30 adsorption candidates, after selecting a subset of structures from 51 initial chemisorption candidates. One also needs to add physisorption here, which adds one candidate for each structural reference, and thus 33 single adsorption cases for recombination to coadsorption and pyrophosphate adsorption. These 33 sites are ten on (1 0 0), ten on n(1 1 0) and 13 on R(1 1 0). The upper limit of combinations for a coadsorption is first estimated by calculating the number of inequivalent site combinations with $\sum_{n=10,10,13} \frac{n(n-1)}{2} = 168$. This higher than the actual number of combinations since, for example on R(1 1 0), ν_1^b, ν_2^{bc} and ν_3^{bcB} all involve site b and thus cannot coexist. Nevertheless, by counting only possible combinations, 107 new coadsorption structures remain. This can be slightly reduced by excluding the less stable site μ_2^a-OH site on n(1 1 0) and R(1 1 0) leaving 86 candidates. The same considerations apply for pyrophosphates $\nu_{xx'}$, with combinations for x and x'. Several more combinations can be excluded due to the limited flexibility of the $H_4P_2O_7$ molecule, but still up to 73 structures would require testing. In the end this leaves us with 180 additional site combinations for coadsorption and pyrophosphates, which would each involve additional sampling for the hydrogen bond network (cf. section 3.2).

We chose to restrict the number of combinations even further, only including those systems systematically, that involve one of the best single orthophosphate adsorption modes. These are combinations of 12 best adsorption geometries with one another, or if this is not possible, involving second best structures of each adsorption mode. Additional calculations were done for the more interesting cases. Overall 118 site combinations were tested. The tested sites are listed in tab. A.2 and A.3 (appendix, p. 156). Based on this exploration a fourth selection criterion was identified:

4. If possible, coadsorbed phosphates are well distributed over the surface.

That means coadsorbed phosphates maximize their distance to one another and avoid neighboring sites if possible. Identifying all important adsorption site candidates on (1 0 0), n(1 1 0) and R(1 1 0) started from single orthophosphate adsorbates, which were evaluated based on their free enthalpies. The best adsorption modes were used to study coadsorption and pyrophosphate, which gave additional insight into coverage effects. Overall 159 of these sites and site combinations were optimized. This way, selection criteria for new adsorption sites were developed, of which only the first is a

Table 3.2.: Most favorable adsorption sites and free enthalpies ΔG and enthalpies ΔH in $\mathrm{kJ\,mol^{-1}}$ at drying conditions, 120°C and 2 kPa H_2O partial pressure.

Facet	Physisorption			Monodentate			Bidentate			Tridentate		
		ΔG	(ΔH)		ΔG	(ΔH)		ΔG	(ΔH)		ΔG	(ΔH)
(1 0 0)	ν_0	−126	(−232)	ν_1^a	−121	(−148)	ν_2^{ab}	−156	(−106)	ν_3^{abc}	−122	(+8)
n(1 1 0)	ν_0	−56	(−147)	ν_1^e	−105	(−121)	ν_2^{de}	−132	(−75)	ν_3^{deA}	−150	(−22)
R(1 1 0)	ν_0	−66	(−79)	ν_1^b	−120	(−62)	ν_2^{eB}	−160	(−30)	ν_3^{bcB}	−168	(+41)

strict limitation (two phosphate oxygen atoms never bind to the same aluminium atom). While – conservatively estimated – up to 203 structures could have been included (i. e. by application of this first criterion), the other three criteria also allow for a more selective approach, within which all plausible sites and site combinations are included in the data set.

3.4. The most stable adsorption modes

As pointed out in sections 3.2 and 3.3 with considerations for creating hydrogen bond networks and with site selection criteria, there are several general rules which can be used to identify a manageable number of candidate structures. Nevertheless, the most stable results comprise exceptional structural features that will be discussed here. It should also be noted that the reported absolute free enthalpy values heavily depend on the environmental conditions, whereas enthalpies are less dependent on temperature. Tab. 3.2 summarizes the results of the four different adsorption modes on the three reference surfaces. The free enthalpy values ΔG and enthalpy values ΔH show very different trends, which can be explained by considering the significant vapor entropy gain after releasing H_2O in the substitution. Comparing systems like, for example n(1 1 0)ν_0 and R(1 1 0)ν_3, one should also consider the amount of water released. For physisorption, no

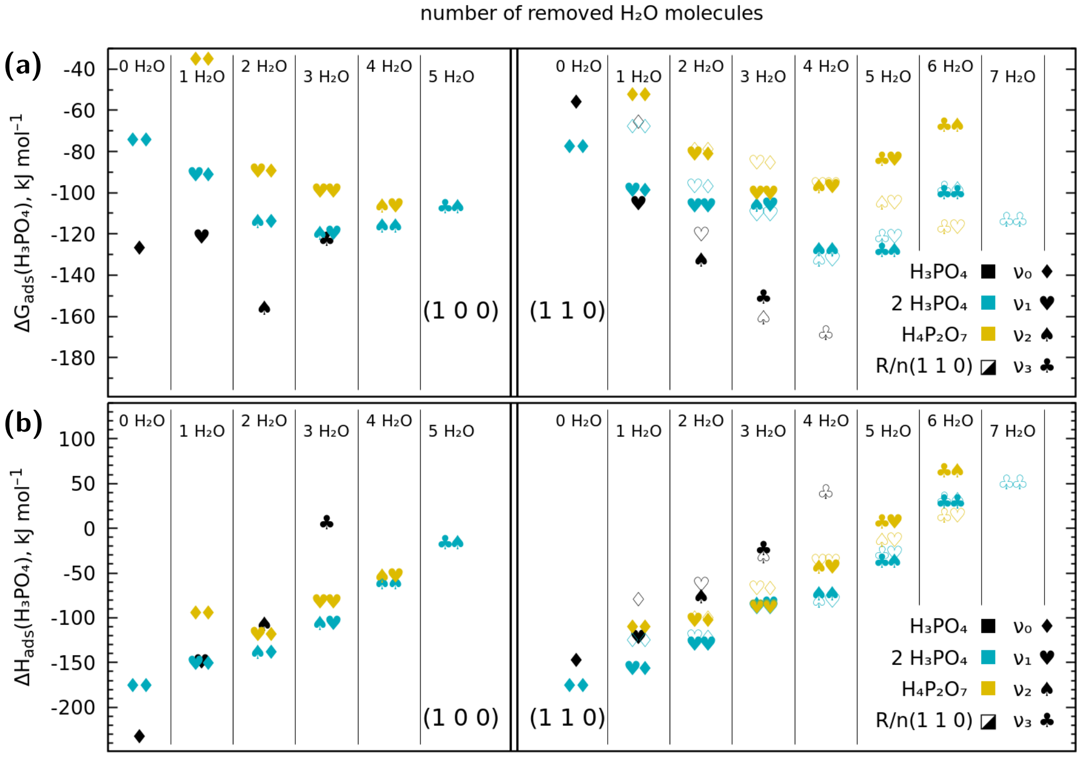


Figure 3.3.: Gibbs free enthalpy (a) and enthalpy (b) of H_3PO_4 adsorption per mole of phosphate at 120°C as a function of the number of released water molecules from the reference surfaces. For R(1 1 0), the common energy reference of the n(1 1 0) surface was used, hence adsorption occurs after removal of one H_2O. Only the lowest enthalpy system of each category is included. Left panels: adsorption on the (1 0 0) surface. Right panels: adsorption on the (1 1 0) facets, n(1 1 0) with filled symbols, R(1 1 0) with empty symbols. Only the free enthalpy of the most favorable system at a given degree of water abstraction was included.

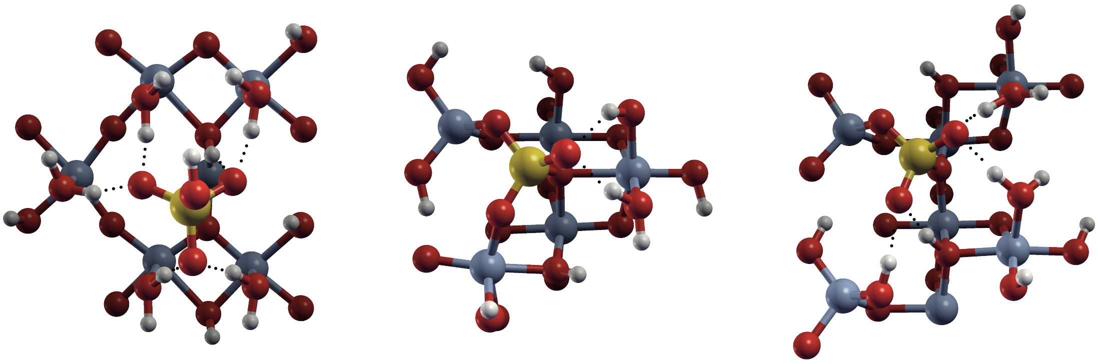
water is released as no substitution occurred but on R(1 1 0)ν_3, three H_2O are released for the ν_3 adsorption and one for surface reconstruction (cf. eq. 3.5). The enthalpy difference between n(1 1 0)ν_0 and R(1 1 0)ν_3 is $-188\,\mathrm{kJ\,mol^{-1}}$ but the free enthalpy $112\,\mathrm{kJ\,mol^{-1}}$, which corresponds to the difference of the entropy contribution $T\cdot\Delta S$. In this example it is $T\cdot\Delta S = 300\,\mathrm{kJ\,mol^{-1}}$ for a vapor entropy gain after releasing four H_2O. When

$(1\,0\,0)\nu_0$ is compared with $\mathrm{R}(1\,1\,0)\nu_3$, the result is similar, with $T\cdot\Delta S = 315\,\mathrm{kJ\,mol^{-1}}$. While identifying the most stable structures, at drying conditions they are ν_2 and ν_3 on $\mathrm{R}(1\,1\,0)$ – indicating that entropy gain by water release is the most decisive term in the free enthalpy results. ΔH and ΔS also slightly depend on the temperature, but the most favorable adsorption sites still remain the same. This balance of enthalpy ΔH and entropy ΔS is linear dependent on the temperature due to $T\cdot\Delta S$, which means that the order by free enthalpy of the tabulated structures in tab. 3.2 is sensitive to the temperature. For example at ambient conditions, i. e. the free enthalpy of $\mathrm{R}(1\,1\,0)\nu_3$ is $118\,\mathrm{kJ\,mol^{-1}}$ and $152\,\mathrm{kJ\,mol^{-1}}$ for $(1\,0\,0)\nu_0$ (cf. tab. A.1, appendix, p. 154). It is also not the case for higher coverages, such as the calculated coadsorption and pyrophosphate adsorption cases. For a general overview in fig. 3.3, ΔG and ΔH were plotted against the number of released H_2O. With eq. 3.3 and 3.5–3.9, the number of released H_2O varies from 0 to 7 relative to the water-covered reference structures. Coadsorption is less favorable than single orthophosphates, which means that phosphate adsorbates compete for surface area, and there is no cooperative adsorption found. This is consistent with the observation, that structures with maximum phosphate distance minimize the free enthalpy. Pyrophosphates are generally less stable even though there are some cases in which pyrophosphates are slightly better stabilized. At drying conditions, up to three or, respectively, four H_2O are substituted on the $(1\,0\,0)$ and $(1\,1\,0)$ facets. Substitution of further H_2O is endergonic at 120°C, but becomes exergonic at higher temperatures (for more data, cf. tab. A.2, appendix, p. 156).

Energy comparisons between the facets $(1\,0\,0)$ and $(1\,1\,0)$ are not ideal for two reasons. First, those energies are based on calculations of two different reference surfaces, neither of which is free of model errors. Second, both systems describe different coverages due to different areas of their surface unit cell. These two aspects are actually very complex to include into a quantitative model, since the coverages on both facets would differ significantly even for slightly different adsorption free enthalpies. It would require to calculate accurate adsorption energies for a range of coverages which would allow us to generate estimates for two local coverages based on the phosphate loading in the sample. In the following sections the coverage on both facets will be considered as equivalent. The coverage effect is qualitatively separated in high and low model coverage domains. Orthophosphate coadsorption and pyrophosphates effectively represent the same coverage, thus the coadsorption is the preferred reference to quantify the stability of pyrophosphates relative to orthophosphates.

3.4.1. Single orthophosphates

Since the order of most stable structures can vary with environmental conditions, the focus on most stable few structures can become arbitrary. In fig. 3.4 three adsorption species from each reference surface were selected based on their exceptional structural features. They are also within the most stable candidates, but at different temperatures. Physisorption is the most stable adsorption mode on (1 0 0) at ambient conditions (see fig. 3.4a). It is adsorbed as HPO_4^{2-} ion, in which the three free P−O oxygen atoms are stabilized by a hydrogen bond pocket consisting of six protons, i. e. two hydrogen bonds per phosphate oxygen. In contrast, the much less stable physisorption cases on n(1 1 0) and R(1 1 0) (cf. tab. 3.2) are $H_2PO_4^-$ ions, stabilized by hydrogen bonds from additional three protons (for two phosphate oxygen, i. e. 1.5 hydrogen bonds each). The high dentation cases n(1 1 0)ν_3^{deA} and R(1 1 0)ν_2^{eB}, where the adsorbed species is fully deprotonated (PO_4^{3-}), are given in fig. 3.4b and 3.4c. Similar to the hydrogen pocket in fig. 3.4a, the free P−O oxygen atoms of the adsorbed species are additionally stabilized by two and four hydrogen bonds from H_2O and OH sites in their vicinity. R(1 1 0)ν_2^{eB} is the only case where such stabilization is possible with two hydrogen bonds stabilizing both remaining free P−O oxygen atoms; other bidentates, like n(1 1 0)ν_2^{de}, are HPO_4^{2-} ions with only one P−O oxygen atom in hydrogen bonds.



(a) $(1\,0\,0)\nu_0$ physisorbed HPO_4^{2-} ion with two hydrogen bonds per phosphate oxygen

(b) $\mathrm{n}(1\,1\,0)\nu_3^{deA}$, tridentated PO_4^{3-} ion; substituted alumina oxygen on $\mu_1^{\mathrm{Al}_O}$, $\mu_1^{\mathrm{Al}_T}$ and μ_3

(c) $\mathrm{R}(1\,1\,0)\nu_2^{eB}$, bidentated PO_4^{3-} ion; substituted alumina oxygen on $\mu_1^{\mathrm{Al}_T}$ and μ_2

Figure 3.4.: Ball-stick models of the most stable cases on n(1 1 0) and R(1 1 0) at 120°C and of (1 0 0) at 25°C. These species are some of the overall best orthophosphate adsorption cases, depending on environmental conditions. $\mu_1^{\mathrm{Al}_T}$ and $\mu_1^{\mathrm{Al}_O}$ are μ_1 sites from tetrahedral or octahedral oxygen atoms. Hydrogen atoms are white, oxygen atoms red, aluminium atoms blue and phosphorus atoms yellow. Hydrogen bonds are illustrated with dots.

It is a remarkable fact that the most stable adsorption mode is found on the reconstructed surface reference, since the hydrated structural reference of n(1 1 0) is $29\,\mathrm{kJ\,mol^{-1}}$ more stable than R(1 1 0) before phosphate adsorption. This stabilization is clearly triggered by the adsorption, but the question is, how exactly. Tab. 3.3 lists all stable bidentate and tridentate adsorption modes ordered by their location on the surface on n(1 1 0) and R(1 1 0). In most of these cases, i. e. ν_2^{bc}, ν_2^{de}, ν_3^{bcE} and ν_3^{deA}, the trend is the same as for the reference, with n(1 1 0) being consistently 16 to $35\,\mathrm{kJ\,mol^{-1}}$ more stable than R(1 1 0). Only ν_3^{bcB} is stabilized by $46\,\mathrm{kJ\,mol^{-1}}$ on R(1 1 0). A second adsorption mode, $\mathrm{R}(1\,1\,0)\nu_2^{eB}$, is considerably more stable than any of the adsorption modes on n(1 1 0), despite the reconstruction. Both sites ν_2^{eB} and ν_3^{bcB} involve the alumina oxygen site B which transforms from a μ_3-O site on n(1 1 0) to a μ_2-OH site on R(1 1 0). For illustration see fig. 3.1 and note that the reconstruction breaks the symmetry in n(1 1 0); thus b'

Table 3.3.: Comparison of bidentate on tridentate free enthalpies, ΔG in $\mathrm{kJ\,mol^{-1}}$, of the same sites on n(1 1 0) and R(1 1 0) at drying conditions, 120°C and 2 kPa H_2O partial pressure. Labels are given for R(1 1 0), on n(1 1 0) c is equivalent to b'.

	ν_2^{bc}	ν_2^{de}	ν_2^{eB}	ν_3^{bcB}	ν_3^{bcE}	ν_3^{deA}
n(1 1 0)	−132	−132	—	−122	−137	−150
R(1 1 0)	−113	−97	**−160**	**−168**	−106	−134

becomes c. It shows, that Wischert's reconstruction is a phenomenon of Digne's (1 1 0) surface model that can occur at low water coverage due to high temperature,[92] or in solution, provided that it can be stabilized with a suitable adsorbate at alumina oxygen site B. In general, it underlines the complexity of these adsorption phenomena, since adsorbate-induced surface reorganization like this one is still very difficult to predict.

3.4.2. Coadsorption and pyrophosphate adsorption

Coadsorption and pyrophosphate enthalpies are determined as described before with eq. 3.6–3.9, where for better comparison with single orthophosphate adsorption, average phosphate adsorption enthalpies are reported. The two best coadsorption and pyrophosphate cases are tabulated in tab. 3.4. Compared to the adsorption of a single orthophosphate on the reference surfaces, coadsorption and pyrophosphate adsorption reach a limit in terms of water abstraction. For single orthophosphate adsorptions, the free enthalpy was dominated by the entropy term, with R(1 1 0)ν_3 (4 H_2O) being the most stable case at 120°C. In comparison, the coadsorption case with most water desorption is R(1 1 0)ν_3+ν_3 (7 H_2O), which is with $\Delta G = -113\,\mathrm{kJ\,mol^{-1}}$ less stable than the most favorable coadsorption cases on n(1 1 0) and R(1 1 0). Removal of 4 or 5 H_2O seems to be the limit for phosphate adsorption at these conditions. A similar trend is observed on the (1 0 0) facet.

Table 3.4.: Free enthalpies of the most favorable adsorption mode for coadsorption and pyrophosphates in comparison to the nextmost stable case with different overall dentation $x+x'$ in $\nu_x+\nu_{x'}$ and $\nu_{xx'}$, sorted by dentation. ΔG in kJ mol^{-1} at drying conditions, 120°C and 2 kPa H_2O partial pressure.

Facet	Orthophosphate coadsorption				Pyrophosphate adsorption			
	sites	ΔG	sites	ΔG	sites	ΔG	sites	ΔG
(1 0 0)	$\nu_2^{ab}+\nu_1^{a}$	−119	$\nu_2^{aa}+\nu_2^{bc}$	−116	$\nu_{11}^{a,a}$	−99	$\nu_{21}^{bc,a}$	−106
n(1 1 0)	$\nu_2^{bc}+\nu_2^{de}$	−127	$\nu_3^{bcE}+\nu_2^{de}$	−128	$\nu_{11}^{e,b\,*}$	−100	$\nu_{21}^{de,b}$	−97
R(1 1 0)	$\nu_2^{eB}+\nu_1^{d}$	−132	$\nu_3^{bcB}+\nu_1^{d}$	−121	$\nu_{21}^{eB,b}$	−105	$\nu_{31}^{bcB,e}$	−117

$^*\nu_{11}^{e,b}$ is a pyrophosphate species that binds with its P–O–P bond positioned over alumina oxygen sites A and B, leaving boundaries of unit cell depictions in fig. 3.1ab.

Regarding adsorption site combinations, it is clear that the second adsorbing phosphate is limited to the remaining sites. One would generally expect that the most favorable sites are occupied first followed by the second most favorable site. However, the sites in tab. 3.3 and tab. 3.4 are not consistent in this regard. In fact, only $(1\,0\,0)\nu_2^{ab}+\nu_1^{a}$ and $\mathrm{n}(1\,1\,0)\nu_2^{bc}+\nu_2^{de}$ specifically fit this expectation. On (1 0 0), the adsorption of two bidentates is only possible by one pattern $\nu_2^{aa'}+\nu_2^{bc}$ that excludes the most stable single bidentate ν_2^{ab}, since the remaining sites a' and c are not direct neighbors. The free adsorption energy of the monodentate $\mathrm{R}(1\,1\,0)\nu_1^{d}$ is $\Delta G = -86\,\mathrm{kJ\,mol^{-1}}$, which makes it very unfavorable on R(1 1 0); ν_1^{c} ($-120\,\mathrm{kJ\,mol^{-1}}$) and ν_1^{e} ($-98\,\mathrm{kJ\,mol^{-1}}$) are better alternatives. Nevertheless, ν_1^{d} is the most stable candidate as coadsorbate with ν_2^{eB} and ν_3^{bcB}, where in each case better sites would be still available. This is likely related to the hydrogen bonds formed with those sites to stabilize ν_2^{eB} and ν_3^{bcB}. The ν_1^{d} adsorption is – for construction of the most favorable site – the best choice, as it does not destabilize the other site. The structures comprise different hydrogen bond patterns but the differences are subtle.

A similar overall trend can be observed for pyrophosphates. The available range of sites

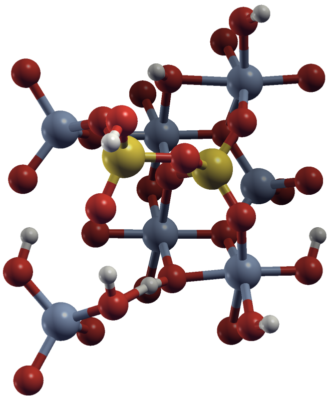


Figure 3.5.: The most stable pyrophosphate species, R(1 1 0)$\nu_{31}^{bcB,e}$ at drying conditions (120°C, 2kPa H_2O partial pressure). With $\Delta G = -117\,\mathrm{kJ\,mol^{-1}}$ it is still clearly less stable than the other coadsorption candidates that were identified.

is more constraint than in case of coadsorbed orthophosphates since both phosphate groups are connected with a bridging P−O−P bond. From their free enthalpies (cf. tab. 3.4), it is evident that pyrophosphates are not better stabilized than coadsorbed orthophosphates at the surface. However, in two cases on R(1 1 0) ($\nu_{31}^{bcB,e}$ and $\nu_{22}^{bc,eB}$), the stabilization is rather close to coadsorbed phosphates. Noticeably, the R(1 1 0) sites B, b, c and e are arranged in such a way that they can bind a pyrophosphate on four sites, while still staying with the selection criteria discussed in section 3.3.

3.5. Summary and perspectives

The result of this structural exploration is a database of overall more than 1000 DFT calculations based on 159 adsorption site candidates. To ensure that the most stable cases were identified, the exploration was done on multiple levels of complexity. With a defined set of structural references, candidate sites had to be defined. This first involved labelling of the relevant surface oxygen sites given in fig. 3.1a, followed by defining a set of relevant adsorption modes based on oxygen site combinations. These adsorption modes were recombined to obtain candidate structures for orthophosphate coadsorption and pyrophosphates. All structures identified in this process required a test for several relevant hydrogen bond networks by applying combinatorics described in section 3.2. Several general trends and features were identified just by comparing their thermodynamic properties. One of the important findings is that the presence of phosphates leads to surface reorganization, here observed by the stabilization of Wischert's reconstruction. It was proposed in several experimental studies,[12,59] but this work is the first successful attempt to reconstruct these effects in computational atomistic models. The database collected as described in this chapter was the basis for further studies. As an attempt to verify the result database, GIPAW chemical shift simulations were conducted for the most stable structures. The results and performance in comparison to the experiment are discussed in chapter 4. With the ambition to better understand the structure- NMR relationship, the chemical shift simulations of this database were used to train a machine learning model. The outcome of this is given in chapter 6. Finally, in order to cover phosphate species at impregnation conditions, the best database structures were used as target geometries in state-of-the-art metadynamics simulations (chapter 5). In particular, high dentation patterns were probed in order to see whether they can form by means of a phosphorylation. Moreover, based on thermodynamic estimates, the condensation of pyrophosphates was excluded within the tested models. This has important consequences, as experimental evidence demonstrates their presence. Pyrophosphate and other polyphosphates either come from the solution, they condensate faster than the most stable adsorption modes can form or they are a consequence of inhomogeneous distribution in the material. These three alternatives are discussed in detail in section 4.3 (p. 4.3), after a detailed analysis of the prediction performance of these models for ^{31}P NMR spectroscopy.

4. Combining NMR and DFT

Experimental and theoretical work have always been complementary. The discussed ^{31}P and ^{27}Al NMR experiments involve – even without computational models – a considerable theoretical foundation in order to infer chemically relevant information from the raw spectral data. Many regular single component systems can be well characterized with different experimental methods, allowing accurate experimentally derived models. In contrast, disordered multicomponent mixtures are usually too complex to permit a complete atomic scale description that is exclusively based on experimental insight. In the past, these knowledge gaps had been commonly filled with more or less speculative hypotheses in lack of better alternatives. Computational modelling is different from the theory needed to interpret experiments, since it is carried out separately. As such, computational results can be evaluated independently, but since they do not involve any measurements of a real chemical system, their relevance is sometimes disputed. Combining NMR cross-polarization experimental studies with computational modelling can therefore improve the discussion at both sides.

In the structural exploration (chapter 3), we presented computational results, identified several trends within the most stable adsorption modes and sites and rationalized them by the balance between enthalpy and H_2O vapor entropy. As described there, the exploration was comprehensive and it is therefore unlikely that the most representative structures were overlooked – but that only includes the three model surfaces tested. From a computational perspective, it is very difficult to proof that the explored phosphate adsorption geometries are relevant models to describe the real systems, thus the structures have to be validated against the experimental results. On the other hand, the unambiguous discussion of the experimental findings requires a lot of results to be reconciled, where drawn figures of adsorption patterns are often speculative. In computational models, however, these presumptions can be evaluated and compared quantitatively; essentially by composing structures *in silico*, and benchmarking it by computation of spectroscopic data from the model geometry. For this work, it is particularly beneficial to compare calculated chemical shifts of different adsorption modes and sites and compare them with the experiment. This is also considerably different to some studies with a strong focus on experimental results and only few computational models that are produced to illustrate the conclusions.

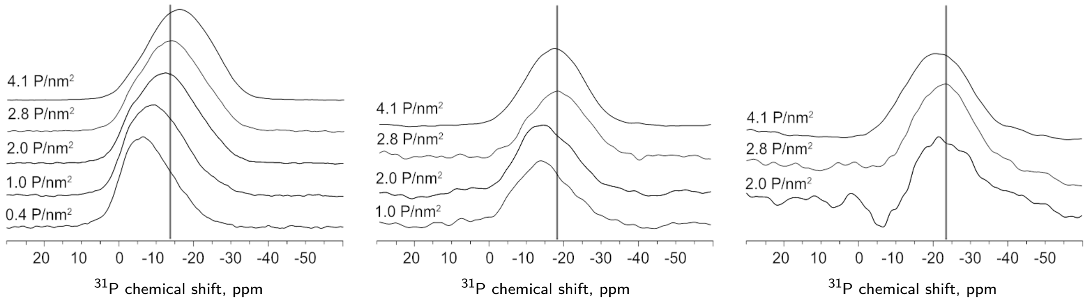
The goal in this chapter is to determine the extent to which the model geometries represent the real phosphate speciation on γ-Al_2O_3. It requires to identify which computational models reproduce the experimental ^{31}P NMR signal band of dried H_3PO_4-impregnated γ-Al_2O_3. This also requires a computational NMR reference, because – equivalently to NMR measurements – a reference substance needs to be computed in order to compare the computational findings with literature data. Since simulations of 85 % phosphoric acid, the ^{31}P NMR standard reference, are not feasible, aluminium phosphate minerals with known crystal structures and NMR iso-shifts were used instead. While a single one could be used as a reference, three compounds – berlinite, variscite and wavellite – were used in order to minimize systematic errors. The reference is explained in detail in appendix A.3. In order to compare the most stable candidates discussed in section 3.4 and 3.4.2, correlation graphs were produced where the chemical shifts from GIPAW calculations are plotted against the phosphate adsorption free enthalpy.

4.1. Experimental results

As discussed in section 1.1.5 (p. 23), DeCanio et al. showed with ^{31}P NMR, that two aluminium phosphate phases can form when preparing samples with 4 to 10 wt% H_3PO_4

into γ-Al_2O_3, which corresponds to coverages of approximately 3.6 to 9 P nm^{-2}. They rationalized the initial signal shift from 1 to 4 wt% (≈0.9–3.6 P nm^{-2}) with polyphosphate formation.[12] Moreover, DeCanio et al. observed a sudden drop in γ-Al_2O_3 pore volume between phosphate loadings of 6 and 10 wt% (≈5.4–9 P nm^{-2}), which is one reason why our experimental collaborators focussed only on loadings up to 4.1 P nm^{-2}. Some experimental work on similar systems also attempted to resolve the adsorption mode of phosphates. Li et al. studied the adsorption of phosphates on boehmite using ^{31}P REAPDOR CP MAS NMR experiments. They identified bidentation (ν_2) as a more favorable adsorption mode than monodentation (ν_1) of orthophosphates on boehmite on dried and wet samples.[72,123] While this is an important finding and a comparable behaviour is expected on γ-Al_2O_3, the authors did not include other adsorption modes like tridentates (ν_3) or physisorption (ν_0). Overall, these studies set a range of species and adsorption modes to be identified, and also a “confined” range of requirements that the computational models need to satisfy. The focus here is on phosphate loadings at which DeCanio et al. proposed that condensed phosphates would influence the ^{31}P NMR signal the most.

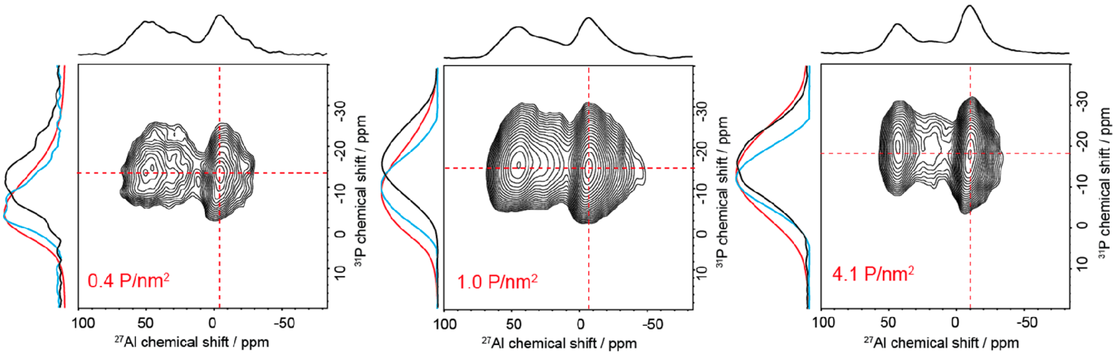
Fig. 4.1 shows NMR spectra acquired by D. Wisser and A. Lesage on γ-Al_2O_3 samples provided by IFPEN based on Dynamic Nuclear Polarization Surface Enhanced NMR Spectroscopy (DNP SENS) techniques.[18] The ^{31}P CP NMR signal in fig. 4.1a confirms the trend observed by DeCanio et al.: an almost featureless signal ranging from 0 to −20 ppm at low coverage. With increasing coverage, the signal broadens and its maximum is shifted to lower values. At the highest experimentally tested coverage of 4.1 P nm^{-2}, the signal band covers an area with an upfield shift to −30 ppm. For the first time they detected the presence of pyrophosphates and other short-chained polyphosphates by using double and triple quantum correlation spectra (fig. 4.1b and 4.1c). The ^{31}Al−^{27}Al-INEPT NMR experiments in fig. 4.1d showed clearly that the phosphates are chemisorbed to the γ-Al_2O_3 surfaces by P−O−Al bonds. However, it is surprising that this apparently



(a) Full ^{31}P NMR signal at varying coverage.

(b) ^{31}P NMR experiment to measure ^{31}P–^{31}P coupled signal selectively.

(c) ^{31}P NMR experiment to measure ^{31}P–^{31}P triple quantum coupling.



(d) ^{31}P–^{27}Al through-bond INEPT correlation spectra. The 1D ^{31}P CP MAS spectra (red line) as well as the projections along the indirect dimension of the 2D through-bond (black line) and through-space (blue line) spectra are reported on the left of the 2D maps. The projection along the ^{27}Al dimension is shown above the plots.

Figure 4.1.: Different DNP enhanced NMR techniques recorded at 9.4 T (400 MHz ^{1}H resonance, 263 GHz microwave frequency). The ^{31}P NMR signal (a) is shifted with the overall phosphate coverage. The results of the 1D INADEQUATE experiments (b) and (c) provide the evidence for the presence of condensed phosphates. Vertical line follows the signal maximum at a coverage of 2.8 P nm^{-2}. The 2D ^{31}P–^{27}Al INEPT correlation experiments (d) lead to new conclusions regarding the nature of P–O–Al bonds with the surface (discussed in main text). NMR spectra obtained at CRMN by D. Wisser and A. Lesage on γ-Al_2O_3 samples provided by IFPEN with the ROAD4CAT project.[18]

does not concern all phosphate centers. The comparison between through-bond and through-space ^{31}Al$-^{31}$P-INEPT experiments (left side of the 2D graphs in fig. 4.1d) reveals that the signal bands deviate significantly, which indicates that the down-field part of the overall signal is represented by physisorbed phosphates or dangling groups of terminally binding polyphosphates. A more subtle observation is that also the ^{27}Al signals are slightly shifted to positive values for increasing phosphate loading. As previous work and these results show, many relevant aspects of phosphate adsorption are accessible by experiments. However, currently there is no alternative to computational methods and simulations in order produce a coherent model which can be benchmarked against the experimental observations. Moreover, future experiments on even more complex systems could rely on such models as foundation.

4.2. Computational chemical shifts

In order to completely characterize phosphate adsorption at the γ-Al_2O_3 surface, one should take into account the following aspects:

- Formation of aluminium phosphate phases[12,59]
- Polyphosphate formation[12]
- Coadsorption or coverage effect (interaction with other phosphates in vicinity)
- Adsorption mode effect (i. e. dentation, for boehmite discussed in ref. [123])
- Adsorption site effect
- Stability (discussed for calcination in ref. [12])

Several of these have been studied before, but it was not yet possible to include these aspects in a coherent model. The coming chapters demonstrate how DFT calculations can enhance our understanding by comparing these results with model chemical shifts. The GIPAW method[66,106] employed in this work explicitly calculates the shielding tensor from the electronic structure for every nucleus of the system. Nevertheless, in the following sections, only the chemical shift of ^{31}P nuclei will be considered.

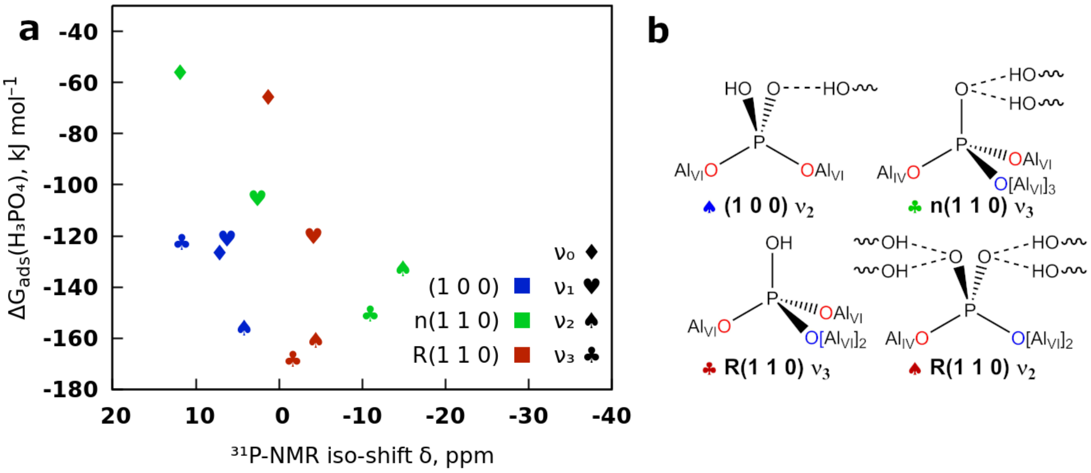


Figure 4.2.: (a) Plot of the relationship between chemical shift and free enthalpy for single adsorbed phosphates (1.5–2.2 P nm^{-2}). (b) Structures of the four most stable adsorption sites. Oxygen sites from H_2O adsorption are labelled in red and alumina oxygen sites in blue.

4.2.1. Chemical shift of orthophosphate

For a single adsorbed orthophosphate species, the coverage of each model facet is determined by the surface area of the unit cell, that is 0.46 nm^2 for the (1 0 0) facet and 0.66 nm^2 for the (1 1 0) facet. This corresponds to local coverages of 2.2 and 1.5 P nm^{-2}. These local coverage values cannot be related to experimental phosphate loadings, as the phosphates are not expected to be equally distributed over the different facets. When comparing the results with experimental data, one also needs to bear in mind that theoretical models are built at the pH corresponding to the point of zero charge. For the computational models, it just means that the calculated systems are uncharged, but for the experimental part it is required to ensure the absence of ions except for the phosphates during sample preparation. GIPAW calculations were conducted for the most stable sites for each adsorption mode – ν_0, ν_1, ν_2 and ν_3 – and structural reference, (1 0 0), n(1 1 0), R(1 1 0). The stability of the modelled phosphate species was already discussed in detail in chapter 3. In this chapter, only the most stable systems are considered for each adsorption mode. In fig. 4.2, the simulated ^{31}P chemical shifts are plotted against the phosphate adsorption free enthalpies defined in section 3.1 (p. 52).

The adsorption on the (1 0 0) facet is less stable than on (1 1 0), thus more phosphates are expected to bind to the latter. The signals of phosphates on (1 0 0) are between 12 and 4 ppm, which is not within the experimental signal range that starts from 0 ppm. Thus, the phosphate does not seem to adsorb on the (1 0 0) at low coverage. For the (1 1 0) facet, the most stable adsorption modes describe the upfield half of the experimental ^{31}P NMR spectrum. The best candidate is the tridentate R(1 1 0)ν_3 with a chemical shift of -2 ppm, which is at the most positive end of the experimental range. The other tridentate, n(1 1 0)ν_3, leads to a chemical shift of -11 ppm, just close to the band maximum of the lowest coverage in fig. 4.1a. The bidentates R(1 1 0)ν_2 and n(1 1 0)ν_2 are both deshielded compared to their tridentate counterparts. Hence, if the amount of bidentate increases, the overall signal would be shifted to lower values. The most stable monodentates (ν_1) and physisorption cases (ν_0) are 20 kJ mol^{-1} less stable than the most favorable site on each reference structure, and in most cases not in the chemical shift experimental range. The fact that the most stable adsorption structures are in agreement with the experiment is a clear validation of our computational method, and allows us to analyze further trends. The next section takes a more elaborate view on ^{31}P chemical shift for coadsorption.

4.2.2. Orthophosphate coadsorption

As explained in section 3.1 (p. 52), the coadsorption free enthalpy is the average free enthalpy of adsorption for both phosphates. This is the only value that can be directly compared with single orthophosphate adsorption. Nevertheless, each coadsorption system produces two results for the GIPAW ^{31}P chemical shift simulations. Fig. 4.3a presents the

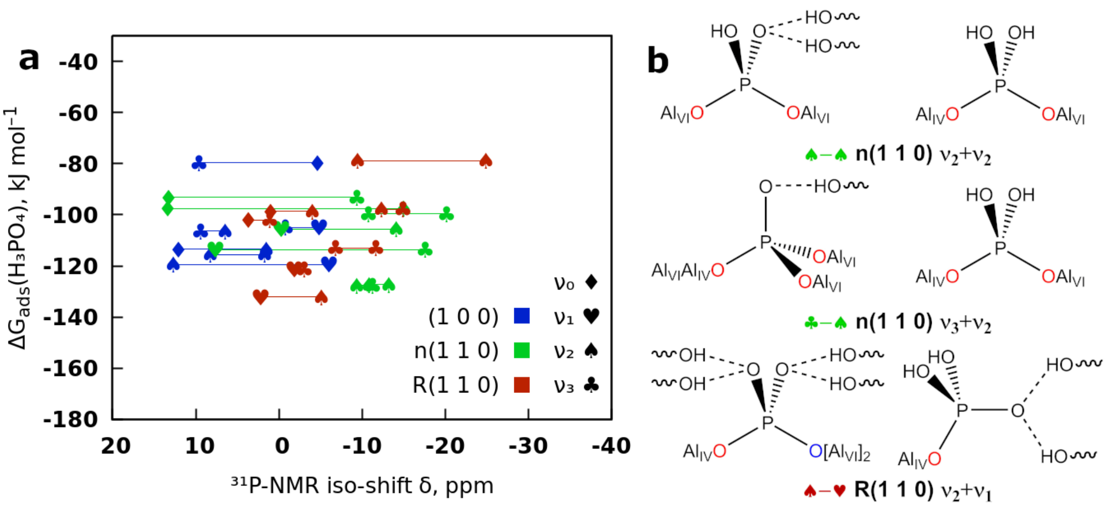


Figure 4.3.: (a) Plot between chemical shift and free enthalpy of orthophosphate coadsorption (3.0–4.3 $\mathrm{P\,nm^{-2}}$). Colors and symbols refer to the reference surfaces (cf. fig 3.1) and adsorption modes, respectively. The chemical shifts of two ^{31}P nucleus in the same system are connected by a horizontal line. (b) Chemical environment of the three most stable adsorption sites combinations.

results for coadsorption in all cases that involve at least one bidentate (ν_2) or tridentate (ν_3), as the other systems are less stable and thus less likely to be present. This does not strictly exclude physisorption and monodentates, only the combinations $\nu_0 + \nu_0$, $\nu_1 + \nu_0$ and $\nu_1 + \nu_1$ are not shown. The most stable adsorption mode combination is R(1 1 0)$\nu_2 + \nu_1$ since the number of abstracted water molecules per surface unit is limited by the balance between entropy gain and enthalpy loss, as discussed in section 3.4. All the three most stable systems (cf. fig. 4.3b) consist of one or two bidentates, whereas only one also consists of one tridentate. This is a stability related change of the adsorption mode, from ν_3 to ν_2 with coverage increase. As proposed in the previous section, this change of the predominant adsorption mode should lead to a chemical shift to lower values. In fact, the chemical shifts of the three systems range from 2 to −13 ppm but with most of signal at lower values, confirming this trend. The (1 0 0) facet, which was excluded based on the modelled chemical shifts for single orthophosphate adsorption, now produces several ^{31}P signals at chemical shifts of coadsorbed monodentates (ν_1) around -5 ppm albeit all of them are coadsorbed with another predicted signal that is outside the experiment range. Even though it should in general be expected to find phosphates on the (1 0 0) facet, as soon as the most stable adsorption sites on (1 1 0) facet are occupied, these monodentation (ν_1) signals could as well be characteristic to any of the monodentates. Monodentates on all facets produce calculated ^{31}P chemical shifts between 6 and -6 ppm. This also concludes our findings for the influence of the adsorption mode. The predominant species, bidentate (ν_2) and tridentate (ν_3), can contribute to an upfield shift of the ^{31}P NMR signal, whereas lower dentation modes (ν_0, ν_1) should have a minor influence to the NMR signal. Another aspect is the decreased free enthalpy difference between adsorption based on the n(1 1 0) reference structure and R(1 1 0). The chemical shifts on n(1 1 0) are clearly at more negative values than those on R(1 1 0). However these differences are not easy to disentangle with the general view on the best adsorption mode combinations, but require a closer look at individual adsorption sites. This is given in fig. 4.4 and will be discussed in the next section.

4.2.3. Coverage and site effect

The previous sections validated the models by comparison with the chemical shift calculations and furthermore analyzed the effect of the predominant ν_2 and ν_3 adsorption modes. In this section, the chemical shifts of the six most stable adsorption candidates at varying coverages will be compared. With data for the orthophosphate adsorption and coadsorption, two different model coverages are already tested. To calculate chemical shifts at lower model coverages than in case of single orthophosphate adsorption, the unit cell surface was increased by constructing $p(\sqrt{2}\times\sqrt{2})$ super cells. GIPAW calculations were carried out on those surface super cells, providing model coverages decreased by 50 %. This means that three model coverages are available on both facets; on (1 0 0) they are 1.1, 2.2 and 4.3 P nm^{-2} whereas on (1 1 0) they are 0.8, 1.5 and 3.0 P nm^{-2}.

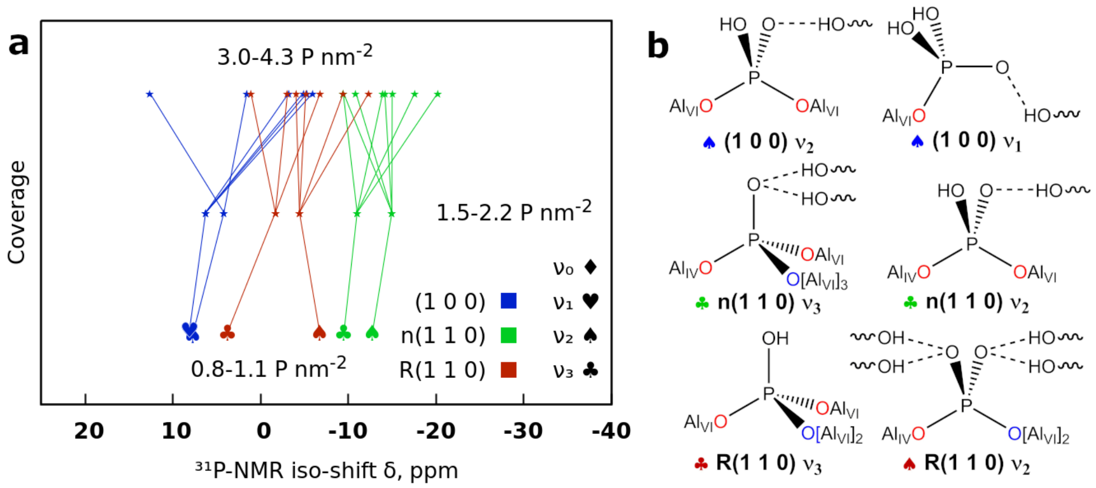


Figure 4.4.: (a) Coverage effect on the ^{31}P chemical shift for six selected sites, i. e. the two most stable adsorption sites of each reference structure: $(1\,0\,0)\nu_1^{a}$, $(1\,0\,0)\nu_2^{ab}$, $\mathrm{n}(1\,1\,0)\nu_2^{de}$, $\mathrm{n}(1\,1\,0)\nu_3^{deA}$, $\mathrm{R}(1\,1\,0)\nu_2^{eB}$ and $\mathrm{R}(1\,1\,0)\nu_2^{bcB}$ (for labels cf. fig. 3.1a, p. 53). The phosphate coverages are given as ranges, because the two model facets have different surface areas. The same sites are connected by lines. At the highest coverage, the site of the coadsorbing phosphate cannot be chosen unambiguously, so several cases were plotted.

The highest coverages (coadsorption) depend on the choice of the other coadsorbing phosphate. Since there is no simple criterion to pick this other phosphate, all adsorption mode combinations were plotted. Fig. 4.4a shows the effect of coverage while keeping the site constant and fig. 4.4b provides structures of the tested sites. Note that this plot has a focus on *sites* which means that all connected points not only share the same structural reference and the adsorption mode, but they are bound to the model surface at the same position and orientation. They are different in where and how their surrounding rearranges depending on the presence of neighboring phosphates. It seems that there is a small trend of decreasing chemical shifts at higher coverage, which could be called "direct" coverage effect. The only exception is the monodentate $(1\,0\,0)\nu_1^{a}$, which shows a strong deshielding for the highest coverage. As this observation is unique to monodentates, which are much less stable based on the thermodynamic results (cf. fig. 4.2), this aspect was not further examined.

The five considered bidentates (ν_2) and tridentates (ν_3) lead to fairly separated chemical shifts all over the range. The chemical shift of each site does not change systematically with increased coverage but might be different within a range of 5–10 ppm depending on their surrounding. The chemical shift difference between the adsorption sites is in comparison much more significant. For all bidentates (ν_2) and tridentates (ν_3), the signals are in distinct areas, even at high coverage where overlapping ranges are overall in the same order. Combined with the stabilization of adsorption on n(1 1 0) sites, this would introduce a significant change of chemical shift to negative values for increased coverage, if these sites become more populated. At low coverage, the predominant adsorption species are found at reconstructed R(1 1 0) surface sites. With increasing the coverage, contributions of adsorption sites identified on the n(1 1 0) structural references become more and more important.

4.2.4. Adsorbed pyrophosphates

Pyrophosphates are the only condensed phosphates discussed in this work, due to the excessively large phase space that would need to be explored otherwise for additional polyphosphates. They are evaluated in a similar way as the coadsorption cases, where also each system produces two chemical shifts and one free enthalpy.

The two most stable pyrophosphates, R(1 1 0)$\nu_{31}^{bcB,e}$ and R(1 1 0)$\nu_{22}^{bc,eB}$ (for labels cf. fig. 3.1a, p. 53), share the same oxygen surface sites and have almost the same free enthalpy. Their calculated signals are in a range from −16 to −27 ppm (cf. fig. 4.5). The third most stable pyrophosphate is on the (1 0 0) facet. The chemical shift of both groups is at about −10 ppm, which is in agreement with the experimental range. Of all calculated chemical shifts on the (1 0 0) facet, pyrophosphates are the only results that align well with the experiment. Hence it is possible that pyrophosphates are more present than

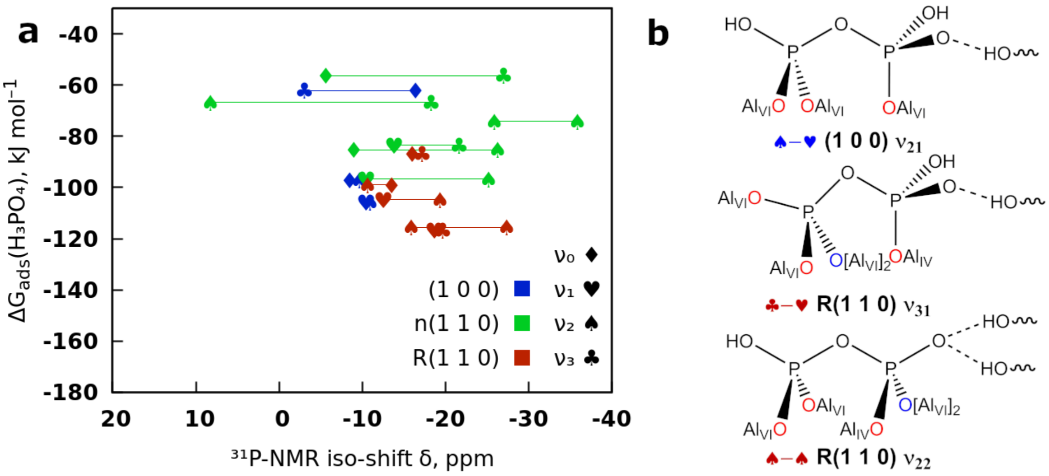


Figure 4.5.: (a) Correlation between chemical shift and stability of pyrophosphates (1.5–2.2 P nm^{-2}). The chemical shift of two ^{31}P nuclei of the pyrophosphate are connected by a horizontal line. (b) Chemical environment of the three most stable adsorption sites.

orthophosphate species on (1 0 0). The thermodynamic prediction on the model surfaces disfavor the presence of pyrophosphate, which is a weakness of the model at this stage. This is likely to improve by exploring higher coverages (4.5-6.5 P nm^{-2}), i. e. a comparison of three orthophosphate coadsorption $\nu_x+\nu_{x'}+\nu_{x''}$ vs. pyrophosphate/orthophosphate coadsorption $\nu_{xx'}+\nu_{x''}$ vs. triphosphates $\nu_{xx'x''}$. Since the most stable adsorption sites are generally occupied with two phosphates, and physisorption is in most cases much less stable, pyrophosphates might be more favorable at this stage. Overall the pyrophosphate adsorption cannot yet be conclusively explained, but different rationales for the presence of pyrophosphates are presented in the discussion in section 4.3.

4.2.5. Terminal pyrophosphates or physisorbed orthophosphates

The ^{27}Al–^{31}P-INEPT NMR experiments (see fig. 4.1d) showed that a significant part of phosphates does not directly bind to the alumina surface. The through-bond INEPT signal projected for ^{31}P nucleus, is shifted by around -6 ppm relative to the through-space INEPT signal, which shows that different phosphate species are involved in both cases. The best candidates in for those nonbonded cases are

- physisorbed phosphates, ν_0 and
- terminal pyrophosphates, ν_{x0}.

The calculated ^{31}P chemical shifts for physisorption are in most cases at significantly more positive values than the experimental range. The few exception that match the experiment are significantly less stable than the former. In contrast both computed ^{31}P NMR chemical shifts for any of the ν_{x0} are in agreement with experiment. While it might still be possible that physisorbed phosphates are misrepresented due to their more dynamic behaviour (without covalent anchors they are more mobile), from the current data, pyrophosphates provide a good rationale for these experimental findings.

4.3. Summary and perspectives

The computed chemical shifts of the most stable model adsorption candidates agree well with the experiments. In particular the most stable adsorption modes – bidentates (ν_2) and tridentates (ν_3) on (1 1 0) – consistently cover experimental range of ^{31}P NMR chemical shifts from 0 to -15. The computed pyrophosphate model systems complement the calculated ^{31}P chemical shifts for orthophosphate by covering a ^{31}P NMR signal range from -10 to -28 ppm. Several different aspects were tested that change with coverage. The preferred adsorption mode shifts from tridentates (ν_3) to bidentates (ν_2) at increased coverage, given the change in thermodynamic stability and simultaneously in agreement with the downfield shift of the ^{31}P NMR signal. Fig. 4.1d clearly shows the site dependence of chemical shifts, whereas the effect of different phosphates and protonation networks does not change the ^{31}P chemical shift in a clear direction. Pyrophosphates and longer polyphosphate chains were for the first time directly detected by our collaborators' ^{31}P INADEQUATE double and triple quantum correlation experiments. The calculated ^{31}P chemical shifts of orthophosphate on the (1 0 0) facet are in most cases at considerably more positive values, indicating that the facet is almost negligible compared to the (1 1 0) facet. Pyrophosphates models on (1 0 0) are in a much better agreement, which is expected due to the downfield chemical shift of condensed phosphates. Similar to the cases on the (1 0 0) facet, the chemical shifts of physisorbed orthophosphates do not agree well with the experiments. While the difference between through-space and through-bond ^{31}P$-$^{27}Al INEPT spectra, is evidence for nonbonded phosphate sites, rather than physisorbed phosphates, those are more likely dangling groups of ν_{x0} bonded pyrophosphates.

The thermodynamic prediction for phosphate adsorption does not permit the presence of pyrophosphates, because coadsorbed orthophosphates are overall more stable. This result is not in alignment with the experimental findings, as pyrophosphate are clearly observed. There are three rationales for this. One possibility is that pyrophosphates come from the solution and adsorbed before being hydrolyzed. This is probable if they were part of the impregnation solution due to their kinetic stability.[45,47] It is not expected that they form in aqueous solution, due to the positive reaction enthalpy;[47] significant phosphate condensation only occurs at high phosphate concentrations with trace amounts of water.[124] This possibility can be tested by reevaluation of the synthesis. Another rational could be found in a more complex relationship between local coverages and phosphate loading. If phosphate chemisorption is rapid and the adsorbates at the solid-liquid interface are too stable, then the adsorption process is exclusively under diffusion control. In turn, phosphates are immobile and remain at the sites they bound to first would lead to an inhomogeneous surface distribution of phosphates, between outermost areas and pores. The same trend is expected if the most stable adsorption sites are concentrated at relatively small facets, which in turn produce high local phosphate coverages there. Since this initial rapid chemisorption would only occur on the reactive surface sites, this effect should be more stronger in case of low phosphate loadings since the remaining free phosphates would adsorb elsewhere after the most stable sites are occupied. Testing this hypothesis would involve additional structure/free enthalpy explorations at higher coverage and likely needs to be extended by different surface models. Atomic scale and mesoscale models for phosphate diffusion in the pores might be needed as well. As last alternative, pyrophosphates may also form at the surface at lower coverage, if they are kinetically more stabilized than coadsorption. However, this also requires that the reaction barrier of the most stable orthophosphate adsorption modes is too high for these cases to occur, which would exclude some of the most stable model adsorption structure based on kinetics.

In general the GIPAW calculations for ^{31}P NMR demonstrate that different structural aspects as well as facet distribution have to be considered simultaneously. The computational models provide a considerably expanded foundation to discuss these effects. However, we are still at the beginning of understanding all aspects of phosphate surface speciation. Reaction kinetics need to be studied in order improve the comparison between orthophosphates and pyrophosphates. This is much more complicated and computationally demanding, but as a first step in this direction, the kinetics of one adsorption reaction $H_2O/\gamma\text{-}Al_2O_3$ interface and several phosphorylation reactions (cf. section 1.1.3, p. 18) will be studied in chapter 5.

5. Towards alumina impregnation

After exploration of phosphate adsorption on dried γ-Al_2O_3, one can define a set of target geometries and ask how they formed through wet incipient impregnation. The most stable adsorption modes in dry samples, bidentates (ν_2) and tridentates (ν_3), may not be the most stable modes during impregnation.[125] However, experiments showed that once impregnated, the phosphates cannot be removed easily without specific leaching treatment. This chapter provides insight into stabilities of the explored adsorption modes in chapter 3 and demonstrates how well-tempered metadynamics (WTmetaD) can be used to explore reaction paths producing some of the described dentations. The most favorable adsorption sites under drying conditions were in focus. In particular, the presence of adsorbed pyrophosphates is evident and they likely have a key role to rationalize the coverage-dependent ^{31}P NMR shift. Here, we want to take a closer look at how the initial adsorption could take place and how pyrophosphate can be involved in some reaction steps. The goal is in particular to find out whether the higher dentation states, i. e. bidentates and tridentates that involve alumina oxygen atoms, are kinetically possible and if they are still stable compared to the other adsorption modes. This chapter also focussed on the (1 1 0) facet, since chapter 4 showed that the (1 0 0) facet is less active for H_3PO_4 adsorption.

The adsorption of orthophosphates can in general be thought of a reaction sequence from physisorption to the highest possible dentation mode:

$$H_3PO_4(aq) \longrightarrow \nu_0 \underset{+^{\star}H_2O}{\overset{-^{\star}H_2O}{\rightleftharpoons}} \nu_1 \underset{+^{\star}H_2O}{\overset{-^{\star}H_2O}{\rightleftharpoons}} \nu_2 \underset{+^{\star}H_2O}{\overset{-^{\star}H_2O}{\rightleftharpoons}} \nu_3 \tag{5.1}$$

Starting from phosphoric acid in solution, the adsorption species are formed in the sequence: physisorption (ν_0), monodentate (ν_1), bidentate (ν_2), and tridentate (ν_3). The first reaction of physisorption of H_3PO_4(aq) from bulk water was not tested here. On the other hand the reaction to form monodentates (ν_1) involves the substitution of one adsorbed $^{\star}H_2O$ molecule as discussed in section 3.1 (p. 52). This can occur a second time, in which a free P−O oxygen in the monodentate substitutes another preadsorbed $^{\star}H_2O$. The preadsorbed water sites are either μ_1-H_2O or μ_1-OH, the latter of which has to be protonated before desorption. Unlike in the drying state, this desorbed water molecules become part of the solution, hence the entropy effect is expected to be significantly lower. The H_2O abstraction mechanism to form tridentates (ν_3) on the (1 1 0) facet is considerable more complex, since the formation of another bond with the γ-Al_2O_3 surface now involves a triangular pattern of surface oxygen sites. On the (1 1 0) facet this always involves at least one alumina oxygen atom.

The next section demonstrates the possibility of predicting barriers with WTmetaD simulations by computationally reproducing the experimentally determined orthophosphate condensation barrier. Then the focus is on the first chemisorption steps for the formation of mono- and bidentates (ν_1, ν_2) from physisorbed orthophosphate (ν_0). At last the question whether tridentates (ν_3) can form at the solid liquid interface is addressed. All WTmetaD calculations were done at a simulation temperature of 330 K with a bias factor γ of 72.9 and gaussian depositions with a height of $3\,\mathrm{kJ\,mol^{-1}}$ every 10 fs if not noted otherwise. Furthermore parallelization using four walkers, with two starting at the initial state and two at the final state of the reaction.

5.1. Phosphate condensation in bulk water

The condensation and hydrolysis of pyrophosphate is the most fundamental phosphorylation reaction (cf. section 1.1.3, p. 18) because it only involves the two smallest polyphosphates, orthophosphate and pyrophosphate. At 120°C, 10 % of water-free orthophosphoric acid is converted to pyrophosphate, and trace amounts of longer polyphosphate chains can be measured.[124] However, pyrophosphate is not thermodynamically stable in aqueous solutions, thus they do not form such solutions, but existing phosphates hydrolyze at very slow rates (more than 21 h half-life at 65°C,[45] or several days at room temperature).[31] In particular the cyclic polymetaphosphates can remain in solution for weeks, before significant decomposition.[31] In this section, the pyrophosphate hydrolysis reaction is discussed in aqueous solution. It is used as a testbed of the method since the barrier and reaction energy are known by experiments for the hydrolysis.[47] Then in the next sections, these findings are compared to phosphate speciation reactions via phosphorylation on γ-Al_2O_3.

Condensation and hydrolysis in phosphorylation reactions are quite complex as they involve two $P-O$ bond formations and may also include several proton shifts, involving the solvation shell.[40,50,51] Several theoretical studies were done using AIMD and metadynamics, but typically within the context of biochemistry, involving nucleoside polyphosphates in presence of magnesium and within enzymes.[49,51] To our best knowledge, the reaction of the pure compounds (the phosphates in water) was not yet simulated by AIMD/WTmetaD. For brevity, the following discussion only refers to the condensation reaction and neglects the hydrolysis since it is equivalent. The system consists of 124 H_2O and two H_3PO_4 which reflects a solution of 1 M phosphoric acid with pH=1-2. It is generally complicated to identify good collective variables (CV). Since the system is symmetric, starting from two orthophosphates, one would ideally find CV that respect this symmetry. However, the symmetry is broken in the course of the reaction, in which only

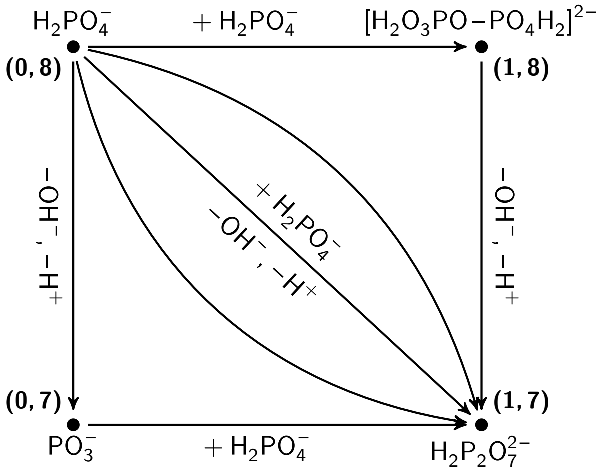
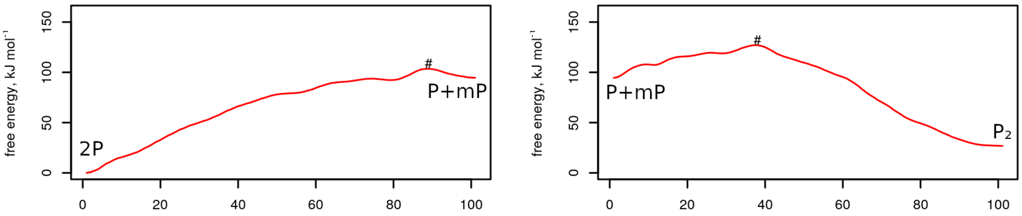


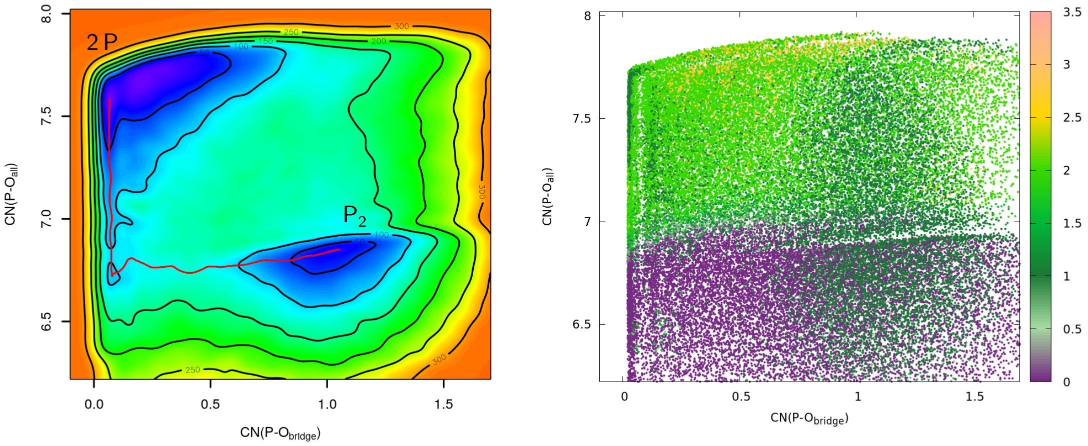
Figure 5.1.: Schematic of the orthophosphate condensation of phosphoric acid in aqueous solution. $H_2PO_4^-$ is more stable than H_3PO_4. The coordinates of the different structures of each path in the basis of CN(P−O_{all}) and CN(P−O_{bridge}) (see main text) on x and y-axis are given as 2D vectors (x, y).

one phosphate provides the bridging oxygen atom. For a two-dimensional well-tempered metadynamics simulation, that was discussed in section 1.1.3 (p. 18), the following two collective variables were identified: (i) a variable which counts the number of phosphorus-bridging oxygen atoms keeps track of the pyrophosphate bond formation, CN(P−O_{bridge}), and (ii) the number of phosphate-binding oxygen atoms, CN(P−O_{all}), counts the total number of oxygen atom binding to phosphorus. The CN(P−O_{bridge}) should vary from 0 (orthophosphate) to 1 (pyrophosphate) and CN(P−O_{all}) from 8 (2 H_3PO_4) to 7 ($H_4P_2O_7$). Both variables are explained in detail in appendix A.4, p. 162.

This means that in the basis of these CV, one can find the orthophosphates at $(0, 8)$ and the pyrophosphate at $(1, 7)$. The scheme of a phosphorylation reaction (cf. section 1.1.3, p. 18) is given in fig. 5.1. three possible reactions can be distinguished: addition-activated, concerted and elimination-activated reaction. Addition-activated means, that the reaction coordinate passes through an addition complex $(1, 8)$; concerted means, no intermediate complex occurs and the reaction passes through a single barrier; the elimination-activated process requires the dissociation of H_2O from the phosphate, forming an intermediate metaphosphate PO_3^- at $(0, 7)$. Fig. 5.2c shows the FES for a two-dimensional WTmetaD simulation based on CN(P−O_{all}) and CN(P−O_{bridge}). The actually observed coordinates are shifted from this idealized integer values, due to the



(a) $2\,H_2PO_4^- \xrightarrow[-H_2O]{} PO_3^- + H_2PO_4^-$ **(b)** $PO_3^- + H_2PO_4^- \longrightarrow H_2P_2O_7^{2-}$



(c) Free energy surface, iso-lines in kJ mol^{-1}. **(d)** Number of H$^+$ in least protonated phosphate

Figure 5.2.: 2D WTmetaD simulation of two H_3PO_4 (2 P) condensation to $H_4P_2O_7$ (P_2) in a box with 124 H_2O molecules. (a) Optimized paths of H_2O elimination to form PO_3^- (metaphosphate, mP) followed by $H_2PO_4^-$ addition. (b) $H_2PO_4^-$ addition to PO_3^-. (c) The free energy surface is defined by the number of bridging oxygen CN($P-O_{bridge}$) and the overall number of phosphorus-coordinating oxygen CN($P-O_{all}$) The well in the top left corner of the graph at $(0.1, 7.6)$ represents the situation with two H_3PO_4 with a total of 8 oxygen binding to phosphorus but none bridging. In the bottom right, a well appears around $(1.0, 6.8)$ which shows a total of 7 phosphorus binding oxygen and 1 bridging as in $H_2P_2O_7^{2-}$. (d) The number of H$^+$ in the least protonated phosphate, 0 (purple), 1 (dark green), 2 (bright green), 3 (yellow).

smooth switching functions. Walls were used to avoid the escape of walkers from the vicinity of this rectangle, starting for values of $\mathrm{CN(P{-}O_{all})}<6.7$ and $\mathrm{CN(P{-}O_{bridge})}>1.2$. These values were chosen after careful analysis of (unbiased) AIMD simulations such that those are not affected by the walls.

As the FES in fig. 5.2c shows, the reaction is activated by initial elimination of water from one phosphate leading to a short-lived metaphosphate ion $\mathrm{PO_3^-}$:

$$2\,\mathrm{H_2PO_4^-} \rightleftharpoons \mathrm{PO_3^-} + \mathrm{H_2PO_4^-} + \mathrm{H_2O} \rightleftharpoons \mathrm{H_2P_2O_7^{2-}} + \mathrm{H_2O} \tag{5.2}$$

With a free energy of $95\,\mathrm{kJ\,mol^{-1}}$, the $\mathrm{PO_3^-}$ (positioned at $(0.1, 6.7)$ in fig. 5.2c) is far more likely to react back to orthophosphate ($F^\ddagger = 9\,\mathrm{kJ\,mol^{-1}}$) than to pyrophosphate ($F^\ddagger = 33\,\mathrm{kJ\,mol^{-1}}$). In fact it cannot be excluded that the small barrier of $F^\ddagger = 9\,\mathrm{kJ\,mol^{-1}}$ vanishes at longer simulation time. The activation by initial addition of water to orthophosphate was also covered at $(1.0, 7.7)$ but with a higher free energy of $119\,\mathrm{kJ\,mol^{-1}}$. Overall, $\mathrm{H_2PO_4^-}$ is $27\,\mathrm{kJ\,mol^{-1}}$ more stable than the $\mathrm{H_2P_2O_7^{2-}}$, which is consistent with experimental findings of $32\,\mathrm{kJ\,mol^{-1}}$.[46] The total barrier for condensation is $F^\ddagger = 127\,\mathrm{kJ\,mol^{-1}}$, and $100\,\mathrm{kJ\,mol^{-1}}$ for the hydrolysis. The hydrolysis barrier is $19\,\mathrm{kJ\,mol^{-1}}$ lower than the result of $G^\ddagger = 119\,\mathrm{kJ\,mol^{-1}}$ reported by Stockbridge and Wolfenden.[47] This is still an excellent agreement, since it is well known, that PBE, the used DFT functional, tends to underestimate barriers and ab initio WTmetaD simulation times are difficult to converge.

Besides P−O bond reformations, the reaction also consists of proton transfer reactions in order to form the activated metaphosphate ion $\mathrm{PO_3^-}$. The protonation states cannot be retrieved from the FES directly, but by evaluation of the biased trajectory. Fig. 5.2d is a scatter plot where each dot refers to the protonation of the *less* protonated phosphate center. This plot clearly shows the most common protonation states in each area of the FES: The orthophosphoric acid is clearly deprotonated once to $\mathrm{H_2PO_4^-}$ and the metaphosphate is PO_3^-; the pyrophosphate is an interesting case, since the protonation varies between 0 and 1, which is mostly an intramolecular protonation:

$$[HO_3P{-}O{-}PO_3H]^{2-} \rightleftharpoons [O_3P{-}O{-}PO_3H_2]^{2-} \tag{5.3}$$

The right-side configuration forms after addition of $H_2PO_4^-$ to PO_3^-, but it is surprisingly common in the simulation. In fig. 5.2d this is visible as overlapping dark green and purple points reflecting protonations of 0 an 1 in one of the phosphate groups The overall protonation of the pyrophosphate ion can be identified in a different scatter plot (not shown) that estimates the protonation of both phosphate centers, which is in nearly all points 2 and validates eq. 5.3. This is also the case for the transition state at the highest barrier around $(0.4, 6.7)$, which validates eq. 5.2 The deprotonation of the orthophosphate occurs in a concerted mechanism with water elimination which is visible in fig. 5.2d as a sharp transition from two protons (bright green dots) to zero (purple dots) at $\mathrm{CN}(\mathrm{P{-}O_{bridge}} \approx 6.8)$.

The simulation reveals a high level of complexity due to the protonation transfer reactions. Those reactions would, for further exploration, involve dedicated CV for different protons. Moreover, the proton transfer in aqueous solution most likely involves one or several H_2O. Prasad et al. indicated that the number of H_2O may vary in unexpected ways which are difficult to explore.[51]

5.2. Orthophosphate adsorption in presence of water

The γ-Al_2O_3 (1 1 0) model facet comprises dentation sites for two sites to form bidentates. Following the labels introduced in fig. 3.1a on page 53, they can be located as neighboring oxygen atoms at sites *de* and *bb'* (or *bc* on R(1 1 0)). The *de* sites bind to tetrahedral and octahedral aluminium atoms (labelled as α and β in fig. 3.1ii and iii, p. 53), whereas *bb'* are sites from two symmetry-equivalent octahedral aluminium atoms (labelled with γ and γ'). The *de* adsorption reaction is much more complicated, due to the tetrahedral aluminium binding site α, potentially involving the adsorption of one or two additional H_2O molecules in an intermediate octahedral complex around the aluminium atom α, similar to those shown in [96]. With these aspects, the formation of the ν_2^{de} bidentate would require a difficult and time-consuming simulation. The *bb'* site, on the other hand, is limited to a much simpler elimination-addition process as it only involves two symmetry-equivalent sites. In this chapter, we focus on the adsorption on the *bb'* site.

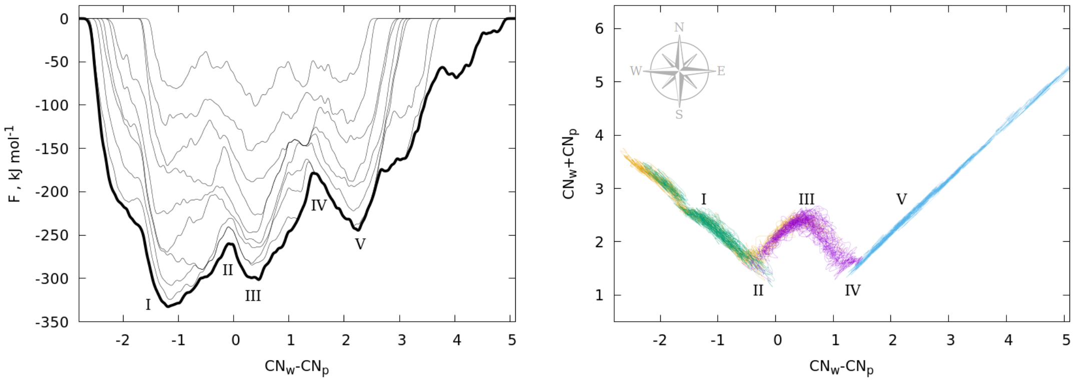
The relevant collective variables for the reaction are CN_w, the coordination number of the two relevant aluminium atoms (with γ-label in 3.1ii on page 53) with water oxygen ($\mathrm{Al}^\gamma{-}\mathrm{O}_w^b$), and CN_p, number of phosphate oxygen atoms binding to the same site b ($\mathrm{Al}^\gamma{-}\mathrm{O}_p^b$). Since the H_3PO_4 adsorption requires the initial desorption of H_2O from b, there is only a single possible pathway to form the monodentate ν_1^b. This monodentate can, again, only form the bidentate $\nu_2^{bb'}$ after desorption of the second H_2O from site b'. Both, the reactions $\nu_0 \longrightarrow \nu_1^b$ and $\nu_1^b \longrightarrow \nu_2^{bb'}$ decrease CN_w by 1 and increase CN_p by 1. Knowing this, CN_w and CN_p can be combined to a single CV:

$$\mathrm{CV} = \mathrm{CN}_w - \mathrm{CN}_p \tag{5.4}$$

For this CV, bidentate, monodentate and physisorption occur at well distinguished values:

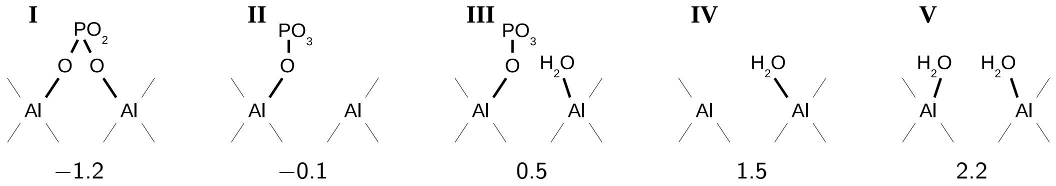
$$\nu_2^{bb'} \longrightarrow \mathrm{CV} = -2\,, \qquad \nu_1^b \longrightarrow \mathrm{CV} = 0 \quad \text{and} \quad \nu_0 \longrightarrow \mathrm{CV} = 2$$

Fig. 5.3 shows the results of a one-dimensional well-tempered metadynamics simulation. At the end of the simulation, the bidentate ($\nu_2^{bb'}$, structure I in fig. 5.3) is clearly the most stable state followed by monodentate (ν_1^b, structure III) and physisorption (ν_0, structure V). The free energies of ν_1^b and $\nu_2^{bb'}$ formation are, -58 and $-88\,\mathrm{kJ\,mol^{-1}}$, respectively, relative to the physisorbed state ν_0. The first and second adsorption barrier



(a) Free energy surface

(b) Visited area for independent CN_w and CN_p



(c) Local geometry at minima and transition states. Protonation state of phosphate neglected in this illustration. The determined of $CN_w - CN_p$ for each geometry are given below each structure. They are shift relative to their ideal values of, from left to right, -2, -1, 0, 1 and 2.

Figure 5.3.: (a) 1D WTmetaD simulation of phosphate adsorption to γ-Al_2O_3 by H_2O substitution at the (1 1 0)*bb'* sites. The CV is the difference of coordination numbers of Al−O with phosphate oxygen and water oxygen: CN_w−CN_p (a) Consecutive FES in steps of 1000 gaussian hills each (hair lines), the last FES is highlighted as thick line. (b) Projection of trajectories on CN_w+CN_p. This reveals the progression of the simulation along CN_w and CN_p (equivalent to diagonals, north-east: CN_w, north-west: CN_p). Four different colors indicate each walker (section 2.2.5), p. 48). (c) Structures at minima and transition states. Both aluminium atoms are octahedral sites. Gaussian hills: $\sigma = 0.04, W_0 = 3\,\mathrm{kJ\,mol^{-1}}$; deposition rate: 100 ps^{-1}.

are 66 kJ mol^{-1} (structure IV in fig. 5.3) and 41 kJ mol^{-1} (structure II) respectively. Thus the adsorption is clearly exergonic which indicates that phosphate rapidly and irreversibly binds to the surface as $\nu_2^{bb'}$. This is interesting since the higher dentations like ν_2 under drying conditions are only stabilized due to the H_2O vapor entropy. Transferred to the solid-liquid interface, this vapor entropy is substituted for a H_2O liquid entropy that contributes the overall FES. At drying conditions the free enthalpy difference between ν_1^b and $\nu_2^{bb'}$ is $32\,\mathrm{kJ\,mol^{-1}}$ (cf. tab. A.1, p. 154), which is almost equal to the to the solid-liquid interface ($30\,\mathrm{kJ\,mol^{-1}}$). While it is possible that $\nu_2^{bb'}$ is also the most stable adsorption mode at the solid-liquid interface it is unexpected to observe it with the same free energy as for drying conditions. One likely reason here is that the WTmetaD simulation is not ideally converged. Nevertheless, the same trend, i. e. stability order is clearly visible for the last 10 ps of the simulation, hence at least the stability is expected to be correct.

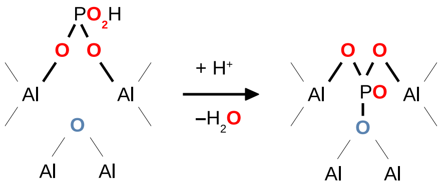
In order to verify the combined CV in eq. 5.4, the positions of the minima are studied in more detail. The minima are not exactly at 2, 0 and -2 because CN_w and CN_p are defined by smooth functions and surrounding water and phosphate oxygen atoms still produce subtle contributions at distances around 4Å, which would definitely not be considered coordination. Note that these smooth functions are necessary in order to generate the forces that pull the to-be-bonded oxygen atoms to its site. It is possible to achieve numeric values close to 2, 0 and -2 by using stiff functions for the coordination, but this is at the expense of a less effective metadynamics sampling, since the forces vanish too quickly around the cutoff distance (cf. section 2.2.4, p. 46) Fig. 5.3c illustrates the adsorption modes and transition states, and provides the actual values at which the adsorption modes occurred. The transition states are expected at $\mathrm{CV}{=}{-1}$ and $\mathrm{CV}{=}1$, that are the positions where H_2O is desorbed, but the phosphate did not yet adsorb to the available coordination site. Consequently, the aluminium atom (label γ in fig. 3.1ii) is pentacoordinated in the transitions states. Fig. 5.3b plots the trajectories projected on $\mathrm{CN}_w{-}\mathrm{CN}_p$ and $\mathrm{CN}_w{+}\mathrm{CN}_p$, which is equivalent to plotting CN_w against CN_p and rotating the plot by 45°. Along this projection the trajectories generate a W-shaped, rectilinear area, in which the edges represent the three adsorption modes connected by two transition state. While the positions of the minima and transition state positions are fluctuating during the simulations (cf. hair lines in fig. 5.3a), the edges of the W-shape in fig. 5.3b are accurately spaced with distances of ΔCV=2 between the adsorption modes. The overall positive shift observed for $\mathrm{CN}_w - \mathrm{CN}_p$ and $\mathrm{CN}_w + \mathrm{CN}_p$ is due to the larger number H_2O contributing to CN_w compared to four oxygen atoms in H_3PO_4. Overall, the one-dimensional WTmetaD successfully sampled the adsorption sequence from physisorption to bidentate.

Compared to these barriers, a large free energy landscape of more than $300\,\mathrm{kJ\,mol^{-1}}$ was carved out throughout the simulation, which is visible at the ordinate of fig. 5.3a. Hence towards the end of it, less probable states like additional hydration barriers of aluminium atoms ($CV > 3$, see fig. 5.3) can be overcome. Besides parallelization, the use of multiple walkers serves the purpose of minimizing oscillations during the exploration of the free energy landscape. Nevertheless, these walkers tend to diverge and explore separate areas, given to the fact that they generate a potential that repels itself and other walkers. In this simulation, four walkers simultaneously explored the FES projection along one CV. In particular for the exploration of the more stable adsorption modes and barrier (structures I-IV) are sampled by three of the walkers, in fig. 5.3a green, purple and yellow, whereas physisorption (structure V) is only sampled by one walker (blue), which also contributes to the transition state (structure IV). At some point, the blue walker starts to determine a new local minimum at $CV \approx 4$, which can occur, given that relative to structure V (ν_0) a potential of $250\,\mathrm{kJ\,mol^{-1}}$ forces this walker to proceed along the CV towards in this case irrelevant areas. In the meantime of the simulation, the remaining ones continue to improve the minima for ν_2 and ν_1 (I and III in fig. 5.3). This problem is intrinsic standard metadynamics (metaD) but also WTmetaD only in parts overcomes this problem. In particular for the overall short simulation times at DFT level, it can introduce artefacts that require excessively long simulation times to “smoothen”. In order to avoid these irrelevant areas of the free energy surface from the beginning of the simulation, additional wall biases can be introduced. These additional biases keep the walkers from diverging into areas of “irrelevant” chemistry by effectively limiting the area covered within the metadynamics simulation. For the other simulations presented in the following sections, such walls were introduced. However, in the region of interest, this does not change the overall conclusion that the adsorption process is exergonic and readily occurs at room temperature. While one can expect a similar trend for the much more complex ν_2^{de} bidentate formation involving two different adsorption sites, one being the tetrahedral aluminium atom α (cf. fig. 3.1ii, p. 53), the result may not be exactly the same. According to new geochemical models based on *in situ* ATR-IR and zetametry,[125] a significant bonding mode of phosphoric acid at the solid-liquid interface is a monodentate binding to tetrahedral aluminium sites. This in general means that adsorption is also favorable on tetrahedral alumina, even though possibly as monodentate. Nevertheless the results in this sections indicate, that the monomer ν_1^d is in a fast equilibrium with the bidentate ν_2^{de}.

5.3. Phosphorylation barrier of the γ-Al_2O_3 surface

Monodentates and most bidentates are readily accessible via substitution of μ_1-OH hydroxy groups from preadsorbed water, as shown in section 5.2. While some of these structures were rather stable, the most stable candidates in the dry state are tridentates and a bidentate, where an alumina oxygen is substituted. However, these alumina oxygen are μ_2-OH, μ_3-OH or μ_3-O, which cannot be substituted directly, since they do not easily desorb from the surface. Thus, it is important to determine how and when they can form. Here, we test the phosphorylation reaction of those sites, which could be similar to the reaction:

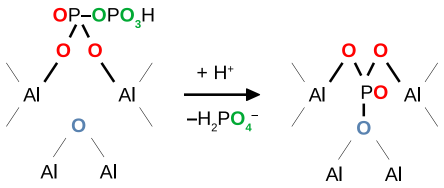
$$\nu_2 + \mu_2\text{-O} \longrightarrow \nu_3 + H_2O \quad :=$$



(5.5)

The phosphate oxygen atoms are highlighted in red and the alumina oxygen in blue. The H_2O leaving the surface was formed from a P–OH moiety. This is similar to the pyrophosphate condensation, which is a phosphorylation between two orthophosphates. It completes the reaction sequence as an exploration of reactions up to ν_3 in eq. 5.1. Since pyrophosphates are also present at the surface, this reaction could also start from an adsorbed pyrophosphate and form an orthophosphate as leaving group instead of water:

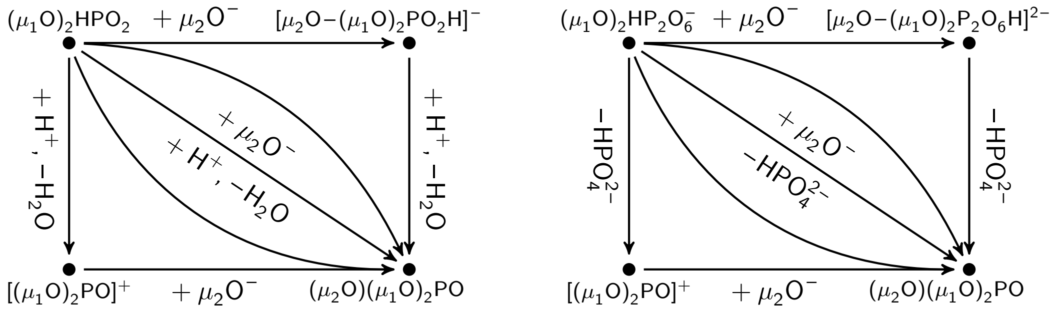
$$\nu_{20} + \mu_2\text{-O} \longrightarrow \nu_3 + \nu_0 \quad := \tag{5.6}$$



The same coloring of oxygen applies here with additional green labels for phosphate oxygen of the leaving orthophosphate. In this reaction, the alumina phosphorylation reaction occurs together with pyrophosphate hydrolysis, however, the only change compared to eq. 5.5 is the leaving group of $H_2PO_4^-$. At least for the simulations, this leaving group cannot diffuse into bulk H_2O, hence it is described as physisorption (ν_0). The presence of pyrophosphate was discussed in detail in chapter 3 and 4, with a proposed rationale for ^{31}P NMR[12] and direct ^{31}P NMR measurements[18] at the dry γ-Al_2O_3 surface. The possible paths for both phosphorylation reactions are illustrated in fig.5.4. It should be noted, that these two reactions are simplified, as they only show one possible scenario for proton transfer reactions. In general, the concerned proton transfer reactions are not known and have to be included in several cases as well. The overall considerations of phosphorylation reactions were explained in more detail in section 1.1.3 (p. 18) In the following sections, these reactions were explored by means of WTmetaD with focus on the P–O bond cleavage and formation. At the dry surfaces, the most stable single orthophosphate adsorption structure is identified as R(1 1 0)ν_3^{bcB} with labels given in fig. 3.1iii (p. 53). The formation of this species involves the alumina oxygen site μ_2^B-OH, that needs to be phosphorylated in order to form this adsorption structure. Similarly on the unreconstructed facet, the most stable structure is n(1 1 0)ν_3^{deA}, requiring the phosphorylation of μ_3^A-O (see fig. 3.1ii). Another special case is R(1 1 0)ν_2^{eB}, which is the second most stable adsorption structure identified. As opposed to n(1 1 0)$\nu_2^{bb'}$, which was discussed to describe the chemisorption from ν_0 to ν_2 in the last section, R(1 1 0)ν_2^{eB} involves the same μ_2^B-OH site as in ν_3^{bcB}. It is therefore more similar to the tridentates, which cannot form by water substitution. Reactions to form any of these adsorption structures can be explored by adapting the reaction in eq. 5.5 for the specific site. In case of n(1 1 0)ν_3^{deA} the μ_2-O is substituted by μ_3^A-O whereas it is substituted by μ_2^B-OH for R(1 1 0)ν_3^{bcB} and ν_2^{eB}. For phosphorylation of μ_2-OH it was thus also necessary to add a CV for deprotonation of this site. Furthermore, the starting structures can be specified: ν_2^{de} for ν_3^{deA}, ν_2^{bc} for ν_3^{bcB} and ν_1^{e} for ν_2^{eB}.

5.3.1. The collective variables and bias

Three distinct sites were chosen to carry out WTmetaD simulations of the surface phosphorylation, using two different reactions (eq. 5.5 and eq. 5.6). The principle is the same in all cases: the collective variables (CV) for the surface phosphorylations in fig. 5.4 are established as coordination (cf. section 2.2.4, p. 46). Oxygen atoms in the systems are grouped in order to create the coordination CV CN_a and CN_w which each follow the bond formation with the specific alumina oxygen (μ_3^A-O or μ_3^B-OH) and bond cleavage with H_2O oxygen. The coordination CN_a quantifies the number of bonds to the alumina oxygen site that is phosphorylated; since it is only one site, CN_a can only vary between 0 and 1. CN_w counts the number of free P−O bonds and all oxygen atoms in H_2O. The values of CN_w depend on the free P−O bonds, which are 2 and 1 for ν_2 and ν_3 structures (also illustrated in eq. 5.5 and fig. 5.4a). For the formation of ν_2^{eB} from ν_1^e the value varies from 3 to 2. In the case of the pyrophosphate reaction (eq. 5.6) CN_w is replaced by the P−O coordination CN_p with the four oxygen atoms from the leaving phosphate group; in all cases CN_p varies between 0 and 1, since pyrophosphate



(a) $\nu_2 + \mu_2O \longrightarrow \nu_3 + H_2O$

(b) $\nu_{20} + \mu_2O \longrightarrow \nu_3 + \nu_0$

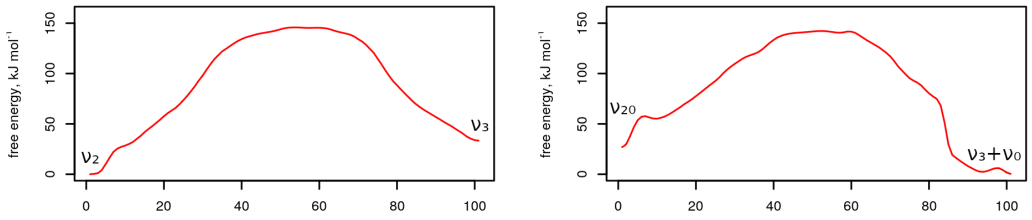
Figure 5.4.: γ-Al_2O_3 surface phosphorylation reaction schematics for the reactions according to (a) eq. 5.5, with the formation of H_2O as leaving group, and (b) eq. 5.6, with instead the formation of a phosphate ion at the surface. Concerning proton transfers and charges, the figures are simplified by only involving μ_xO sites with $x = 1$ for water substitution sites (cf. section 5.2) and $x = 2$ for alumina oxygen, each assumed to carry one negative formal charge. The released HPO_4^{2-} can be protonated to $H_2PO_4^-$ but it is assumed that it stays close to the interface. The desorption into bulk H_2O is neglected.

is characterized by a single P−O−P bridge, which is hydrolyzed during the reaction. Regardless of the different integer values of the different coordinations, their range is always $\Delta CN = 1$ and equivalent WTmetaD simulations can be carried out to explore the FES of those phosphorylations. As before, the minima of the resulting FES are not exactly at the described integer values, due to the smooth switching functions employed to define a P−O coordination. The bias is applied by depositing gaussians in the visited FES regions, thereby "flooding" minima and automatically exploring the FES area of interest. Here, a gaussian height of $3\,\mathrm{kJ\,mol^{-1}}$ was used, and for CN_a, CN_w and CN_p gaussian widths of 0.04. In section 5.3.3 an additional CV was required for deprotonation of μ_2^B-OH. The additional dimension was included with a gaussian width of 0.15, and the gaussian height increased to 5 kJ mol^{-1}. Tab. 5.1 provides an overview of the WTmetaD simulation times and the evaluated reaction free energies and free energy barriers.

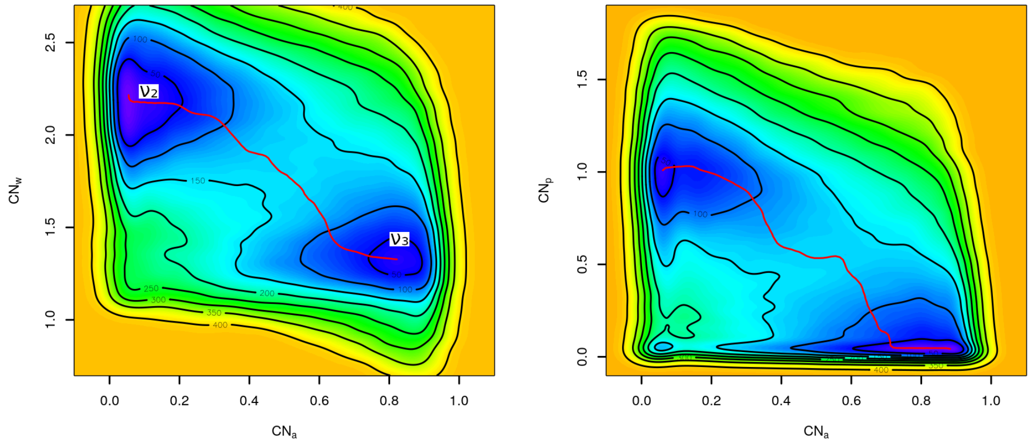
5.3.2. Tridentate ν_3^{deA} on n(1 1 0)

The results of two WTmetaD simulations for the formation of n(1 1 0)ν_3^{deA} are given in fig. 5.5. Both simulations follow the schemes in fig. 5.4, starting either from a preadsorbed orthophosphate ν_2^{de} or pyrophosphate ν_{20}^{de}. The lowest reaction path is a concerted reaction regardless of the leaving group. This conclusion is not only based on the fact that the reaction path is approximately described as diagonal through the two graphs but also because it does not include intermediates in contrast to the phosphate condensation. It is clearly visible in fig. 5.5b that the phosphorylation starting from orthophosphate is endergonic (33 kJ mol^{-1} cf. tab. 5.1), while the reaction with pyrophosphate in fig. 5.5d is exergonic (-27 kJ mol^{-1}). The free energy barriers are 146 kJ mol^{-1} for the orthophosphate and 115 kJ mol^{-1} for pyrophosphate. Overall the barrier is decreased when considering pyrophosphate but it is still slightly higher than the estimate for phosphate hydrolysis (100 kJ mol^{-1}). The relevant comparison is the hydrolysis, since the pyrophosphate is split into orthophosphates (while phosphorylation, thus no H_2O is formed here).

Regarding the elimination-activated path observed for phosphate condensation in bulk liquid, it is not expected to occur based on the high free energy region at $(0.1, 1.3)$ in coordinates of CN_a and CN_w of fig. 5.5b. The free energy in the same region in fig. 5.5d seems to be lower, but one has to keep in mind that, the phosphorylation reaction with a phosphate leaving group is in competition with the analogous reaction of leaving water. While sampling CN_a, thereby cleaving the bond without immediate addition of phosphate, would lead to the elimination state around $(0.1, 0.0)$ in coordinates of CN_a and CN_p. This reacts spontaneously with water and even in another alumina oxygen site, thereby escaping the relevant FES region, i. e. the state at $(0.1, 0.0)$ is usually not a metaphosphate



(a) $\nu_2^{de} + \mu_3^A\text{-O} \rightleftharpoons \nu_3^{deA} + \mathrm{H_2O}$ **(c)** $\nu_{20}^{de} + \mu_3^A\text{-O} \rightleftharpoons \nu_3^{deA} + \nu_0$



(b) FES of (a), iso-lines in $\mathrm{kJ\,mol^{-1}}$ **(d)** FES of (b), iso-lines in $\mathrm{kJ\,mol^{-1}}$

Figure 5.5.: Two 2D WTmetaD simulations for formation of ν_3^{deA} on n(1 1 0). Bias along coordination number of P−O$_{\text{alumina}}$ (CN$_a$), P−O$_{\text{water}}$ (CN$_w$) and P−O$_{\text{phosphate}}$ (CN$_p$). Iso-lines in free energy surface in $\mathrm{kJ\,mol^{-1}}$. (a): Minimum energy path of the phosphorylation reaction of alumina oxygen “A” (cf. fig. 3.1ii, p. 53), starting from the bidentate ν_2^{de} with $\mathrm{H_2O}$ as leaving group. (b): Free energy surface of (a). (c): Minimum energy path of the same phosphorylation with a ν_{20} pyrophosphate, such that the leaving group is orthophosphate. (d): Free energy surface of (c). See eq. 5.5 and 5.6 on page 102 for the reaction mechanism.

structure like $[(\mu_1\mathrm{O})_2\mathrm{PO})]^+$ (cf. fig. 5.4b). It is therefore not possible to make reliable statements about the elimination path, however, the concerted reaction seems to be favored by the system. Additional wall biases were included that artificially disfavor water adsorption in order to circumvent this problem, however these wall biases did not entirely suppress the competing reactions. In consequence, the all biases keep the walkers of the simulation ”on track” and allow to sample the desired area but the elimination-activated

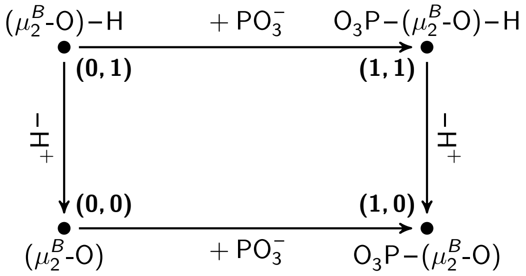


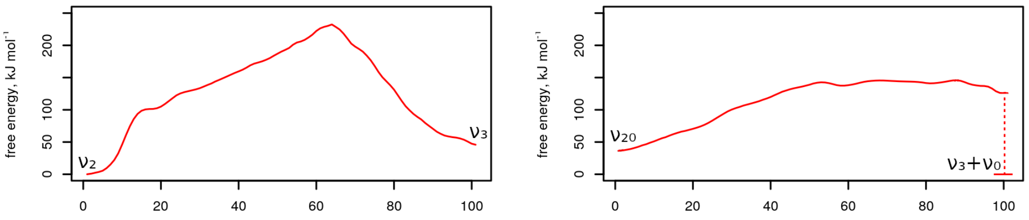
Figure 5.6.: The local coordination of μ_2^B was biased with two CV, CN_a (x-axis) and CN_h (y-axis), in order to describe the deprotonation and phosphorylation of this site. The coordinates (x, y) correspond to CN_a and CN_h.

mechanism for pyrophosphate cannot be sampled reliably. It must be emphasized that this problem is intrinsic to the tested reaction for this system and independent of the added walls. This concerns all surface reactions involving pyrophosphate when using the chosen collective variables.

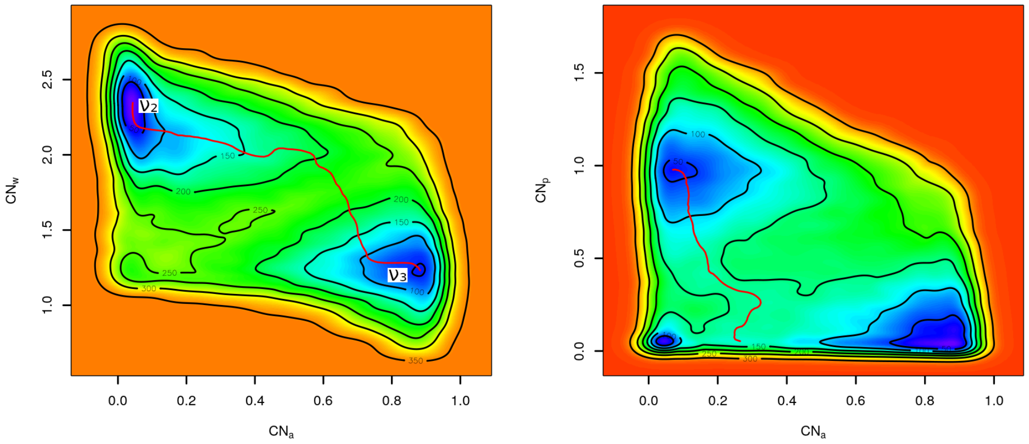
5.3.3. Bidentate ν_2^{eB} and tridentate ν_3^{bcB} on R(1 1 0)

The two most stable adsorption sites found at drying conditions are R(1 1 0)ν_3^{bcB} and R(1 1 0)ν_2^{eB}. They involve a different alumina site, the hydroxide μ_2^B-OH, which is always protonated at the water interface. This is in contrast to μ_3^A from n(1 1 0)ν_3^{deA} discussed in section 5.3.2; this site can be protonated as μ_3^A-OH but it is not required to be. Unbiased AIMD simulations, which consist of 2×2 unit cells with symmetry-equivalent pairs of A/A', in total 8 sites representing the behavior of A, showed that some sites are protonated but there are always several unprotonated sites. Here, the protonation of μ_3^A was neglected, assuming that some of these sites are always deprotonated. The alumina oxygen site μ_2^B is however always protonated – it rapidly recovers this protonation in AIMD simulations, if a starting geometry involving μ_2^B-O was chosen.

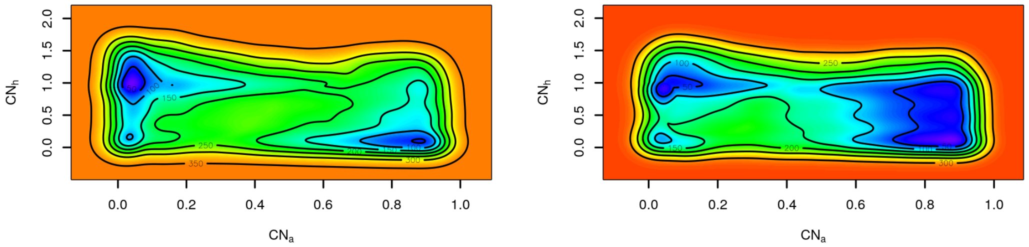
This requires a deprotonation during the reaction in order to form μ_2^B-O$-$PO$_3$. It was included as a third CV, CN_h which is used to bias the μ_2^B-O$-$H coordination. The WTmetaD simulation is thus three-dimensional which may lead to considerably slower convergence. The three-dimensional FES cannot be plotted directly, thus the extra CN_h



(a) $\nu_2^{bc} + \mu_2^B\text{-OH} \rightleftharpoons \nu_3^{bcB} + \mathrm{H_2O}$ **(d)** $\nu_{20}^{bc} + \mu_2^B\text{-OH} \rightleftharpoons \nu_3^{bcB} + \nu_0$



(b) projected FES of (a), iso-lines in kJ mol^{-1} **(e)** projected FES of (d), iso-lines in kJ mol^{-1}

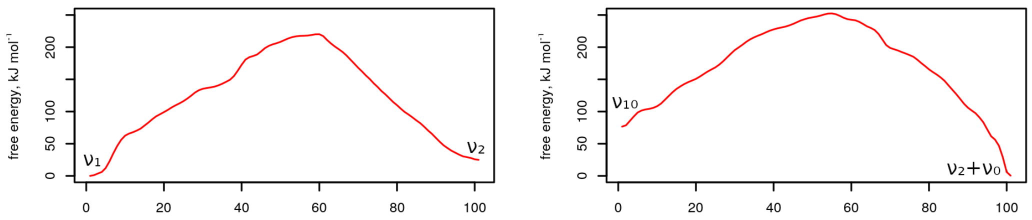


(c) projected FES of (a) for proton transfer **(f)** projected FES of (d) for proton transfer

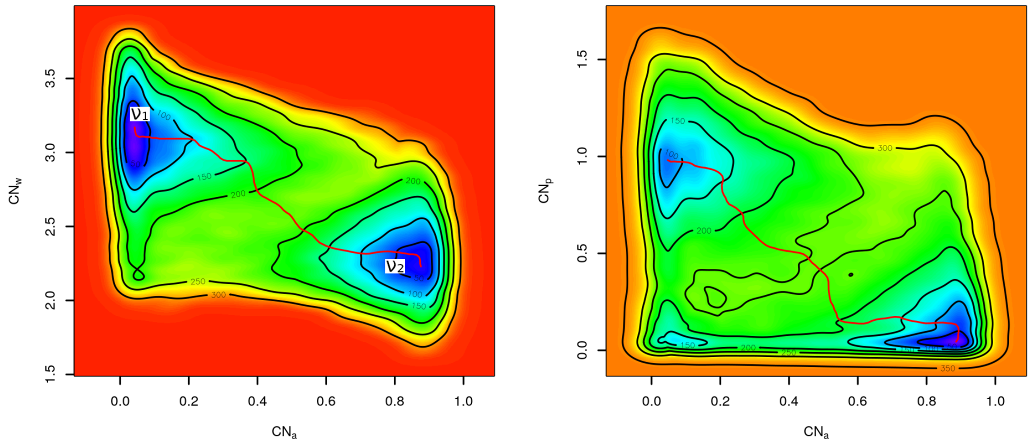
Figure 5.7.: Two 3D WTmetaD simulations as candidate reactions to form ν_3^{bcB} on R(1 1 0). Bias along coordination number of $\mathrm{P{-}O_{alumina}}$ (CN_a), $\mathrm{P{-}O_{water}}$ (CN_w), $\mathrm{P{-}O_{phosphate}}$ (CN_p) and μ_2^{B}-O−H (CN_h). Iso-lines in free energy surface in kJ mol^{-1}. (a): Minimum energy path of the phosphorylation of alumina oxygen “B” (cf. fig. 3.1ii, p. 53), starting from the bidentate ν_2^{bc} with $\mathrm{H_2O}$ as leaving group. (b): Free energy surface of (a). (c): Free energy surface of proton transfer in (a). (d): Minimum energy path of the same phosphorylation with a ν_{20} pyrophosphate, such that the leaving group is orthophosphate. Path incomplete in order to include correct transition state (software limitation). (e): Free energy surface of (d). (f): Free energy surface of proton transfer in (d). See eq. 5.5 and 5.6 on page 102 for the reaction mechanism.

dimension was integrated out. The result is a two-dimensional plot of CN_a and CN_w which is shown in fig. 5.7b. Even though CN_h was biased, it is here treated like one (of many) hidden variables. Fig. 5.7b is equivalent to the phosphorylation scheme in fig. 5.4a. From this it is immediately visible, that this reaction is concerted like the phosphorylation of n(1 1 0)μ_3^A for the formation of ν_3^{deA}. Moreover, the third dimension provides information about the protonation state along the reaction path. For this a different two-dimensional FES was extracted from the simulation, this time by integrating out CN_w. That way, CN_h and CN_a are kept, which describes the local coordination of the μ_2^B site. The scheme for the different coordination states is given in fig. 5.6. While the reaction path in the phosphorylation FES in fig. 5.7b roughly describes a diagonal, indicating a concerted phosphorylation, reaction path in fig. 5.7c goes through the corners. That means, the deprotonation occurs independently from the phosphorylation. The reaction is possible by initial deprotonation as well as deprotonation after phosphorylation, where the latter is preferred, since the lowest free energy path leads through point $(1, 1)$ in fig. 5.7c, which corresponds to $\mathrm{O_3P{-}}(\mu_2^B\text{-O}){-}\mathrm{H}$. In the phosphorylation starting from pyrophosphate, shown in fig. 5.7e and fig. 5.7f, the same trends are observed and the tendency to deprotonate after phosphorylation is even more pronounced. This is most likely due to the more stable orthophosphate leaving group, which provides the necessary free energy to bind to μ_2^B-OH directly. The same trend is observed for the phosphorylation reaction to form ν_2^{eB} from the monodentated orthophosphate ν_1^e and the pyrophosphate ν_{10}^e, shown in fig. 5.8

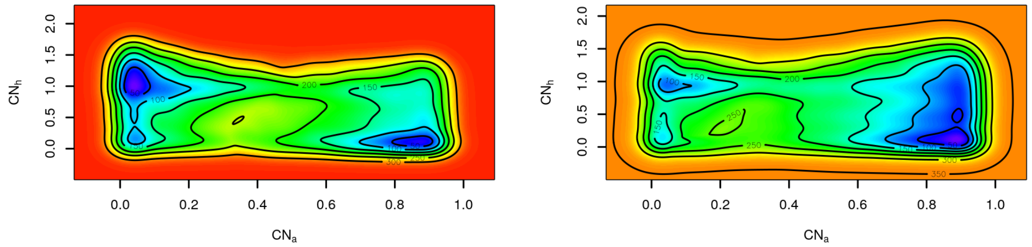
As summarized in tab. 5.1, the reaction free energies also follow the same trend in all four cases of formation of ν_2^{eB} and ν_3^{bcB} from orthophosphate and pyrophosphate. The reactions of orthophosphate in fig. 5.7b and fig. 5.8b are endergonic and have a very high barrier of $232\,\mathrm{kJ\,mol^{-1}}$ and $220\,\mathrm{kJ\,mol^{-1}}$. The similar reactions of pyrophosphate in fig. 5.7e and fig. 5.8e are exergonic and have lower barriers of $109\,\mathrm{kJ\,mol^{-1}}$ and $176\,\mathrm{kJ\,mol^{-1}}$. In particular the value of $109\,\mathrm{kJ\,mol^{-1}}$ is surprisingly low, however, here



(a) $\nu_1^e + \mu_2^B\text{-OH} \rightleftharpoons \nu_2^{eB} + \mathrm{H_2O}$ **(d)** $\nu_{10}^e + \mu_2^B\text{-OH} \rightleftharpoons \nu_2^{eB} + \nu_0$



(b) projected FES of (a), iso-lines in kJ mol^{-1} **(e)** projected FES of (d), iso-lines in kJ mol^{-1}



(c) projected FES of (a) for proton transfer **(f)** projected FES of (d) for proton transfer

Figure 5.8.: Two 3D WTmetaD simulations as candidate reactions to form ν_2^{eB} on R(1 1 0). Bias along coordination number of P−O$_{\text{alumina}}$ (CN$_a$), P−O$_{\text{water}}$ (CN$_w$), P−O$_{\text{phosphate}}$ (CN$_p$) and μ_2^{B}-O−H (CN$_h$). Iso-lines in free energy surface in kJ mol^{-1}. (a): Minimum energy path of the phosphorylation of alumina oxygen “B” (cf. fig. 3.1ii, p. 53), starting from the bidentate ν_2^e with $\mathrm{H_2O}$ as leaving group. (b): Free energy surface of (a). (c): Free energy surface of proton transfer in (a). (d): Minimum energy path of the same phosphorylation with a ν_{20} pyrophosphate, such that the leaving group is orthophosphate. (e): Free energy surface of (d). (f): Free energy surface of proton transfer in (d).

it has to be considered that this WTmetaD simulation was much shorter than the others. As a consequence of the third CV, CN_h, the overall convergence time of these simulations increases by an order of magnitude. Moreover, the carried out simulations were shorter than the simulations for $\mathrm{n}(1\,1\,0)\nu_3^{deA}$ formation (cf. tab. 5.1). Consequently, the convergence of the presented FES differs from one another. The big differences in the barriers indicate that those are strongly affected by this, whereas the reaction free energies are consistent. This is actually expected, since the bias potential generated WTmetaD does not fully compensate for the minima in the FES ("flat target"), but converges to a "well-tempered target". It is a target which depends on the bias factor and allows sampling of the FES by flattening it up to the point where thermal overcomes the barrier (see eq. 2.11, p. 49 for details). As consequence, minima are still better sampled than transition states, because they are still visited more often throughout the simulation. With the limited simulation time, this shows in much less accurate barriers than reaction energies.

Regarding the apparent preference of $\nu_{10}^{e} \rightarrow \nu_2^{eB} + \nu_0$ (fig. 5.8e) for an elimination-activated reaction: this is rather likely in this case, but for the same reason as in the previous section, the minimum at $(0, 0)$ cannot be reliably sampled due to competing reactions Therefore, this possibility was not evaluated.

5.3.4. Summary and perspectives

In this chapter, well-tempered metadynamics (WTmetaD) simulations were used to explore the formation of several phosphate surface species on γ-$\mathrm{Al_2O_3}(1\,1\,0)$ that were identified as the most stable model adsorption structures at drying conditions: $\mathrm{R}(1\,1\,0)\nu_3^{bcB}$, $\mathrm{R}(1\,1\,0)\nu_2^{eB}$ and $\mathrm{n}(1\,1\,0)\nu_3^{deA}$ (cf. fig. 3.1, p. 53). The idea was a comprehensive study of surface phosphorylation reactions targeting the aforementioned most stable of orthophosphate species. The precursor of these reactions are orthophosphates at the next lower

Table 5.1.: WTmetaD results of the six conducted explorations for γ-Al_2O_3 surface phosphorylation: Reaction free energies and free energy barriers by the end of the simulations and the number of included CV and simulation size in hills are given. The leaving groups indicate, which of eq. 5.5 and eq. 5.6 are used. HPO_4^{2-} is considered physisorbed (ν_0). The hills implicitly include a simulation time, since 100 hills are added per picosecond, however, this time is the sum of four walkers producing 4 biased trajectories in parallel.

	$n(1\,1\,0)\nu_3^{deA}$		$R(1\,1\,0)\nu_3^{bcB}$		$R(1\,1\,0)\nu_2^{eB}$	
leaving group :	H_2O	HPO_4^{2-}	H_2O	HPO_4^{2-}	H_2O	HPO_4^{2-}
ΔF_r, $\mathrm{kJ\,mol^{-1}}$:	33	-27	46	-37	25	-77
$\Delta F^{\ddagger}$, $\mathrm{kJ\,mol^{-1}}$:	146	115	232	109	220	176
number of CV :	2	2	3	3	3	3
hills $\times$ 1000 :	29	27	19	6	19	18

dentation, only involving ν_1 adsorption sites which are accessible by water substitution, $R(1\,1\,0)\nu_2^{bc}$, $R(1\,1\,0)\nu_1^{e}$ and $n(1\,1\,0)\nu_2^{de}$, and the related pyrophosphate species, $R(1\,1\,0)\nu_{20}^{bc}$, $R(1\,1\,0)\nu_{10}^{e}$ and $n(1\,1\,0)\nu_{20}^{de}$. As demonstrated in section 5.2, these types of chemisorptions in form of bidentates (ν_2) and monodentates (ν_1) are the predominant species. For benchmarking the method and as a reference for comparisons, the condensation of two orthophosphates to pyrophosphate was simulated in bulk water. The focus in all cases was also on the activation mechanism of the reaction, simultaneously covering elimination-activated, concerted and addition-activated mechanisms. Based on the results of seven simulations, the addition-activated mechanism can be excluded, whereas the elimination-activated path was only observed in bulk water. The concerted reaction is observed as the only mechanism of surface phosphorylation, however, convergence of the barriers was only satisfactory in the two simulations on $n(1\,1\,0)$: $\nu_2^{de} \rightarrow \nu_3^{deA} + H_2O$ and $\nu_{20}^{de} \rightarrow \nu_3^{deA} + \nu_0$. In the four simulations on $R(1\,1\,0)$ the barriers are much higher, but this is most likely due to inaccurate sampling of those more demanding three-dimensional

WTmetaD simulations. The reaction free energy is not affected by this and shows the same trend as for ν_3^{deA}: the surface phosphorylation using an orthophosphate precursor is endergonic, whereas the phosphorylation using pyrophosphate is exergonic. The barrier for the pyrophosphate-based phosphorylation is with $115\,\mathrm{kJ\,mol^{-1}}$ similar the related pyrophosphate hydrolysis in bulk water ($100\,\mathrm{kJ\,mol^{-1}}$), which means that surface phosphorylation may already occur slowly during impregnation, provided that pyrophosphate or other polyphosphates are present in solution. However, it is overall unlikely to be a driving factor of surface speciation during impregnation at room temperature, which confirms that the chemisorption described in section 5.2 indeed leads to the most relevant species during impregnation: bidentates and monodentates. Even though the three-dimensional WTmetaD simulations on R(1 1 0) (fig. 5.7 and fig. fig:reB) are not accurately converged, they show trends regarding the site deprotonation of μ_2^B-OH: the deprotonation typically occurs after phosphorylation, which means the deprotonation does not occur readily by itself.

While the results in this chapter certainly do not directly support the presence of high dentations involving the phosphorylation of alumina oxygen sites, they are not excluded either. For example in Wischert's reconstruction,[92] a water adsorption site is formally removed, and the R(1 1 0)μ_2^B becomes more reactive (much more basic), than n(1 1 0)μ_3^B (cf. fig. 3.1a, p. 53). More elaborate reconstructions may effectively allow the exchange of alumina oxygen sites at a lower barrier than phosphorylation, but such reactions are unknown and difficult to explore. As it seems, the surface species may change considerably during the drying step. AIMD simulations of those surfaces could be used to explore formation of such species under these conditions. Simulations of the dry surface are difficult to achieve without ambiguity, since the protonation states cannot change without solvent contact. The presented results also do not yet include the coverage considerations, in fact the model surface coverage is with $0.38\,\mathrm{P\,nm^{-2}}$ lower than most experiments. Further simulations of thin layers (monolayer, double layer) of different phosphoric acid content may be use to complete the atomistic description by filling the void between impregnation step and drying as well as allowing to test for surface coverage effects. This requires the use of more complex collective variables (CV) such as those described in appendix A.4 (p. 162). Even though open questions remain, the most likely surface species were identified as bidentates and potentially monodentates which form exclusively by water substitution from μ_1 sites. These are ν_2^{bc} and ν_2^{de} but it excludes the most stable bidentate at drying conditions, ν_2^{eB} as it binds to an alumina oxygen site.

The formation of higher dentations involving alumina oxygen sites can be excluded during impregnation since they are endergonic at reaction conditions, except in the presence of polyphosphates in solution. As alternative to pyrophosphates during impregnation, the same types of surface phosphorylation may occur at the drying conditions, since temperatures are higher and a drop in H_2O solvent should activate the reaction. It would be interesting to explore the drying process with a similar set of calculations in which the phosphorylation reaction is explored for orthophosphates. Different phosphate loadings could actually induce further reactions, including polyphosphate condensation. Chapter 3 indicated, which model coverages would have to be reached.

6. Machine learning NMR chemical shifts

During the last decade, the field of machine learning got a tremendous amount of attention in all material sciences.[126–128] The two main reasons for this growing interest are (i) the availability of computing resources for this type of research and (ii) the availability of steadily growing databases. Still, the field is very new and at this time, without generally established strategies.

There are very different problems that could be addressed by machine learning methods. The overall idea is to build a structure-activity relationship model on complex data sets with many variables, where subtle correlations can be exploited in an automated way, without the need of human intervention to discover and quantify them. This, however, does not necessary mean that the prediction model cannot anymore be interpreted, but that interpretation follows the model construction. Interpretation of the prediction scheme is usually not straightforward and in many cases not necessary. In such cases, machine learning models are used as "black box". A good example for this would be a machine learning interpolation of the potential energy surface (PES) for geometry optimization.[129] The success of this method is easily understood when considering that most geometry optimization algorithms are at best based on a parabolic approximation to the real potential hyperplane, which only applies in close proximity to the real minimum. The machine learning approximation for the PES is used to update the structure, and the structures improve the estimate. The final result is not further transferable however, at least in case of ref. [129], because the machine learning estimates are based on cartesian coordinates, which are difficult to generalize. For that purpose, dedicated chemical descriptors exist which are explained in section 6.1.

Long before computational NMR predictions, empirical relations between ^{31}P NMR and structural parameters were tested.[76,130] In case of ^{31}P NMR, typically only single parameters like average bond distance *or* bond angles were correlated with an NMR chemical shift. Dawson and Asbrook unravelled the combined effect P−O bond distances and P−O−Al angles in aluminium phosphates by screening calculations at DFT level and GIPAW chemical shift simulation and multivariate regression. Their results demonstrated that the NMR prediction depends on multiple structural features, of which P−O bonds and P−O−Al angles are the leading contributions. The authors proposed several regression functions with mean absolute errors (MAE) of 0.49 ppm – much more accurate than simpler correlations of ^{31}P NMR with either P−O bonds (quadratic fit: MAE = 3.9 ppm) or P−O−Al angles (linear fit: MAE = 7.1 ppm), that both were used in previous works.[130,131] Their work method can be applied to a range of aluminium phosphates, however, it cannot be transferred to more complex chemical environments that involve bonds other than a local structure of $[\mathrm{P(OAl)_4}]$. Dawson and Asbrook's first study was even more specific by only studying aluminium phosphate clusters with tetrahedral aluminium atoms, which they extended for more complex compounds involving octahedral alumina.[69,75]

Chaker et al. recently identified a reliable method to learn chemical shifts from ab initio calculations by combining several local chemical descriptors with different machine learning methods.[77] Their work was focussed on the solid-state NMR chemical shift prediction of ^{29}Si, ^{17}O, ^{23}Na and ^{27}Al nuclei in aluminosilica glasses based on DFT geometry data and GIPAW calculations. The authors combined three atom centered descriptors with seven common machine learning algorithms and found that the SOAP descriptor (smooth overlap of atomic positions) in combination with linear ridge regression (LRR) produces accurate results for their purpose (^{29}Si: RMSE = 1.3; ^{17}O: RMSE = 1.5; ^{23}Na: RMSE = 1.5; ^{27}Al: RMSE = 1.5) This chapter applies the methodology of Chaker et al. to the calculated ^{31}P chemical shifts results from the database generated in chapter 3. As starting point prediction schemes based on SOAP/LRR were used to generate a prediction model for ^{31}P chemical shifts. This work was only possible in collaboration with Tao Jiang, Research Engineer at ENSL, who carried out the machine learning parametrizations and helped during evaluation.

6.1. Chemical descriptors

The value of machine learning frameworks is to accept arbitrarily complex input data and systematically process it in order to extract relevant information. Those involve regression models, i. e. machine learning algorithms like LRR, neural networks or boosted trees. For a table with property columns and one set of data points per line, machine learning algorithms are typically applied directly, for example to train a prediction algorithm for one property based on correlation with the others. More complex, unstructured data, such as pixel information in images, require considerable adjustments in order to preprocess the information before using it as input for a machine learning algorithm. For pixel images, those might be scanning of local image snippets on a grid which is necessary to learn features irrespective of their global position in an image. Similar considerations have to be made when processing atomistic data. There are two common types of encoding atoms in chemistry: first by defining a connectivity graph, and second by explicitly providing atomic coordinates in space. The former is most common for molecules – essentially IUPAC names of molecules are an encoding of such graphs; the latter are typically geometry representations which could be a computed or for instance crystallographic positions. Even though connectivity graphs contain much less information than the atomic coordinates, Jonas and Kuhn recently used an experimental database of 43k molecules to successfully generate accurate prediction schemes for ^{1}H and ^{13}C NMR chemical shifts using graph neural networks, with RMSE = 1.2 ppm (^{13}C).[132]

Efforts towards empiric NMR chemical shift predictions are not new, HOSE codes are well established for ^{13}C and ^{1}H NMR for more than 30 years (albeit lower accuracy, for ^{13}C RMSE = 4.2 ppm).[132,133] West explored similar graph neural network methods experimental ^{31}P NMR data in 1993, but did not reach satisfactory accuracy (at best about 20 ppm).[76] The author rationalized this with errors in the included databases and weaknesses in the applied descriptors and machine learning methods at the time. This is consistent with Jonas and Kuhn, who also filtered the available database, removing about 25 % based on various criteria and added several input features (like hybridization and aromaticity) for each atom explicitly.[132] It shows that the applied neural networks do not automatically derive such information, even though most of them are implicitly contained in the molecular graph. As an alternative to the connectivity graph, one can also consider atomic representations as lists of cartesian coordinates, which is physically much more accurate since these contain exact positions and with it bond distances and angles. However, even bonds are not defined in a coordinate list, this is left to be recognized by an algorithm. In some cases, machine learning algorithms can use them directly, like in interpolation schemes which produce estimates on few data points,[129] but in most cases one needs to process the atom list in a way that allows for generalization over a wide range of different systems with different size. Those generalizations are chemical descriptors where the structural information in the atom list is encoded in a well defined form. A descriptor must fulfill the following requirements:[134]

- It must faithfully represent the properties of interest. No two different structures should generate the same descriptor.
- It should be continuous and differentiable to quantify structural similarities.
- The size and dimensionality of the descriptor needs to be constant, independent of the input size.

- The descriptor should be invariant to rotations, translations, inversion, and list permutations.

Furthermore, the descriptor can be global, capturing the whole geometry or local, atom-centered, in which case it is constructed to capture a chemical environment. It may or may not be suitable for periodic structures like crystals. A simple global descriptor is the Coulomb matrix, a matrix that contains two-body terms $\frac{Z_i Z_j}{|\mathbf{r}_i - \mathbf{r}_j|}$ of the Coulomb repulsion of the nuclei. The Coulomb matrix is a unique representation of the system of interest (which means that it is “faithful”) and it is invariant to rotation, translation and inversion by construction, and can be made invariant to list permutations by sorting (but at the cost of differentiability).[135] Several extensions of the Coulomb matrix for periodic structures exist,[135,136] as well as derived representations like the Bag of Bonds.[137] The size of a Coulomb matrix depends on the number of atoms; a constant descriptor size can be achieved by zero-padding. A different global descriptor is the many-body tensor representation (MBTR), which uses bond lengths and angles based on the chemical element. It uses one-body (g_1, elements), two-body (g_2, distance or inverse distances) and three-body terms (g_3, angles in degree or cosine). These motifs are grouped by elements, element pairs and all possible element triples, and then added together on a range $\mathcal{D}$. The scalar values are then broadened with a gaussian kernel density estimate and the range is sampled on a numerical grid, similar to a histogram:

$$\begin{aligned} \mathcal{D}(\chi) &= \sum_{g_n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\chi - g)^2}{2\sigma^2} \right), \text{ with} \qquad (6.1) \\ g &= \left\{ g_1(Z_i),\ g_2(\mathbf{r}_i, \mathbf{r}_j),\ g_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) \right\} \qquad (6.2) \end{aligned}$$

where χ is a grid point and σ is the broadening. Altogether the ranges of g_1, g_2 and g_3 define the descriptor. g_1 is equivalent to the system stoichiometry, g_2 and g_3 characterize element distances and angles in a way that is independent of the number of atoms in the system. It is straightforward to derive a local variant of MBTR (LMBTR) by defining a central atom and using a subset of ranges: g_1 is removed, since stoichiometry is difficult to establish as a local system property, and only g_2 and g_3 involving the central atom are included. The central atom is typically treated separately, independent from the element.

Since the NMR chemical shift is a local property of the chemical shielding by the local electron density around a nucleus, local descriptors are a natural choice. A popular local descriptor is the smooth overlap of atomic positions (SOAP).[138] It encodes the local environment of a central atom as smoothened atomic positions by a local expansion in orthonormal functions based on spherical harmonics Y_{lm} and radial basis functions g_n. The coefficients c_{nlm} of these basis functions for one atom sort are defined as inner product:

$$c_{nlm} = \iiint_V dV\ g_n(r) Y_{lm}(\theta, \phi) \rho(\mathbf{r}) \tag{6.3}$$

Those basis functions make the SOAP descriptor naturally a faithful descriptor of the local environment, which is differentiable, invariant to translation and permutation, and of constant size, based on a finite basis set and the number of involved elements. In order to achieve rotational invariance, the SOAP descriptor defines a partial power spectrum vector $\mathbf{p}$ of the angular momenta of all atoms in the vicinity. An element $p_{nn'l}^{Z_1 Z_2}$ of this vector is calculated by:

$$p_{nn'l}^{Z_1 Z_2} = \sqrt{\frac{8\pi^2}{2l+1}} \sum_m c_{nlm}^{Z_1} \cdot c_{n'lm}^{Z_2} \tag{6.4}$$

Note that in the implementation used here, real-valued (tesseral) spherical harmonics were used;[136] for complex-valued spherical harmonic, the multiplication is with the complex conjugate $c_{nlm}^{Z_1\,*}$. There are other local descriptors like atom-centered symmetry functions (ACSF) or angular and radial distribution functions (ARDF),[139,140] which use a different methodology to sample the distribution of atoms in a vicinity. Both can be used to construct continuous differentiable, invariant descriptors of constant size. The ACSF descriptor is different to the other descriptors, since its components are all user-defined allowing more flexibility, but it leaves it to the user to ensure faithfulness.[136] ARDF provides a consistent framework similar to SOAP, however, the author promises a much higher efficiency for systems involving multiple chemical species.[140]

Chaker et al. tested SOAP, ARDF and a simple implementation of ACSF involving only two-body terms for performance in predicting NMR chemical shifts. They demonstrated that of these descriptors, SOAP clearly leads to the most accurate prediction models. In this work, the SOAP descriptor was compared with the LMBTR in order to test the performance of this new descriptor. In comparison to SOAP, LMBTR explicitly encodes three-body terms including all atoms within its vicinity. This should be advantageous in comparison to SOAP, since angles, such as the P−O−Al angle discussed by Dawson and Ashbrook[74] needed to accurately predict the ^{31}P NMR chemical shift.

6.2. Machine learning methods

Overall, the two descriptors SOAP and LMBTR were combined with three machine learning methods: linear ridge regression (LRR),[141] the multilayer perceptron (MLP)[142] and extreme gradient boosting (XGBoost).[143] LRR is a simple technique in which a regularization term is added to a least square fit:

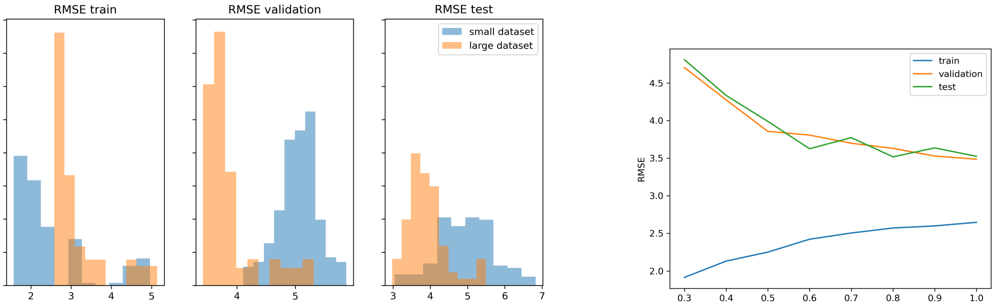
$$\mathrm{LRR} \;=\; \min\left[\underbrace{\sum_i \Big(y_i - \sum_k (m_k x_{ik} - b)\Big)}_{\text{linear least squares}} + \underbrace{\lambda \sum_k m_k^2}_{\text{regularization}}\right] \tag{6.5}$$

where x_{ik} are k independent variables for data points i for regression and y_i values are to be predicted. The linear least square fit part needs to be overdetermined ($i \gg k$) in order to quantify uncertainties. On the other hand, least square fit is not possible if the system is underdetermined ($i < k$). The regularization term is effectively a bias, leading to a systematic deviation from the least square fit results, but it provides solutions to the minimization in eq. 6.5 even if the system is underdetermined. While this seems ill-advised for the low-dimensional case or in case the independent variables are truly independent,[144] ridge regression can be very useful to identify the most important components in underdetermined data sets, provided that the variables are strongly correlated, i. e. linear dependent.[145] Like indicated, LRR works extraordinary well for linear regressions in which the input is heavily correlated, but it fails where non-linear components are crucial. The parameter λ is a so-called hyperparameter, since it is fixed by the user, as opposed to the parameters m_k and b which are optimized in the regression.

MLP and XGBoost are more advanced techniques; the former is an artificial neural network and the latter based on boosted decision tree algorithms. They offer more flexibility and are suitable for very complex data structures. It means that while LRR cannot take into account nonlinear components, and hence is inherently limited, the prediction accuracy of MLP and XGBoost can be arbitrarily improved by adding more data (and adding parameters). MLP requires defining a neural network topology between input neurons (defined by the feature array) and the output neurons (here, a single one that returns a chemical shift estimate). This can be very complex, as one can choose the number of neuron layers as well as the dimensions of each layer and the connectivity between all the neurons in every layer. For simplicity, here the number of layers is limited to a single, fully connected layer. Only the number of neurons was varied.

6.3. The data set

For the presented results, the DFT GIPAW calculations of the phosphate species at the γ-Al_2O_3 model surfaces were used to train different machine learning algorithms to predict the ^{31}P chemical shift based on the geometry. The original set of GIPAW calculations only included the most favorable cases of the sampled hydrogen bond networks described in section 3.2 (p. 58), which were 361 data points in total including



(a) RMSE comparison of smaller data set which only includes the most favorable adsorption sites (361 data points) and the full data set (1227 data points).

(b) Randomized fractions of large data set. RMSE of training, validation and test set.

Figure 6.1.: RMSE change with respect to data set size, in case of the use of on SOAP/LRR. (a) Scaled histograms of RMSE (x-axis) for small an large data set; y-axis in arbitrary units. (b) RMSE (y-axis) dependent on the data set size (x-axis).

single adsorption, coadsorption and pyrophosphates. An even smaller subset of those was discussed in chapter 4. This number of data points is small for machine learning purposes, hence additional GIPAW calculations were carried out for the less stable hydrogen bond networks from the previous sampling. Only orthophosphate coadsorption and pyrophosphate adsorption cases were used as they produce two data points per calculation. With the additional calculations, the number of data points was increased to 1227. All data points were used without filtering outliers. While it does not seem to disturb machine learning results, in one (faulty) system a peroxide geometry $PO_3(O_2)^{3-}$ was identified, which apparently did not substantially affect the results, and was therefore kept. Fig. 6.1a provides a comparison between the original set of 361 data points of the most stable structures, and the extended set of 1227 data points. A core issue of machine learning is overfitting, which means that a machine learning algorithm can produce accurate fits of the data, but the result is not transferable to new data. The RMSE of the training set actually increases, since with more data points, the degree of overfitting is reduced. The test and validation sets (explanation is given the next section) are on the other hand significantly improved. This effect on the training accuracy is due to the fact that the same method needs to find a regression on more data points for a constant set of features, which overall reduces overfitting, but also shows that one can only reach limited accuracy with the used algorithm based on SOAP/LRR. Fig. 6.1b shows a more systematic plot of this size effect. The size effect was in this case determined by picking data points from the large data set at random, which produces clear trends. The RMSE of training and test/validation converges with additional data, but the RMSE is still larger than zero. Based on the full data set, the RMSE converges between 2.7 and 3.6 ppm. This simply reflects the systematic error introduced by the machine learning model itself, in this case SOAP/LRR. Consequently, in order to improve the prediction, the machine learning model has to be changed, either by an improved structural descriptor or a better fitting algorithm.

6.4. Parametrization

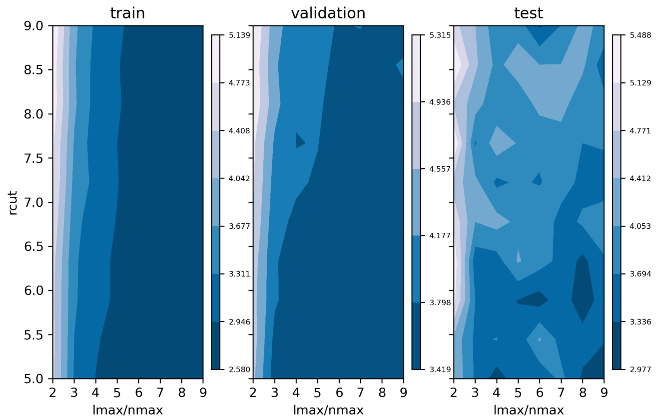
Parametrization in machine learning is a global optimization problem in which hyperparameters of an algorithm are varied in order to maximize prediction performance. In this work, it means to minimize the RMSE of the machine learning predictions of ^{31}P GIPAW results. By combining the descriptors SOAP and LMBTR with the machine learning algorithms LRR, MLP and XGBoost six different methods were generated. Besides the hyperparameters, the descriptors also depend on parameters. Since the parameters depend on one another in non-trivial ways, for all six combinations test calculations were carried out. To keep track of overfitting, cross-validation and testing were done as described by Chaker et al.[77] A test set of 10 % of the database was separated at random. For cross-validation, the remaining 90 % were split in 5 groups, 72 % for training and 18 % for validation. Training and validation were conducted for 5 times, i. e. in every case and the fraction for the validation set was swapped in the groups. Based on this procedure, descriptor parameters and machine learning hyperparameters were scanned.

6.4.1. Descriptor parameters

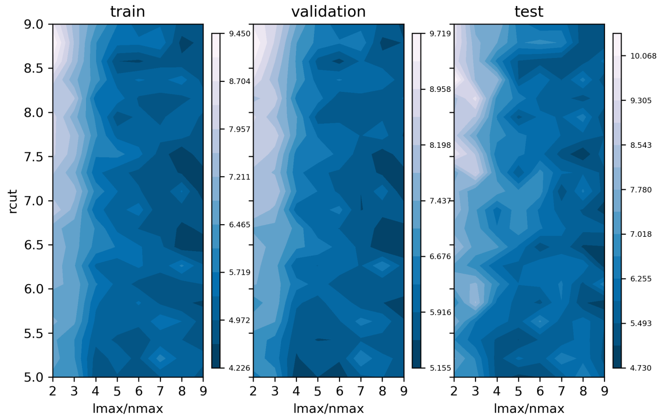
The SOAP descriptor can be manipulated by defining different basis sets of the radial (nmax) and angular (lmax) components. Furthermore, one has to define a radial cutoff which restricts vicinity and with it the number of included atoms. Lastly the atomic positions are smeared out by a gaussian kernel which can be adjusted by a parameter sigma. We tested the radial cutoff in a range from 5 to 9Å and equal values for nmax and lmax from 2 to 9. This corresponds to a feature size of 108 for lmax/nmax = 2 and 4752 for lmax/nmax = 9. Fig. 6.2 contains the results for the tests of SOAP parameters. From the ranges of the RMSE, it becomes clear that the LRR produces the most accurate predictions. Training and validation graphs of MLP and XGBoost contain notable noise, which indicates that both methods need more data points in order to work reliably. This is consistent with findings by Chaker et al.,[77] who also observed overall lower performance for more complex methods.

For LMBTR, a slightly different approach was used. In contrast to SOAP, in LMBTR ranges and grid resolution of such ranges have to be defined for the two-body and three-body terms. The two-body terms were defined in inverse units to improve resolution for small distances from 0.182 to 1.2Å^{-1}, which corresponds to a cutoff value at 5.5Å. two-body (n_2) and three-body (n_3) ranges were tested with numeric grid densities from 20 to 100 corresponding to descriptor sizes of 800 to 4000 features. Fig. 6.3 contains the results for the tests of parameters. For $n_2 < 50$, or $n_3 < 50$ more irregularities are observed. In particular, LMBTR/MLP leads to bad predictions if n_2 and n_3 are not balanced. LRR is as before better than XGBoost, which is itself better than MLP.

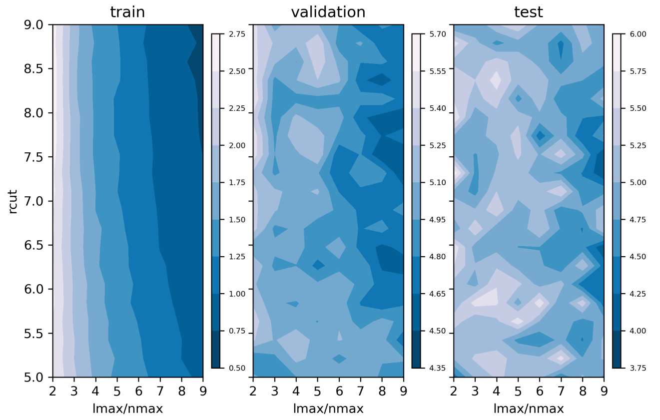
Since 28 % of the database were used for test and validation, just about 880 data points are included in each training, which means that the system was in most cases underdetermined for regression, but the features were strongly correlated. The parameter scans of LMBTR seemed overall more noisy than the results for SOAP; this is misleading



(a) SOAP/LRR, λ = 0.75

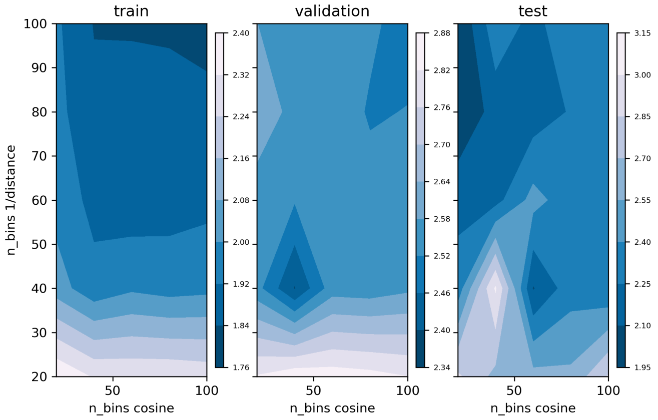


(b) SOAP/MLP, 100 neurons, 1 hidden layer

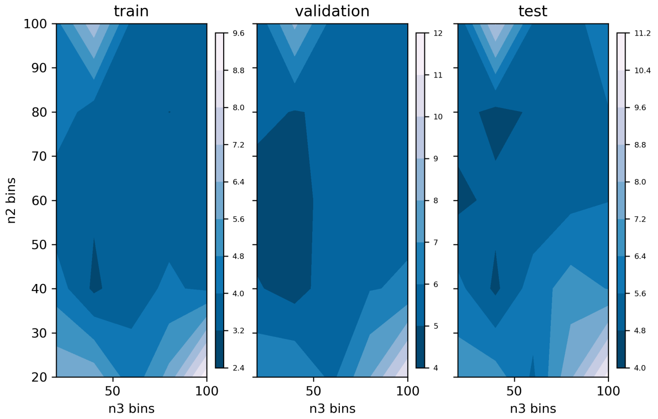


(c) SOAP/XGBoost, 500 trees, depth = 1

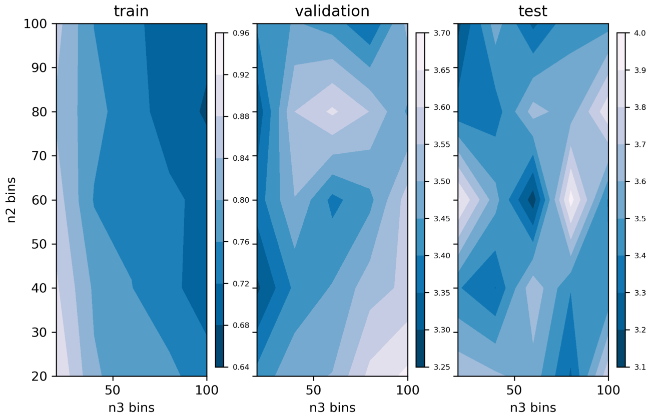
Figure 6.2.: Test of the SOAP descriptor against three different machine learning algorithms for (a) LRR, (b) MLP, (c) XGBoost; rcut and lmax/nmax see the text. Comparison of training data (left), validation (middle) and test data (right). Color coded range describes the RMSE, ranges defined individual for each graph. Validation and test RMSE were higher for MLP and XGBoost.



(a) LMBTR/LRR, $\lambda = 0.02$

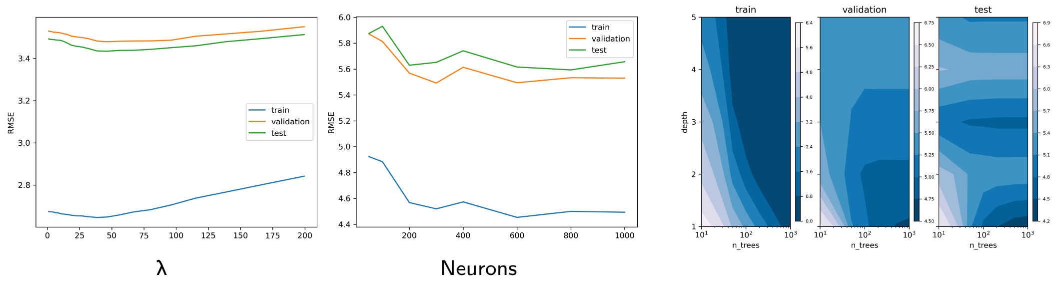


(b) LMBTR/MLP, 100 neurons, 1 hidden layer



(c) LMBTR/XGBoost, 500 trees, depth = 1

Figure 6.3.: Test of the LMBTR descriptor against three different machine learning algorithms for (a) LRR, (b) MLP, (c) XGBoost; n_2 bins and n_3 bins, see text. Comparison of training data (left), validation (middle) and test data (right). Color coded range describes the RMSE, ranges defined individual for each graph. Validation and test RMSE were higher for MLP and XGBoost. The RMSE ranges of LRR and XGBoost were narrow as n_2 bins and n_3 bins are well converged.



(a) LRR **(b)** MLP **(c)** XGBoost

Figure 6.4.: Fine tuning the parameters of LRR, MLP and XGBoost with the SOAP descriptor. The λ parameter is the only adjustable parameter for LRR. The graph in (a) shows the fitting result at rcut = 5.9 and lmax/nmax = 8; λ = 40 minimizes training, validation and test set RMSE. The MLP consists of a predefined topology of neurons, (b) plots the *best* fitting result by SOAP/MLP parameters at each data point. Beyond 200 neurons the accuracy does not improve for the test and validation sets. The tree depth and number of trees of XGBoost are indications of the permitted complexity of the model. (c) plots of the *best* fitting results for SOAP/XGBoost parameters at each data point for tree depth and number of trees. Tree depths beyond 1 improved training but deteriorated prediction quality.

in case of LRR and XGBoost, since the RMSE range for both methods is very narrow. In contrast, the MLP neural network was too small for the much larger feature input, leading to substantial fluctuations.

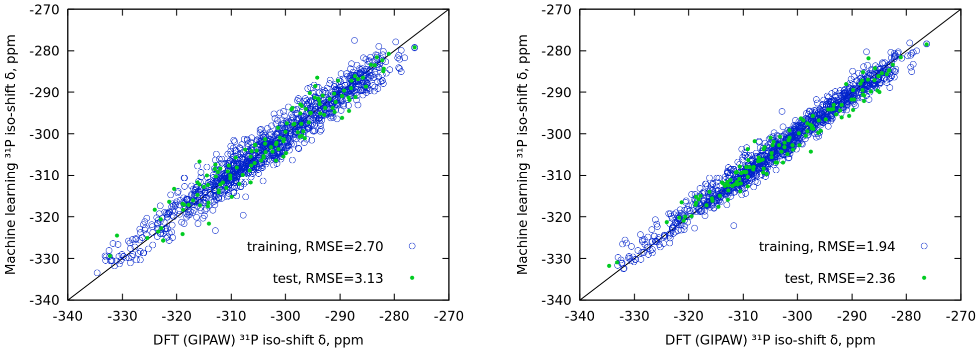
6.4.2. Hyperparameters

For LRR, the regularization parameter λ was scanned, MLP was limited to a single hidden layer, for which the number of all-to-all connected neurons was varied, and for XGBoost, both the tree depth and number of trees were scanned. The scans of all three hyperparameters give different insight into the machine learning performance.

The effect of different values for λ were tested at fixed SOAP parameters (rcut = 5.9, lmax/nmax = 8) (fig. 6.4a), whereas the hyperparameters for MLP and XGBoost were tested together with the SOAP parameters by performing a three-dimensional scan of the SOAP parameters and each hyperparameter. Fig. 6.4b and c provide the best RMSE results of the SOAP parameter scan for a given hyperparameter. As expected, the lowest RMSE were observed for the training set, while RMSE for validation and test set are similar, which means that in all cases a certain degree of overfitting was observed due to the finite database. The 'J' shaped curves for increasing λ indicate an initial performance improvement followed by monotonous increase of RMSE where regularization overly restricts the regression procedure. MLP converges with respect to the number of neurons while in contrast, increasing tree depth quickly leads to overfitting with XGBoost. Regarding overfitting, it also seems clear that it cannot be entirely avoided for a given data set with LRR and MLP. The difference between training RMSE and validation/testing RMSE was in both cases nearly constant 0.9 ppm. With XGBoost, the tree depth rapidly improved the training RMSE, while testing/validation RMSE only improves only up to a tree depth of 3. Increasing the tree depth therefore can improve the prediction performance, but it quickly leads to overfitting.

6.5. SOAP vs. LMBTR

The previous sections have shown that for the conducted tests, LRR outperforms the two more complex machine learning algorithms. This is consistent with the results of Chaker et al.[77] who found that LRR works better than any other tested algorithm for chemical shift estimations. Their data also indicate that similar algorithms like gaussian kernel ridge regression (GKRR) and elastic net regression (ENR) produce similar results with SOAP, while different local descriptors with LRR, like Behler-Parinello symmetry functions (BPSF) and the angular and radial distribution function (ARDF), were clearly inferior to SOAP. However, our results show that the prediction accuracy is improved by replacing SOAP/LRR with LMBTR/LRR. These trends imply that the limitation in the



(a) SOAP/LRR, $\lambda = 40$ **(b)** LMBTR/LRR, $\lambda = 0.02$

Figure 6.5.: Performance of machine learning result with respect to GIPAW training data. LMBTR/LRR produces more accurate predictions, than SOAP/LRR. Value for λ has been optimized separately.

prediction scheme is based on the descriptor, SOAP, rather than the algorithm, LRR. Fig. 6.5 provides a direct comparison between both methods. As opposed to the previous cross-validation scheme, the machine learning scheme uses 90 % of all data for training and 10 % for testing, thus test and training RMSE are more similar than in section 6.3. Due to overfitting, the training RMSE is still lower than the validation/test RMSE. As shown in fig. 6.1b, the training RMSE and validation/test RMSE asymptotically converge with the data set size. While it is certainly possible to extrapolate the RMSE trend, at this stage we can assume that the training RMSE as optimistic and test RMSE as conservative estimate of the accuracy of the same method for a much larger data set. The two values provide a lower and upper boundary to the real RMSE. Since the test RMSE value of LMBTR/LRR is lower than the training RMSE of SOAP/LRR, we can conclude that LMBTR/LRR is a better prediction scheme than SOAP/LRR. This trend is consistent for all the tested machine learning algorithms, which demonstrates that the LMBTR descriptor is more suitable for predicting ^{31}P NMR chemical shifts. For chemical shift estimates, the descriptor which provides the most relevant information should perform best. It is not trivial to compare both descriptors based on their features,

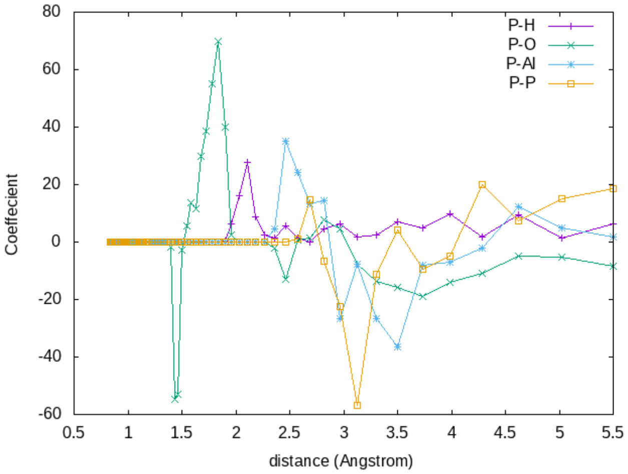


Figure 6.6.: LMBTR/LRR coefficients between the central phosphorus atom and distances for all elements in the system.
Each point is an LMBTR g_2 grid point (cf. eq. 6.1), lines added for guidance. The y-axis describes the slope (m_k) in eq. 6.5. Positive coefficients describe higher NMR shielding compared to the average structures.

since their methodologies are very different. However, both provide distance and angular information. SOAP encodes it as two-dimensional power spectrum based on gaussian radial functions and spherical harmonics, hence distances and angles are grid positions in the spectrum. This is equivalent to other descriptors which include only two-body terms, including angular information but only from the central atom X with the motif A-X-B, even though implicitly this also includes other angles. LMBTR also encodes angles around neighboring atoms such as X-A-B directly. Simple machine learning algorithms like LRR cannot recognize implicit angular information, thus the SOAP descriptor is inherently restricted by these aspects. This is also in agreement with findings by Dawson and Ashbrook who parametrized distances and angles manually using a multivariate regression.[74] LMBTR can be used as one possible way of generalizing their findings.

As discussed earlier, both SOAP and LMBTR discard information compared to the original geometry by smoothening atomic positions, and forming smooth distributions from distances and angles. Fig. 6.6 is a plot of the radial part of the numerical grid. The P$-$O distance around 1.5Å reflects P$-$O bonds. The smallest P$-$O bonds are correlated with deshielding which is due to few significantly distorted phosphate adsorbates with free P−O bonds. They are actually too unstable to be relevant here, and should thus be excluded from the data set. At slightly higher values, the P−O bonds lead to slightly positive contributions from P−O−X. The large values between 1.7 and 2.0Å are again exceptionally distorted geometries. For the smallest P−H, P−Al and P−P distances, another positive correlation is observed, which is followed by an even stronger negative correlation at a slightly larger distance in case of P−Al and P−P. This is related to the P−O−(P, Al) angles that are directly connected with P−(P, Al) distances. The lowest P−Al and P−P distances at 2.5 and 2.7Å minimize P−O−Al and P−O−P angles, which leads to high amounts of electron density around phosphorus. At larger distances, which are 3.1Å for P−P, 3.0Å for P−$\mathrm{Al_{IV}}$ and 3.5Å for P−$\mathrm{Al_{VI}}$, the angles approach 180° which leads to deshielding of phosphorus due to bond competition in the O-p_σ orbital of the phosphate (electron pulling of Lewis acidic moieties is not effective at small angles). At further distances the effects are overall more subtle except for the P−O distance, which leads to a deshielding from 3.1 to 4.5Å.

6.6. LRR performance

It is important to note that the chemical descriptors SOAP and LMBTR are both derived from the geometry and encoded in a numerical form that allows to unambiguously capture a wide range of chemical environments. For this, both methods are excessively large, saving hundreds to thousands of values as a fingerprint of generally less than 20 atoms in the vicinity. It is therefore expected that these values, when used as independent variables for regression, are strongly correlated. As discussed, ridge regression methods perform very well in automatically identifying such correlations and scaling such correlated features simultaneously. It seems that NMR chemical shift predictions are in large parts driven by a linear correlation to structural features. Up to this point, it was more important to identify the most important structural features in order to generate an ^{31}P NMR prediction scheme, than to find the most suitable algorithm. Dawson et al.[69,74,75] started from hand-selected structural features like bond distances and angles, whereas Chaker et al.[77] combined recently developed local descriptors and various more complex machine learning algorithms. Here, we also tested two algorithms despite LRR, but in agreement with Chaker et al., we realized that those do not outperform simpler regression schemes. Nevertheless, a more elaborate machine learning algorithm should outperform LRR if the descriptor captures all the relevant structural motifs, as well as potentially other quantities such as atomic charges (e. g. by Bader or Hirshfeld) Successful examples for the use of neural networks for NMR prediction exist in the literature, however, those also heavily rely on the design of the descriptor.[132]

6.7. Summary and perspectives

This machine learning project is not yet complete; however, it already provides valuable insights into the challenges for building a machine learning-based ^{31}P NMR chemical shift predictor. Recent work showed that both complex algorithms like neural networks and simple regression schemes like LRR likewise allow for accurate machine learning chemical shift predictions,[77,132] but this largely depends on the descriptors and data set size. The comprehensive tests by Chaker et al. indicate that the choice of a good descriptor, in their case SOAP, leads to the biggest improvement in NMR prediction accuracy. The simplest machine learning method, LRR, produced the best result. The success of LRR is overall surprising, but it might mean that better descriptors have to be identified before more sophisticated machine learning methods can improve the prediction model. In this study, it was demonstrated that with LMBTR, a clear improvement is observed for all machine learning methods tested. One rationale for this improvement is that LMBTR describes angles and distances independently.

At the current state many perspectives for future work are possible. One important task is to further ensure the consistence of the database. Optimizing all hyperparameters of the tested descriptor/algorithm combinations is necessary in order to complete this work. Further adjustments to the LMBTR descriptor like exchanging the representation of the angular part (g_3), using the angle directly instead of the cosine, has now proven to be more accurate. The search could also include squared components of the g_2 component of the descriptor, in alignment with the work of Dawson and Ashbrook.[74] In order to improve the system description, atomic charges might improve accuracy. They can be derived from DFT results as Bader or Hirshfeld charges. Testing more complex algorithms could also improve the prediction performance. Elastic net regression might improve the learning model, since the method is more suited for removing irrelevant components. Since the XGBoost worked best with the shallowest trees, so-called stumps (tree depth $= 1$), a simpler boosted tree method, AdaBoost, could be an alternative.

Based on the available GIPAW data, a similar machine learning method could also be applied to ^{27}Al NMR. This would provide insights into NMR shifts of aluminium surface species. Overall, using machine learning for ^{31}P chemical shifts allows for a different pathway of evaluating the chemical shifts at drying conditions. By evaluating the chemical shift in whole AIMD trajectories, it might be possible to quantify the contributions of each adsorption mode individually based on the experimental band instead of the model free enthalpy estimates.

Conclusions and Perspectives

The main goal of this work was to generate a better understanding of how phosphates interact with γ-Al_2O_3 surfaces at the atomic scale. This γ-$Al_2O_3/H_3PO_4/H_2O$ system is of special interest in the context of industrial heterogeneous catalysis in which γ-Al_2O_3 is a widely used support material. In addition, this work can be instructive in improving our understanding of different phosphate-mineral interactions and reactivity. The strategy here was to first explore the phase space for the most relevant candidates by systematic geometry generation at the DFT level of theory based on three established surface models, and including other aspects like hydrogen bond networks (cf. chapter 3). The geometry generation was carried out in two stages: first, testing all orthophosphate adsorption modes based on substitutions of water adsorption sites; then, the most stable single orthophosphate adsorption modes were again used for input generation in order to generate datasets of orthophosphate coadsorption, mimicking a higher phosphate coverage, and pyrophosphate adsorption. To the best of our knowledge, this is the first phosphate adsorption study that includes changing coverages and phosphate condensation phenomena of surface species. Altogether, this involved more than thousand surface geometries after detailed analysis of the problem and filtering potential redundancies. Most of those came from different hydrogen bond networks involving the same adsorption site. Data for 159 sites are provided in appendix A.2, with discussion for a small subset of the most stable candidates. These are based on more than a thousand geometry optimizations in order to explore the hydrogen bond network in each case. On the (1 1 0) facet, phosphate species stabilize the reconstruction of the surface, which was earlier identified at high-temperature models in which undersaturation of water adsorbates is accommodated by shifting a surface aluminium atom from an octahedral to a tetrahedral site, thereby reducing the degree of undersaturation. The stabilization by phosphates is clearly different, since the discussed systems are always coordinatively saturated. Hence it could also be expected that the reconstruction is (in the absence of chemisorbed phosphates) less stable at drying conditions. We also identified an alumina oxygen site, which is destabilized in the reconstruction and which is much more stable after phosphorylation. This initial exploration in chapter 3 was the basis for all following chapters, which was also a motivation for the phase space search that was carried out.

While exclusively theoretical models can improve our understanding of the more complex real systems, its foundation should be regularly justified on experimental grounds. In chapter 4, magnetic shielding tensors of the explored model surface species were calculated using the GIPAW method. The magnetic shielding was used in order to estimate experimental ^{31}P chemical shifts. Overall, the model structures could reproduce most of the experimental observations. ^{31}P chemical shifts of phosphate species are well-known to produce a broad signal band that is broadened and shifted to lower values when increasing phosphate loading. The comparison between the coverage-dependent chemical shifts allowed us to elaborate on changes that actually occur at the atomic scale. We can distinguish surface between the interaction effects from (i) the adsorption site, (ii) the adsorption mode and (iii) direct coverage effects (influence of phosphate increase in the vicinity of the adsorbed phosphate). The strongest effect identified was the site effect (i), which means that the relative weight of the site distribution changes with the phosphate loading. The free enthalpy estimates from DFT confirm this in comparison between single orthophosphate adsorption versus coadsorption of two orthophosphates on the (1 1 0) facet. After this the adsorption mode (ii) influences the chemical shift, in particular bidentates (ν_2) show lower chemical shift values than tridentates (ν_3). For the direct coverage effect (iii), additional chemical shift calculations with larger unit cells (lower coverage) were carried out for the selected adsorption sites. The interaction of orthophosphate with the surface can rationalize ^{31}P chemical shifts down to $-15\,\mathrm{ppm}$, but the experimental chemical shift bands can reach values down to around $-30\,\mathrm{ppm}$.[12,18,59] Two explanations for this trend exist: polyphosphate formation and aluminium phosphate formation. Our experimental collaborators recently produced evidence for the presence of polyphosphates at dry surfaces. Pyrophosphate adsorbates can actually fill the gap from -15 to $-30\,\mathrm{ppm}$ based on the presented DFT models; however, based on the free enthalpy estimates, they might not be stable enough. Moreover, the presence of terminally bound pyrophosphates can also rationalize discrepancies in two-dimensional $^{31}\mathrm{P}-{}^{27}\mathrm{Al}$ NMR INEPT experiments, which show two different distributions of phosphate species for through-bond and through-space correlation. The alternative rationale for this discrepancy are physisorbed (ν_0) orthophosphates, although those are much less stable than the chemisorbed species on (1 1 0). Aluminium phosphate formation, which was discussed earlier by Van Eck et al.[59] and DeCanio et al.[12] also depends on the overall phosphate loading. Our experimental data and our computational models show that up to a phosphate loading of 4.5 wt% or $4.1\,\mathrm{P\,nm^{-2}}$, the adsorption site distribution and polyphosphate condensation are crucial to rationalize the change of chemical shift for different phosphate loadings.

An important open question from chapters 3 and 4 is how the most stable adsorption structure can actually form at the involved experimental conditions. The initial impregnation is done in aqueous solution, in which phosphate from bulk water reaches the surface through the solid-liquid interface. Do the most stable structures at drying conditions already form during impregnation or does the solid-liquid interface produce different adsorption modes? This was studied in chapter 5. As an initial benchmark, the reaction mechanism of orthophosphate condensation to pyrophosphate was tested by well-tempered metadynamics. Even though the reaction was not yet simulated by this approach, its reaction free energy and the barriers are known experimentally and were successfully reproduced with this method. While the initial adsorption could also be assumed to happen, based on experimental observation, it was tested by a well-tempered metadynamics exploration. It was confirmed for octahedral aluminium sites, that water substitution by phosphates readily occurs for the explored case, leading to strongly adsorbed bidentates (ν_2). The phosphorylation of alumina oxygen sites in order to form tridentates was explored by well-tempered metadynamics for the three most stable adsorption sites identified at drying conditions on the (1 1 0) facet. It also included one case of a bidentate constructed from an alumina oxygen. Two configurations for the reactions were used for each site: preadsorbed orthophosphates $\nu_2 \rightleftharpoons \nu_3 + H_2O$ and pyrophosphates $\nu_{20} \rightleftharpoons \nu_3 + \nu_0$, which only differ in the leaving group, i. e. water and another orthophosphate. The bidentate (ν_2) starts from ν_1 and ν_{10}. The trend in all cases is that the reaction free energy is positive for all cases involving orthophosphates, and negative for the reaction starting from pyrophosphate. This means that, based on reaction energies, none of the reaction products are stable, as even after the pyrophosphate hydrolysis, the phosphorylated alumina oxygen can be hydrolyzed as well, i. e. $\nu_{20} \xrightarrow[-\nu_0]{} \nu_3 \longrightarrow \nu_2$. Only two of the six surface phosphorylation reaction barriers were converged; compared to pyrophosphate hydrolysis in water, with a calculated value of $100\,\mathrm{kJ\,mol^{-1}}$, they showed a higher phosphorylation barrier ($115\,\mathrm{kJ\,mol^{-1}}$ starting from n(1 1 0)ν_{20}, $146\,\mathrm{kJ\,mol^{-1}}$ starting from n(1 1 0)ν2). Consequently, since pyrophosphate hydrolysis is a slow reaction at room temperature, taking overall several days, higher dentations cannot even form temporarily during impregnation based on such barriers. These barriers also indicate that, once formed, the tridentates should be stable enough to resist hydrolysis for some time. It is also consistent with the results of Van Cleve et al. who demonstrated how phosphonate adsorption on γ-Al_2O_3 improves hydrothermal stability.[3] Overall, dentations involving alumina oxygens do not occur at the solid-liquid interface, but they are more likely under drying conditions. Once formed, these bonds are quite stable. At this point, the mechanistic studies under drying conditions are a very interesting topic for future studies.

As a result, we clearly see that higher dentations involving alumina oxygen are limited to drying conditions. This fact can also hold as an *a posteriori* justification for using 120°C as reference of the dry model surfaces (chapter 3) for comparison with the experimental ^{31}P MAS NMR results. Those experiments were done with samples dried at 120°C but they were kept at room temperature in contact with air. Hence, it was an open question which temperature is the better reference. In fact, the most stable adsorption structure identified was $(1\,0\,0)\nu_0$ at room temperature, but this structure is on a minority facet and does not reproduce the ^{31}P chemical shift (chapter 4). Only by assuming the thermodynamic stability ordering at the drying conditions themselves, free enthalpy estimates and experimental results are consistent. The barriers identified in chapter 5 rationalize why this is the case: at room temperature, surface hydration of P−O−Al bonds is kinetically inhibited.

The last chapter, chapter 6, demonstrates that it is possible to generate efficient prediction schemes of ^{31}P chemical shifts based on atomic positions based on linear ridge regression, a simple machine learning regression algorithm. This study was only possible due to the large structural exploration carried out in chapter 3 (p. 51), which allowed to generate 1227 data points for ^{31}P GIPAW chemical shifts as the basis for this work. To the best of our knowledge, this work is the first application of machine learning on^{31}P NMR of phosphate surface species; nevertheless, several previous studies dealt with bulk materials involving many more data points per GIPAW calculation or cluster models.[69,74,75,77] The recent local chemical descriptor LMBTR (local many-body tensor representation)[146] was applied for the regression, leading to a significant improvement over one of the most popular local descriptors (smooth overlap of atomic positions, SOAP).[138] This opens the possibility for future work: a study of the chemical shift including dynamic properties of the system, potentially quantifying the proportions at which different adsorption modes contribute to the ^{31}P NMR signal. Those can be sampled by (biased) molecular dynamics, for which GIPAW calculations would be prohibitively expensive. A similar study could also be conducted for ^{27}Al NMR in order to computationally reproduce the two-dimensional INEPT correlation NMR experiments.

Overall, the discussed structures are the first realistic adsorption models available for the γ-Al_2O_3/H_3PO_4/H_2O system at the time, and as such they should be used as the basis for future DFT studies of heterogeneous catalyst models involving γ-Al_2O_3, phosphate additives and other active compounds. Their plausibility was demonstrated by comparison with ^{31}P NMR experimental results (chapter 4, p. 75), with a perspective on further quantification of the spectra by generating a machine learning NMR prediction scheme and using it to test NMR chemical shifts on trajectories of molecular dynamics. The metadynamics simulations presented in chapter 5 (p. 91) allowed to disentangle which adsorption structures occurred at which of the two tested conditions, dried or wet surface. In the context of future research, this work vastly improves our understanding of the effects of phosphates at γ-Al_2O_3 surfaces which paves the way to construct new, more holistic models of heterogeneous catalysts.

A. Appendix

A.1. Electronic structure methods

Today, computational chemists have access to a wide variety of methods in order to produce atomistic models for their systems of interest. These methods can largely vary in computational expense and general accuracy as well as in robustness and practicality. To start from the fundamentals, the most basic approximations are introduced in this section with a short description of the underlying physics, followed by an outline of a few notable methods. Since quantum mechanical particles are represented by wave functions which are continuous probability amplitudes, the underlying set of basis functions is a crucial part for the accuracy of the systems. Both aspects are explained in more detail below.

A.1.1. The total energy

A system's energy is a conserved quantity of any physical system from large interacting bodies down to molecules and crystals. The latter are described in atomistic models of atom nuclei and electrons. An accurate determination of the behavior of either of these particles requires quantum mechanics. The total energy of a quantum mechanical system is given by the Schrödinger's equation which defines the Hamilton operator $\hat{H}$

that is applied on a wave function $|\Psi\rangle$ in order to determine the total energy, $\mathbf{E}$, of the system:[147]

$$\hat{H}\,|\Psi(\mathbf{r}_\mathrm{N},\mathbf{r}_\mathrm{e})\rangle \;=\; \mathbf{E}\,|\Psi(\mathbf{r}_\mathrm{N},\mathbf{r}_\mathrm{e})\rangle \tag{A.1}$$

For non-stationary states with its time-dependent form,

$$i\hbar\frac{d}{dt}\,|\Psi(\mathbf{r}_\mathrm{e},\mathbf{r}_\mathrm{N},t)\rangle \;=\; \hat{H}\,|\Psi(\mathbf{r}_\mathrm{e},\mathbf{r}_\mathrm{N},t)\rangle \tag{A.2}$$

The wave function $|\Psi\rangle$ is a descriptor that – in accordance with Heisenberg's uncertainty principle – represents the many-body system of particles by probability amplitudes of their positions $\mathbf{r}_\mathrm{e}$ (electrons) and $\mathbf{r}_\mathrm{N}$ (nuclei). By taking advantage of the Dirac notation,[148] we can use a short form to multiply with the conjugate transpose of the wave function (implying that it is represented by some matrix), in order to separate observables and wave functions. For example, eq. A.1 becomes

$$\langle\Psi(\mathbf{r}_\mathrm{N},\mathbf{r}_\mathrm{e})|\,\hat{H}\,|\Psi(\mathbf{r}_\mathrm{N},\mathbf{r}_\mathrm{e})\rangle \;=\; \mathbf{E} \tag{A.3}$$

Analogously to classical mechanics, the Hamilton operator for charged particles can be written as a sum of all operators that determine the kinetic energy of translation $\hat{T}$ and Coulomb potential energy $\hat{V}$ operators for every particle and particle pair:[149][148]

$$\hat{H} \;=\; \sum_i \hat{T}_i + \sum_i \hat{V}_i \;=\; -\frac{\hbar^2}{2}\sum_i \frac{\Delta_i}{m_i} + \frac{1}{4\pi\epsilon_0}\sum_i\sum_{j<i}\frac{q_i q_j}{|r_i - r_j|} \tag{A.4}$$

Since the Coulomb terms are sums over particle pairs, they rapidly grow with the number of particles. If we are furthermore interested in dynamic information of, i. e. structural changes over time, we would have to solve the time-dependent Schrödinger equation. As first approximation, we separate nuclear and electronic terms:

$$\hat{H} \;=\; \underbrace{\hat{T}_{\mathrm{N}} + \hat{V}_{\mathrm{NN}}}_{\text{nuclei}} + \underbrace{\hat{T}_{\mathrm{e}} + \hat{V}_{\mathrm{Ne}} + \hat{V}_{\mathrm{ee}}}_{\text{electrons}} \tag{A.5}$$

The terms each refer to the sum of operators of involved electrons e and nuclei N, i. e. $\hat{V}_{\mathrm{NN}}$ refers to the electrostatic repulsion of all nuclei. Nuclear movement is much smaller than electronic movement given their mass ratio of $m_{\mathrm{N}}/m_{\mathrm{e}} > 10^3$, as considered in the Born-Oppenheimer approximation.[149] Nuclei are then given as fixed point charges, and $\hat{T}_{\mathrm{N}}$ is zero by that definition. $\hat{V}_{\mathrm{NN}}$ is included to the electron part as constant term. Even though this is not strictly necessary, it simplifies the construction of energy derivatives like forces and Hessian matrices. The full Schrödinger equation can be approximated by its (stationary) electronic part and an independent term $\mathbf{T_N}$ for nuclear motion.

$$\langle \Psi(\mathbf{r}_{\mathrm{N}}, \mathbf{r}_{\mathrm{e}})| \, \hat{H} \, |\Psi(\mathbf{r}_{\mathrm{N}}, \mathbf{r}_{\mathrm{e}})\rangle \;\approx\; \mathbf{T_N} + \langle \psi_e(\mathbf{r}_{\mathrm{e}})| \, \hat{H}_{\mathrm{e}}(\mathbf{r}_{\mathrm{N}}) \, |\psi_e(\mathbf{r}_{\mathrm{e}})\rangle \tag{A.6}$$

In prospect of molecular dynamics, this equation provides the total energy of a dynamic system, where the nuclear positions change over time, $\mathbf{r}_{\mathrm{N}}(t)$, while the electronic ground state is stationary, that means the electrons instantly adapt to the moving coordinates. In general $\mathbf{T_N}$ is the expectation value of a nuclear wave function that is only defined by nuclear movement in the potential given by the electronic wave function and $\hat{V}_{\mathrm{NN}}$. In order to solve the electronic part of the Schrödinger equation, a mathematical expression for the wave function is needed.

The first approximative solution was given by Hartree in 1927 a year after publication of the Schrödinger equation itself.[150][148] He expressed the many-electron wave function as product of single-electron wave functions, the Hartree product. Since it violates the principle of antisymmetry of proton exchange in the wave function,[151] it was extended in the following few years by constructing a determinant of Hartree products, the Slater determinant, which is a construction that satisfies this principle. This method, later referred to as Hartree-Fock method, does not directly produce solutions to a ground state wave functions but is capable of iteratively improving a trial wave function by optimizing the one-electron states or orbitals in the average electric field of the other orbitals. Hence it is also called self-consistent field (SCF) method. The most critical approximation here stems from the fact that each electron orbital is expressed in the mean-field of the *other* electrons. In consequence, electron correlation phenomena other than covalent bonds, like van-der-Waals interaction cannot be reproduced. To alleviate this problem, many improvements were proposed, for example based on perturbation theory, where the Hartree-Fock solution which is exact for single electrons is improved by adding perturbation terms to approximate the neglected Coulomb correlation effects. Most noteworthy is Møller-Plesset perturbation method since it can efficiently produce accurate results by even current standards,[152] while it was published in 1934 where calculations were limited to the simplest atomic or diatomic cases. Only in 1951 the Hartree-Fock method was supplied with an improved mathematical framework by Roothaan as one of the first steps that allowed it to be run on computers.[153] The Møller-Plesset perturbation method is the earliest *post*-Hartree-Fock method and was long seen as "gold-standard" for *ab-initio* models of organic molecules, however its approximation is only accurate for large energy gaps between occupied and virtual orbitals, and, as opposed to the Hartree-Fock method, it is not variational. As a consequence, perturbation methods are difficult to improve other than by including higher order correction terms. It is common to interpret the additional terms as excitations with first order terms as single excitations, second order terms as double excitations, etc, however, rather than real excitation phenomena, they are terms formed by excited state Slater determinants from the ground state optimized Hartree-Fock orbitals. Since the single excitations do not overlap with the ground state, the first perturbative correction is achieved by inclusion of double excitations, within the second order Møller-Plesset perturbation method (MP2). Higher order corrections like triple excitations in MP3 or beyond may improve the accuracy of the result at computational expense and numerical stability. The fact that virtual excited states are included also leads to high demands for the used basis functions, needing a more “complete basis” set in order to be accurate.

In modern jargon, Hartree-Fock and in particular *post*-Hartree-Fock methods termed as wave function methods. They include methods that can be exact in principle, like configuration interaction and are rather computationally demanding, limiting its applicability to a smaller scale. The reason is the electron-electron repulsion term which rapidly increases computational cost with the number of electrons which leads to an asymptotic computation time of $\mathcal{O}(n^4)$. Many alternatives were proposed, which are related with wave function methods in the sense that they solve the electronic Schrödinger equation while replacing some of the terms in eq. A.5 by numerically simpler ones. The most common is based on density functional theory (DFT) in which the explicit calculation of the largest sum $\hat{V}_{\mathrm{ee}}$ can be substituted. While otherwise iteratively minimizing the total energy. Density functional theory (DFT) is presented in the following section in more detail. *Empirical* or *semiempirical* methods can be characterized by replacing the iterative procedure by empirical parameters, for example to build a wave function and access orbital energies and total energies by a single diagonalization, or by replacing the whole electronic wave function in eq. A.6 by classic spring potentials and force fields that mimic forces of the electron structure on the nuclei.

A.1.2. Density functional theory

DFT arisen with the study of the inhomogeneous electron gas of Hohenberg and Kohn, who proved that the density distribution ρ_0 can completely describe the physics of a quantum mechanical ground state system, like the electronic wave function $|\psi_e(\mathbf{r}_\mathrm{e})\rangle$.[154] The Hohenberg and Kohn theorem was reformulated and generalized over the years,[155] but it is just a proof of existence and does not give any guidance on how to replace a wave function by a density. It was also demonstrated to break down on excited states;[156] while this does not disqualify DFT as a method, it shows its limits. It nevertheless proved to be superior in terms of computational efficiency and comparable accuracy.

The general idea is that the ground state density of a system is, like the total energy, an observable of the system. It can be determined from the wave function by summation over the orbital densities:

$$\rho(\mathbf{r}) = \sum_i |\varphi_i(\mathbf{r})|^2 \tag{A.7}$$

According to the Hohenberg and Kohn theorem, in principle, this relationship can be reversed, i. e. $|\psi_e(\mathbf{r}_\mathrm{e})\rangle = |\psi_e[\rho_0]\rangle$, where the brackets indicate that the wave function is a functional of the ground state density. If the wave function is fully determined by its density, the Schrödinger equation can be expressed as density functional as well:

$$\mathbf{E}_{el} \;=\; \langle\psi[\rho_0]|\,\hat{H}_e\,|\psi_e[\rho_0]\rangle \;=\; \mathbf{E}_{el}[\rho_0] \tag{A.8}$$

This implies that for most cases should be possible to replace a many body wave function, a complex 3N-dimensional construction comprised of densities for every particle N, it contains by one single 3-dimensional density distribution. Consequently calculations based on the density are much less time consuming, and, regarding numerical stability for convergence, much more robust. Most of the terms shown in eq. A.4 and eq A.5 are Coulomb interactions, which can be expressed as integrals over the density:[157]

$$\begin{aligned}\mathbf{E}_{el}[\rho_0] &= T_\mathrm{e}[\rho_0] + V_\mathrm{Ne}[\rho_0] + V_\mathrm{ee}[\rho_0] \\ &= T_\mathrm{e}[\rho_0] + \int d\mathbf{r}\,\mathcal{V}_\mathrm{N}(\mathbf{r})\rho_0(\mathbf{r}) + \frac{e^2}{2}\iint d\mathbf{r}d\mathbf{r}'\,\frac{\rho_0(\mathbf{r})\rho_0(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\end{aligned} \tag{A.9}$$

$\mathcal{V}_\mathrm{N}$ is the static Coulomb field of the nuclei, which is commonly generalized as external field. The integration over $\mathbf{r}$ and $\mathbf{r}'$ is over the density and pair density and is usually implemented as numerical integration. With these integrals it is possible to calculate the potential energy components of the electronic structure exactly, and only based on the electron density. However, there is currently no exact density functional known for the kinetic energy $T_{\mathrm{e}}[\rho_0]$. Thomas and Fermi independently found an analytical approximation for the kinetic energy in 1927:[158,159]

$$T_{\mathrm{e}}^{\mathrm{TF}}[\rho_0] = \frac{3h^2}{40m_e}\left(\frac{3}{\pi}\right)^{\frac{2}{3}} \int [\rho_0(\mathbf{r})]^{\frac{5}{3}} \, d\mathbf{r} \tag{A.10}$$

While this model was also extended to account for exchange energy correctly,[160] the approximation is not even accurate enough to predict covalent bonds in molecules.[159] This problem can in principle be overcome by finding a better approximation for $T_{\mathrm{e}}[\rho_0]$. Recently Snyder et al. demonstrated that it is possible to find an improved kinetic energy term by machine learning,[161] and several other groups explored different possible machine learning approaches.[162,163] While these may improve in the future, and allow systematic improvement of a "pure" kinetic energy functional, the most established method involves the same mean-field kinetic energy terms as in the Hartree-Fock method. Kohn and Sham combined the DFT terms in eq. A.9 with a wave function build from single Slater determinant of non-interacting electrons in a fictitious potential. The Kohn-Sham kinetic energy can be expressed by the orbitals $|\varphi_i(\mathbf{r})\rangle$ of this wave function:

$$T_{\mathrm{e}}^{\mathrm{KS}}[\rho_0] = \sum_i \langle\varphi_i(\mathbf{r})| - \frac{\hbar^2\Delta}{2m}|\varphi_i(\mathbf{r})\rangle \tag{A.11}$$

The Kohn-Sham kinetic energy is much more accurate than any analytical approximation, but introduces the subtle errors, inherited from the mean field approximation. It is generally collected in an unknown term, the exchange-correlation potential $V_{xc}[\rho_0]$, which is empirically approximated based on the density. It is the only approximative term of the Kohn-Sham energy:

$$\mathbf{E}_{el}^{\mathrm{KS}}[\rho_0] = T_{\mathrm{e}}[\rho_0] + V_{\mathrm{Ne}}[\rho_0] + V_{\mathrm{ee}}[\rho_0] + V_{xc}[\rho_0] \tag{A.12}$$

The non-interacting electrons in the wave function part are produced by SCF minimization. Their orbital energy levels ε_i are given by:

$$\varepsilon_i = \langle \varphi_i(\mathbf{r}) | \big[-\frac{\hbar^2 \Delta}{2m} + \mathcal{V}_s(\mathbf{r}) \big] | \varphi_i(\mathbf{r}) \rangle \tag{A.13}$$

The fictitious potential $\mathcal{V}_s(\mathbf{r})$ depends on the density itself:

$$\mathcal{V}_s(\mathbf{r}) = \mathcal{V}_{\mathrm{N}}(\mathbf{r}) + e^2 \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, d\mathbf{r}' + \mathcal{V}_{xc}(\mathbf{r}) \tag{A.14}$$

Starting from an initial guess, the density and orbitals are iteratively optimized, and new densities produced by eq. A.7. Kohn-Sham DFT is a compromise between wave function method and *pure* DFT. While the wave function based kinetic energy introduces a single reference determinant, it allows for an accurate estimate of the kinetic energy; hence the correction added in $\mathcal{V}_{xc}(\mathbf{r})$ is smaller and would therefore introduce smaller inaccuracies. The mix between both sides sides also leads to a clear computational advantage, since the $\hat{V}_{\mathrm{ee}}$ part could be substituted by a much smaller integral $V_{\mathrm{ee}}[\rho_0]$. Its asymptotic computation times scales with $\mathcal{O}(n^3)$ for a system of size n. Several approximations are available for the exchange-correlation potential which can be ordered by a dependence on the local density, its first derivatives, and further corrective measures: the local density approximation (LDA) only includes terms that scale with the density, while general gradient approximations (GGAs) include density changes and meta-GGAs add kinetic energy corrections to $V_{xc}[\rho]$. As alternative, so-called hybrid DFT methods substitute varying amounts of their approximate exchange by exact exchange from Hartree-Fock terms. Perdew described this process of including increasingly complex terms into a successively more accurate method with the figure of speech of Jacob's ladder.[164] In more recent years the focus was on improving the description of dispersion interaction, which is included in wave function methods like MP2 but at high computational cost. Grimme et al. created several method agnostic correction potentials which, irrespective of the functional, add an empirical term based on the atomic positions.[89] These schemes aim to determine long range correlation effects by parametrized force fields independent of the density. The method was also adapted for charge-dependent polarizability by estimating atom charges from the density.[165] On the other hand Klimeš et al. built whole new functionals from scratch.[166] Overall, DFT methods are still evolving, with continuous development of more accurate functionals and efficient implementations.

A.1.3. Basis functions and pseudopotentials

The electronic structure is described by a set of mathematical functions that do not only have to be well suited to describe its topology but they also need to ensure numerical stability for ground state optimization. At the same time it must be possible to evaluate their accuracy and transferability, while in particular small basis sets should be reasonably accurate.

In computational chemistry, which emerged from wave function methods, in the past almost exclusively Gaussian orbitals were used. Gaussian orbitals are localized at the atomic positions and are parametrized to describe atomic orbitals and molecular orbitals alike. Angular components are included by multiplying with spherical harmonics. They are commonly "contracted" in fixed linear combinations to approximate more physically sound Slater-type orbitals without increasing the number of available basis functions. Since Gaussian orbitals are not orthogonal, but the determined molecular orbitals must be orthogonal, one has to ensure, that linear dependencies are avoided. This is unproblematic as long as the Gaussian basis set does not include too diffuse basis functions.

In periodic boundary conditions the most "natural" functions are plane waves (PW). They are likely the most popular basis set in computational solid state physics for decades. Plane waves are similar to a particle in a box, where a quantum mechanical particle is described by a superposition of a infinite discrete set of trigonometric functions. In contrast to Gaussian basis sets, PW basis sets are defined by the unit cell dimensions and independent of the number of atoms in the system. To actually allow computations with PW basis sets, it has to be truncated to a finite size, usually by choosing a maximum frequency. Plane waves are ideal to describe smooth continuous densities, however, local density spikes like atom nuclei would require a very large number of plane waves. For the interfacial systems described here, it is also important to note that even vacuum volume in surface models contributes to the basis set size.

To effectively save computation time without affecting the accuracy of the chemically most relevant valence states (which are relatively smooth), the core electrons are substituted by pseudopotentials. Pseudopotentials are in essence radial functions that describe the same electrostatic field as the nucleus at long distances, while the core region is optimized in such a way that the remaining electron orbitals as well as the overall density remains the same outside of the core region given by the pseudopotential. Thus, effectively the number of explicitly calculated electrons is reduced to valence electrons. Calculations with Gaussian basis sets can be done without pseudopotentials, but in particular heavy elements are not accurately represented without relativistic models of the core electrons. To avoid more costly relativistic methods, the pseudopotentials are constructed including relativistic effects. For PW methods, pseudopotentials are crucial in order to smoothen the electron density in close proximity to the nuclei. This in fact changes the density description in close vicinity to the nucleus without physical justification. A simple method for this is the use of norm-conserving pseudopotentials.[167] While the shape of the density close to the nuclei is changed, the number of valence electrons is maintained. Ideally, one would be able to recover the *all-electron* density after carrying out pseudopotential calculations. This can be in principle be achieved within Blöchl's projector-augmented wave (PAW) formalism.[101] The method uses a much smaller PW basis to built Kohn-Sham orbitals than for the electron density. With the method, all-electron densities can be produced from pseudo-orbitals. Nevertheless the most common application of PAW is as theoretical generalization of Vanderbilt's ultrasoft pseudopotentials (USPP),[168] as for example implemented in VASP.[102] The motivation of USPP is, rather than constructing an all-electron density, to soften the requirement for the orbital plane waves beyond limitations due to norm conservation by introducing local components to the plane wave orbitals.

In Kohn-Sham DFT, the kinetic energy is calculated by the construction of non-interacting Kohn-Sham orbitals in an effective field that is depends on the electron density. The other terms of the total energy are directly determined as functionals of the electron density. For the choice of basis functions and pseudopotentials, this combination of orbitals and electron density leads to inconsistent numeric requirements. Calculations based on the density improve with a larger, more flexible basis set, but the computational cost for orbital energy calculations in plane wave basis functions increases rapidly, ultimately leading to the bottleneck of this method. As described above, within Gaussian basis sets this problem is addressed by keeping the number of available basis functions constant, and improving their quality in contractions that are more suitable to describe orbitals than single Gaussian orbital functions. For the PAW method, a small PW basis set is used for description of the orbitals, and a larger auxiliary basis for the density.

The Gaussian and plane wave (GPW) method does the same separation between Kohn-Sham orbitals and density calculations, where the former is calculated with Gaussian orbitals and the latter – after transformation of the Gaussian basis – in an auxiliary plane wave basis. The Gaussian basis sets are adapted for the use of the same pseudopotential as the auxiliary basis.[104] Compared to PW methods, GPW is similar to the use of norm-conserving pseudopotentials, both basis sets rely on the same pseudopotentials. However it is much more efficient and scalable since orbitals and densities can be much better evaluated in those basis sets, while additional computational load by transformation is usually negligible.

A.2. Database tables

Table A.1.: Free enthalpy of adsorption of phosphate in kJ mol^{-1} and 31P NMR shift in ppm with respect to to 85 % H_3PO_4 for the (at 393 K) most stable system of every tested single adsorption site, labelled according to fig. 3.1a (p. 53).

			ΔG_{ads}		^{31}P NMR
facet	site	(25°C)	(120°C)	(300°C)	
(1 0 0)	ν_0	−152	−126	−78	7.1
(1 0 0)	ν_1^a	−128	−121	−108	6.3
(1 0 0)	ν_1^b	−87	−83	−75	13.5
(1 0 0)	ν_1^c	−95	−88	−74	13.5
(1 0 0)	$\nu_2^{aa'}$	−99	−113	−138	5.7
(1 0 0)	ν_2^{ab}	−144	−156	−178	4.2
(1 0 0)	ν_2^{bc}	−126	−140	−166	6.3
(1 0 0)	ν_2^{cb}	−97	−111	−136	13.0
(1 0 0)	$\nu_3^{aa'B}$	−41	−72	−129	5.4
(1 0 0)	$\nu_3^{aa'C}$	−76	−107	−165	5.7
(1 0 0)	ν_3^{abA}	−81	−112	−170	8.2
(1 0 0)	ν_3^{abc}	−91	−122	−181	11.7
(1 0 0)	ν_3^{acb}	−31	−63	−123	13.3
n(1 1 0)	ν_0	−78	−56	−15	11.9
n(1 1 0)	ν_1^a	−81	−76	−67	5.6
n(1 1 0)	ν_1^b	−103	−100	−95	1.5
n(1 1 0)	ν_1^d	−91	−85	−75	8.6
n(1 1 0)	ν_1^e	−109	−105	−98	2.6
n(1 1 0)	ν_2^{ab}	−60	−74	−100	1.9
n(1 1 0)	$\nu_2^{bb'}$	−118	−132	−158	−8.6
n(1 1 0)	ν_2^{de}	−119	−132	−159	−15.0
n(1 1 0)	$\nu_2^{ee'}$	−98	−112	−138	10.9
n(1 1 0)	$\nu_3^{bb'B}$	−90	−122	−182	−4.9
n(1 1 0)	$\nu_3^{bb'E}$	−106	−137	−195	−6.5
n(1 1 0)	ν_3^{deA}	−119	−150	−209	−11.0
R(1 1 0)	ν_0	−69	−66	−59	1.3
R(1 1 0)	ν_1^a	−72	−85	−111	4.1
R(1 1 0)	ν_1^b	−96	−112	−143	−1.0

continued on next page

Table A.1: continued from previous page

			ΔG_{ads}		^{31}P NMR
facet	site	(25°C)	(120°C)	(300°C)	
R(1 1 0)	ν_1^{c}	−106	−120	−147	−4.1
R(1 1 0)	ν_1^{d}	−72	−86	−112	−3.6
R(1 1 0)	ν_1^{e}	−83	−98	−125	3.7
R(1 1 0)	ν_2^{ab}	−75	−108	−170	−0.1
R(1 1 0)	ν_2^{bc}	−79	−113	−177	−5.4
R(1 1 0)	ν_2^{cB}	−72	−103	−159	−0.6
R(1 1 0)	ν_2^{ca}	−30	−63	−126	−1.7
R(1 1 0)	ν_2^{de}	−63	−97	−161	−1.1
R(1 1 0)	ν_2^{eB}	−129	−160	−219	−4.4
R(1 1 0)	ν_3^{abA}	4	−46	−139	−5.5
R(1 1 0)	ν_3^{bcB}	−118	−168	−263	−1.7
R(1 1 0)	ν_3^{bcE}	−54	−106	−203	−13.6
R(1 1 0)	ν_3^{cEF}	6	−45	−141	−11.4
R(1 1 0)	ν_3^{deA}	−85	−134	−227	−7.2
R(1 1 0)	ν_3^{deD}	0	−50	−144	1.9

Table A.2.: Average free enthalpy of adsorption of two orthophosphates in kJ mol^{-1} and ^{31}P NMR shift in ppm with respect to to 85 % H_3PO_4 for the (at 393 K) most stable system of every tested single adsorption site, labelled according to fig. 3.1a (p. 53).

		ΔG_{ads}			^{31}P NMR	
facet	sites	(25°C)	(120°C)	(300°C)	1st	2nd
(1 0 0)	$\nu_0+\nu_0$	−98	−74	−27	2.1	1.3
(1 0 0)	$\nu_1^a+\nu_0$	−93	−77	−47	−3.2	5.5
(1 0 0)	$\nu_1^b+\nu_0$	−105	−91	−64	−2.0	4.1
(1 0 0)	$\nu_1^c+\nu_0$	−84	−69	−41	−8.8	9.0
(1 0 0)	$\nu_1^a+\nu_1^a$	−82	−75	−62	5.5	−6.2
(1 0 0)	$\nu_1^a+\nu_1^b$	−102	−97	−86	11.6	−5.3
(1 0 0)	$\nu_1^a+\nu_1^c$	−90	−85	−75	−11.1	4.9
(1 0 0)	$\nu_2^{ab}+\nu_0$	−107	−101	−88	1.6	−6.0
(1 0 0)	$\nu_2^{bc}+\nu_0$	−119	−114	−102	1.5	12.1
(1 0 0)	$\nu_2^{ab}+\nu_1^a$	−116	−119	−126	12.7	−6.0
(1 0 0)	$\nu_2^{ab}+\nu_1^c$	−91	−95	−101	8.8	0.4
(1 0 0)	$\nu_2^{bc}+\nu_1^a$	−105	−110	−118	−2.6	−0.6
(1 0 0)	$\nu_2^{aa}+\nu_2^{bc}$	−102	−116	−141	1.7	8.3
(1 0 0)	$\nu_2^{aa}+\nu_2^{cb}$	−66	−80	−106	8.8	12.4
(1 0 0)	$\nu_3^{abc}+\nu_0$	−77	−80	−85	9.6	−4.7
(1 0 0)	$\nu_3^{abc}+\nu_1^a$	−91	−105	−132	−0.8	−4.9
(1 0 0)	$\nu_3^{aaC}+\nu_2^{bc}$	−84	−106	−148	9.4	6.4
(1 0 0)	$\nu_3^{aaC}+\nu_2^{cb}$	−76	−98	−139	1.0	7.7
n(1 1 0)	$\nu_0+\nu_0$	−101	−77	−33	4.7	0.1
n(1 1 0)	$\nu_1^b+\nu_0$	−112	−99	−72	10.6	8.1
n(1 1 0)	$\nu_1^e+\nu_0$	−89	−75	−48	−10.7	0.3
n(1 1 0)	$\nu_1^b+\nu_1^{e'}$	−111	−106	−96	−8.3	3.4
n(1 1 0)	$\nu_1^e+\nu_1^{e'}$	−83	−78	−69	12.3	9.6
n(1 1 0)	$\nu_2^{de}+\nu_0$	−102	−97	−88	−15.0	13.4
n(1 1 0)	$\nu_2^{bb'}+\nu_1^e$	−89	−92	−99	−13.7	9.3
n(1 1 0)	$\nu_2^{de}+\nu_1^b$	−101	−106	−115	−14.2	−0.2
n(1 1 0)	$\nu_2^{bb'}+\nu_2^{de}$	−114	−127	−153	−13.2	−10.9
n(1 1 0)	$\nu_3^{deA}+\nu_0$	−89	−93	−101	−9.4	13.3
n(1 1 0)	$\nu_3^{bb'E}+\nu_1^d$	−76	−90	−116	−6.4	0.0

continued on next page

Table A.2: continued from previous page

		$\Delta \mathrm{G}_{\mathrm{ads}}$			^{31}P NMR	
facet	sites	(25°C)	(120°C)	(300°C)	1st	2nd
n(1 1 0)	$\nu_3^{bb'E}+\nu_1^{e}$	−90	−103	−126	−9.2	5.2
n(1 1 0)	$\nu_3^{deA}+\nu_1^{b}$	−101	−114	−138	−17.6	7.7
n(1 1 0)	$\nu_3^{deA}+\nu_1^{b'}$	−97	−110	−134	−11.7	5.9
n(1 1 0)	$\nu_3^{deA}+\nu_1^{f}$	−95	−108	−132	−7.2	0.8
n(1 1 0)	$\nu_3^{bb'E}+\nu_2^{de}$	−106	−128	−169	−11.3	−9.4
n(1 1 0)	$\nu_3^{deA}+\nu_2^{bb'}$	−105	−127	−169	−13.9	−7.9
n(1 1 0)	$\nu_3^{deA}+\nu_3^{bb'E}$	−68	−100	−159	−10.8	−20.2
R(1 1 0)	$\nu_0+\nu_0$	−76	−61	−34	3.9	8.7
R(1 1 0)	$\nu_0+\nu_0$	−81	−68	−41	2.8	12.5
R(1 1 0)	$\nu_1^{c}+\nu_0$	−102	−97	−85	−2.0	−2.5
R(1 1 0)	$\nu_1^{b}+\nu_1^{c}$	−71	−75	−83	10.3	−7.5
R(1 1 0)	$\nu_1^{c}+\nu_1^{d}$	−97	−101	−109	16.5	−6.6
R(1 1 0)	$\nu_1^{c}+\nu_1^{e}$	−105	−110	−120	5.0	−3.0
R(1 1 0)	$\nu_2^{bc}+\nu_0$	−75	−81	−91	−16.8	7.2
R(1 1 0)	$\nu_2^{eB}+\nu_0$	−95	−99	−105	−4.0	1.1
R(1 1 0)	$\nu_2^{bc}+\nu_1^{d}$	−93	−108	−136	−15.2	−9.7
R(1 1 0)	$\nu_2^{bc}+\nu_1^{e}$	−95	−108	−133	−26.5	4.2
R(1 1 0)	$\nu_2^{eB}+\nu_1^{b}$	−83	−96	−121	−9.8	8.3
R(1 1 0)	$\nu_2^{eB}+\nu_1^{c}$	−100	−113	−139	−8.3	−4.7
R(1 1 0)	$\nu_2^{eB}+\nu_1^{d}$	−120	−132	−155	−5.1	2.2
R(1 1 0)	$\nu_2^{bc}+\nu_2^{eB}$	−56	−79	−121	−24.9	−9.4
R(1 1 0)	$\nu_3^{bcB}+\nu_0$	−89	−102	−126	1.1	3.6
R(1 1 0)	$\nu_3^{bcB}+\nu_1^{a}$	−96	−119	−161	0.8	3.3
R(1 1 0)	$\nu_3^{bcB}+\nu_1^{d}$	−99	−121	−164	−3.1	−1.9
R(1 1 0)	$\nu_3^{bcB}+\nu_1^{e}$	−87	−109	−152	−2.7	8.5
R(1 1 0)	$\nu_3^{bcE}+\nu_2^{eB}$	−36	−68	−127	−8.7	−15.4
R(1 1 0)	$\nu_3^{deA}+\nu_2^{eB}$	−66	−98	−157	−14.9	−12.3
R(1 1 0)	$\nu_3^{bcB}+\nu_3^{deA}$	−73	−113	−188	−11.7	−6.8
R(1 1 0)	$\nu_3^{bcE}+\nu_3^{deA}$	−36	−76	−153	−20.5	−16.5

Table A.3.: Free enthalpy of surface pyrophosphate formation in kJ mol^{-1} and ^{31}P NMR shift in ppm with respect to to 85 % H_3PO_4 for the (at 393 K) most stable system of every tested single adsorption site, labelled according to fig. 3.1a (p. 53). Free enthalpy reported with respect to a single orthophosphate in gas phase, in order to allow comparison with orthophosphate adsorption.

			ΔG_{ads}		^{31}P NMR	
facet	site	(25°C)	(120°C)	(300°C)	1st	2nd
(1 0 0)	ν_{00}	−49	−35	−8	−7.9	1.7
(1 0 0)	ν_{10}^{a}	−96	−89	−76	−19.2	−23.8
(1 0 0)	$\nu_{11}^{a,a}$	−95	−99	−107	−18.0	−10.2
(1 0 0)	$\nu_{11}^{a,b}$	−74	−78	−86	−13.9	−10.7
(1 0 0)	$\nu_{11}^{a,c}$	−79	−82	−87	−22.3	−10.7
(1 0 0)	ν_{20}^{ab}	−94	−97	−104	−9.7	−8.5
(1 0 0)	$\nu_{21}^{bc,a}$	−93	−106	−130	−11.0	−10.4
(1 0 0)	$\nu_{21}^{cb,a}$	−84	−97	−122	−4.0	−12.7
(1 0 0)	ν_{30}^{abc}	−49	−62	−88	−3.0	−16.4
n(1 1 0)	ν_{00}	−66	−52	−26	−5.8	−7.6
n(1 1 0)	ν_{10}^{a}	−63	−58	−49	−18.7	−11.1
n(1 1 0)	ν_{10}^{b}	−83	−78	−70	−8.7	−7.0
n(1 1 0)	ν_{10}^{d}	−78	−73	−62	−16.7	−1.8
n(1 1 0)	ν_{10}^{e}	−86	−81	−71	−10.3	−5.5
n(1 1 0)	$\nu_{11}^{a,d}$	−81	−85	−92	−10.4	−8.9
n(1 1 0)	$\nu_{11}^{b,b'}$	−94	−99	−109	−17.3	−21.2
n(1 1 0)	$\nu_{11}^{b,d}$	−51	−56	−64	−17.2	−11.5
n(1 1 0)	$\nu_{11}^{b',a}$	−69	−72	−79	−17.5	−14.6
n(1 1 0)	$\nu_{11}^{d,e}$	−83	−87	−95	−12.9	−18.4
n(1 1 0)	$\nu_{11}^{e,b}$	−97	−100	−106	−5.7	−13.5
n(1 1 0)	$\nu_{11}^{e,e'}$	−89	−94	−102	−13.6	−15.4
n(1 1 0)	ν_{20}^{ab}	−48	−51	−58	−6.7	−9.4
n(1 1 0)	$\nu_{20}^{bb'}$	−63	−68	−79	−23.6	−8.8
n(1 1 0)	ν_{20}^{de}	−82	−85	−92	−26.3	−9.0
n(1 1 0)	$\nu_{21}^{bb',e}$	−75	−88	−113	−32.6	−4.8
n(1 1 0)	$\nu_{21}^{de,a}$	−69	−81	−105	−33.0	−14.0
n(1 1 0)	$\nu_{21}^{de,b}$	−84	−97	−122	−25.2	−10.5
n(1 1 0)	$\nu_{22}^{bb',ee'}$	−46	−69	−112	2.7	−34.3

continued on next page

Table A.3: continued from previous page

		ΔG_{ads}			^{31}P NMR	
facet	site	(25°C)	(120°C)	(300°C)	1st	2nd
n(1 1 0)	$\nu_{22}^{de,ab}$	−51	−74	−117	−25.9	−35.9
n(1 1 0)	$\nu_{22}^{ee',bc}$	−35	−59	−103	−32.6	4.7
n(1 1 0)	$\nu_{30}^{bb'E}$	−20	−34	−59	−16.1	−5.0
n(1 1 0)	ν_{30}^{deA}	−43	−56	−83	−27.0	−5.6
n(1 1 0)	$\nu_{31}^{bb'E,e}$	−61	−84	−125	−21.6	−13.8
n(1 1 0)	$\nu_{31}^{deA,b}$	−57	−79	−121	−25.2	−8.6
n(1 1 0)	$\nu_{32}^{bb'B,ee'}$	−35	−67	−127	−18.3	8.3
n(1 1 0)	$\nu_{32}^{bb'E,ee'}$	4	−28	−89	−14.1	12.0
n(1 1 0)	$\nu_{32}^{deA,ab}$	−16	−47	−107	−21.7	−13.7
R(1 1 0)	ν_{00}	−84	−79	−69	−6.3	−13.6
R(1 1 0)	ν_{10}^{c}	−81	−85	−94	−14.0	−15.1
R(1 1 0)	$\nu_{11}^{b,c}$	−82	−96	−123	−15.8	−25.6
R(1 1 0)	$\nu_{11}^{c,d}$	−68	−82	−107	−30.0	−21.0
R(1 1 0)	$\nu_{11}^{c,e}$	−54	−67	−92	−17.4	−21.4
R(1 1 0)	ν_{20}^{bc}	−72	−86	−112	−35.3	−16.2
R(1 1 0)	ν_{20}^{eB}	−86	−99	−123	−10.6	−13.5
R(1 1 0)	$\nu_{21}^{bc,B}$	−9	−31	−72	−27.8	−1.3
R(1 1 0)	$\nu_{21}^{bc,E}$	19	−3	−44	−19.7	−14.3
R(1 1 0)	$\nu_{21}^{bc,F}$	−11	−33	−75	−20.5	−6.2
R(1 1 0)	$\nu_{21}^{bc,e}$	−64	−87	−131	−30.4	−12.2
R(1 1 0)	$\nu_{21}^{bc,e}$	−54	−77	−122	−35.5	−23.2
R(1 1 0)	$\nu_{21}^{eB,b}$	−83	−105	−147	−19.3	−12.5
R(1 1 0)	$\nu_{21}^{eB,c}$	−70	−92	−134	−13.7	−10.4
R(1 1 0)	$\nu_{21}^{eB,d}$	−78	−100	−142	−17.8	−19.0
R(1 1 0)	$\nu_{22}^{eB,bc}$	−84	−116	−175	−27.4	−15.9
R(1 1 0)	ν_{30}^{bcB}	−66	−87	−127	−17.2	−16.0
R(1 1 0)	ν_{30}^{bcE}	−58	−79	−119	−20.4	−10.4
R(1 1 0)	$\nu_{31}^{bcB,e}$	−85	−117	−176	−19.7	−18.7
R(1 1 0)	$\nu_{31}^{bcE,e}$	−38	−70	−130	−22.3	−21.3
R(1 1 0)	$\nu_{32}^{deA,ab}$	55	13	−66	−17.4	−15.6

A.3. Computational ^{31}P NMR reference

Experimental NMR data is by convention tabulated with respect to a well defined reference compound. These compounds are usually liquids as NMR spectroscopy is particularly suitable for molecules in solution. In case of ^{31}P NMR, 85 % H_3PO_4 is used as reference which defines the origin of the ^{31}P NMR scale by *convention.* GIPAW calculations calculate the chemical shift from the electronic structure by estimating the shielding tensor of the all-electron wave function. Its reference for the chemical shift is hence the "naked" nucleus, P^{31+}, which is experimentally impractical. Thus, in order to compare experiment and theory, a similar reference needs to be established. Computational modelling is much easier for static structures, like isolated molecules (gas phase) or crystalline solids. Here aluminium phosphate minerals were considered, since they should have a rather similar bonding as the discussed surface adsorptions. Three minerals were considered: berlinite, variscite and wavellite. All of them are characterized by X-ray crystallography and solid state MAS NMR.[58,169–171] Berlinite is the only water-free aluminium phosphate with a $AlPO_4$ stoichiometry and which has the same structure as β-quartz (both aluminium and phosphorus are in tetrahedral coordination). Variscite ($AlPO_4 \cdot H_2O$) is structurally similar to berlinite in that it consists of a three-dimensional mesh of alternating P and Al units; the Al units are in octahedral coordination due to hydration. Wavellite is an aluminium phosphate hydroxide, $Al_3(PO_4)_2(OH)_3 \cdot 5\,H_2O$, with a reduced relative phosphate content. Hence, wavellite might be structurally most similar to the discussed surface adsorption states. The phosphate sites in all three minerals are equivalent, producing only a single signal.

It is well known that NMR results are very sensitive to the local geometry. Three different geometries were tested starting from the experimental X-ray crystal structure. Since X-ray diffraction cannot resolve hydrogen positions, they were optimized at the DFT level. In case of Wavellite, free water molecules are in cavities, which were reported

Table A.4.: Tested ^{31}P NMR references at different levels of theory. The experimental MAS NMR *iso*-shifts are given with respect to 85 % H_3PO_4, while the computational results are given as calculated. The zero-shifts are the difference between experimental and computational value, i. e. the extrapolated values for 85 % H_3PO_4. **A**: GIPAW calculation at PBE-D3 level on the experimental X-ray structure; **B**: Full optimization and GIPAW calculation at PBE-D3 level; **C**: Full optimization and GIPAW calculation at TPSS-D3(BJ) level; **D**: Full optimization at TPSS-D3(BJ) level, GIPAW calculation at PBE-D3 level.

	^{31}P *iso*-shift, δ			zero-shift to 85 % H_3PO_4, δ			
	Berlinite	Variscite	Wavellite	Berlinite	Variscite	Wavellite	average
exp.	−25.3	−19.2	−11.2				
A	−331.4	−324.7	−318.6	−306.1	−305.5	−307.4	−306.4
B	−322.7	−313.7	−305.3	−297.4	−294.5	−294.1	−295.3
C	−325.1	−318.6	−314.1	−299.8	−299.4	−302.9	−300.7
D	−324.3	−318.7	−310.4	−299.0	−299.5	−299.2	−299.2

as 1/2 occupancies. These were also fully relaxed at the DFT level. The other two geometry tests were full optimizations at DFT level, one with the standard functional of this work, PBE-D3, and another test of TPSS-D3(BJ), a more accurate but also more costly method. Tab. A.4 summarizes the results for the GIPAW calculations. A limitation of the VASP code is the lack of pseudopotentials for any functional that is *not* PBE. This problem might have unpredictable effects when using a different functional to optimize the pseudized orbitals, since we then mix the description of core-electrons (described by PBE) and valence electrons (described by TPSS). Grimme's dispersion corrections were used in all tests.[89] Hence for the test with TPSS both cases are presented, one that uses TPSS-D3(BJ) only for geometry optimization, and another that is based on the electronic structure of TPSS (last two rows in tab. A.4).

Overall, the different structures lead to significant *iso*-shift variations, up to 13.3 ppm for wavellite and 11.1 on average. While the clear choice is to use the same functional that is also used throughout the work (PBE-D3) the structural effects provide information about the overall accuracy that can be expected. The deviations within the zero-shift should be negligible if any of the three minerals provided an accurate reference system. For PBE-D3 the deviations from berlinite to variscite and wavellite is notable with 3 ppm. PBE results based on the TPSS-D3(BJ) geometry seems to be most accurate, however, when using the electronic structure of TPSS, the zero-shift prediction of wavellite changes by 3.7 ppm compared to the PBE electronic structure on the TPSS geometry. Tab. A.4 summarizes the results for the GIPAW calculations. It is likely that zero-shift of the TPSS-D3(BJ) structure, that is 299.2 ppm, is the most accurate estimate produced here. However, since the surface structures are computed at PBE-D3 level, they should follow the same trend and be slightly shifted with respect to a more accurate result. Consequently, all GIPAW results of the adsorbed phosphate species were shifted by 295.3 ppm. Tab. A.4 summarizes the results for the GIPAW calculations.

A.4. Metadynamics: complex collective variables

Basic collective variables (CVs) are scalar values, like the coordination discussed in section 2.2.4, p. 46. For complex reactions like the condensation of two orthophosphates (cf. section 5.1, p. 93) requires as a bare minimum a CV which can produce the bridging P−O−P bond in $H_4P_2O_7$ and a different CV that follows the H_2O abstraction. Since it is not known which orthophosphate will provide the oxygen atom in the P−O−P bridge and which abstracts H_2O, one ideally uses CVs that do not decide how to break this symmetry. Similar to the coordinations definition as sum in eq. 2.7, bridging atoms within a group k connecting two structures represented by the atoms i and j can be counted with:

$$b = \sum_i \sum_j \sum_k s_{ik} \cdot s_{jk} \tag{A.15}$$

Here, s_{ik} and s_{jk} are switching functions as defined in eq. 2.8. The product of both elements is only 1 if both switching functions are describing a bond, i. e. atom k connects i and j, 0 otherwise.

As an alternative to counting the number of bonds, one can count the number atoms binding to another group of atoms. For the orthophosphate condensation, $2\,\mathrm{H_3PO_4}$ and $\mathrm{H_4P_2O_7}$ both produce eight P−O bonds, thus the simple coordination as defined in section 2.2.4) does not work. However, the number of oxygen atoms differs from 8 to 7. These numbers are accessible in PLUMED[114–116] with the keyword COORDINATIONNUMBER, introduced in ref. [172]. Instead of summing the whole coordination within one or between two groups of atoms, as described in section 2.2.4, the coordination number cn_i of each atom i in one group is calculated and accessible for further operation:

$$cn_i = \sum_j s(r_{ij}) \tag{A.16}$$

Any scalar quantities can be derived from these values, like e. g. the mean value or the quantity of cn_i smaller than a particular value, or within a range. In order to reach a smooth transition from 0 to 1 bond, a different switching function (based on a gaussian curve) was used:

$$s_g(x) = \begin{cases} \exp\left(-\frac{(x-d_0)^2}{2x_0^2}\right) , & \text{if} \quad x > d_0 \\ 1 , & \text{otherwise} \end{cases} \tag{A.17}$$

The smoothness is set with x_0 as gaussian width. This switching function vanishes rapidly with the difference $x-d_0$, whereas a switching function such as in eq. 2.8 produces a “tail end” which slowly converges with 0 (see also fig. 2.1, p. 46). Such a tail end is undesirable as we want to use the result from a switching function in cn_i and this does not permit convergence with 0 or 1 as a result, since if $cn_i = 1$, then $s(cn_i) > 0$.

For counting phosphorus binding oxygen atoms n, cn_i of 1 or larger were counted:

$$n \;=\; \sum_i 1 - s_g(cn_i)\;, \qquad\qquad x_0 = 0.2\,, \quad d_0 = 0.4 \tag{A.18}$$

Here, the quantity of cn_i of one or more was used in order to count the oxygen atoms that bind to phosphorus, irrespective of the number of P−O bonds. The parameters x_0 and d_0 were chosen such that $s_g(1) \approx 0$. To carry out a WTmetaD calculation for exploration of the orthophosphate condensation, CVs based on eq. A.15 and eq. A.18 were used. These variables are defined independently and satisfy the indistinguishability of the two phosphate centers.

Bibliography

[1] Jaap A. Bergwerff, Tom Visser, G. Leliveld, Brenda D. Rossenaar, Krijn P. de Jong, and Bert M. Weckhuysen. “Envisaging the Physicochemical Processes during the Preparation of Supported Catalysts: Raman Microscopy on the Impregnation of Mo onto Al_2O_3 Extrudates”. In: *Journal of the American Chemical Society* 126.44 (Nov. 1, 2004), pp. 14548–14556.

[2] H. Toulhoat and P. Raybaud. *Catalysis by Transition Metal Sulphides: From Molecular Theory to Industrial Application.* Paris: ED TECHNIP, Nov. 20, 2013.

[3] Tim Van Cleve, Devon Underhill, Mariana Veiga Rodrigues, Carsten Sievers, and J. Will Medlin. “Enhanced Hydrothermal Stability of γ-Al_2O_3 Catalyst Supports with Alkyl Phosphonate Coatings”. In: *Langmuir* 34.12 (Mar. 27, 2018), pp. 3619–3625.

[4] Ryuichiro Iwamoto and Jean Grimblot. “Influence of Phosphorus on the Properties of Alumina-Based Hydrotreating Catalysts”. In: *Advances in Catalysis.* Ed. by Werner O. Haag, Bruce C. Gates, and Helmut Knözinger. Vol. 44. Academic Press, Jan. 1, 1999, pp. 417–503.

[5] L. van Haandel, G. M. Bremmer, E. J. M. Hensen, and Th. Weber. “The effect of organic additives and phosphoric acid on sulfidation and activity of (Co)Mo/Al_2O_3 hydrodesulfurization catalysts”. In: *Journal of Catalysis* 351 (July 1, 2017), pp. 95–106.

[6] J. A. R. Van Veen, P. A. J. M. Hendriks, E. J. G. M. Romers, and R. R. Andrea. “Chemistry of phosphomolybdate adsorption on alumina surfaces. 1. The molybdate/alumina system”. In: *The Journal of Physical Chemistry* 94.13 (June 1, 1990), pp. 5275–5282.

[7] J. A. R. Van Veen, P. A. J. M. Hendriks, R. R. Andrea, E. J. G. M. Romers, and A. E. Wilson. “Chemistry of phosphomolybdate adsorption on alumina surfaces. 2. The molybdate/phosphated alumina and phosphomolybdate/alumina systems”. In: *The Journal of Physical Chemistry* 94.13 (June 1, 1990), pp. 5282–5285.

[8] Jaap A. Bergwerff, Leon G. A. van de Water, Tom Visser, Peter de Peinder, Bob R. G. Leliveld, Krijn P. de Jong, and Bert M. Weckhuysen. “Spatially Resolved Raman and UV-visible-NIR Spectroscopy on the Preparation of Supported Catalyst Bodies: Controlling the Formation of $H_2PMo_{11}CoO_{40}^{5-}$ Inside Al_2O_3 Pellets During Impregnation”. In: *Chemistry – A European Journal* 11.16 (2005), pp. 4591–4601.

[9] L. Catita, A. -A. Quoineaud, D. Espinat, C. Pichon, and O. Delpoux. “Application of Magnetic Resonance Imaging and Raman Imaging to study the impact of phosphorus in impregnation of hydrotreatment catalysts”. In: *Applied Catalysis A: General* 547 (Oct. 25, 2017), pp. 164–175.

[10] A. Griboval, P. Blanchard, E. Payen, M. Fournier, and J. L. Dubois. “Alumina supported HDS catalysts prepared by impregnation with new heteropolycompounds. Comparison with catalysts prepared by conventional Co–Mo–P coimpregnation”. In: *Catalysis Today* 45.1 (Oct. 19, 1998), pp. 277–283.

[11] A. Griboval, P. Blanchard, L. Gengembre, E. Payen, M. Fournier, J. L. Dubois, and J. R. Bernard. “Hydrotreatment Catalysts Prepared with Heteropolycompound: Characterisation of the Oxidic Precursors”. In: *Journal of Catalysis* 188.1 (Nov. 15, 1999), pp. 102–110.

[12] E. C. DeCanio, J. C. Edwards, T. R. Scalzo, D. A. Storm, and J. W. Bruno. “FT-IR and Solid-State NMR Investigation of Phosphorus Promoted Hydrotreating Catalyst Precursors.” In: *ChemInform* 23.11 (1992).

[13] Daniele Nicosia and Roel Prins. “The effect of phosphate and glycol on the sulfidation mechanism of CoMo/Al_2O_3 hydrotreating catalysts: an in situ QEXAFS study”. In: *Journal of Catalysis* 231.2 (Apr. 25, 2005), pp. 259–268.

[14] Ting-Ting Zheng, Zhong-Xi Sun, Xiao-Fang Yang, and Allan Holmgren. “Sorption of phosphate onto mesoporous γ-alumina studied with in-situ ATR-FTIR spectroscopy”. In: *Chemistry Central Journal* 6.1 (Apr. 3, 2012), p. 26.

[15] Mirella Del Nero, Catherine Galindo, Rémi Barillon, Eric Halter, and Benoit Madé. “Surface reactivity of α-Al_2O_3 and mechanisms of phosphate sorption: In situ ATR-FTIR spectroscopy and ζ potential studies”. In: *Journal of Colloid and Interface Science* 342.2 (Feb. 15, 2010), pp. 437–444.

[16] Chao Ren, Zhiwu Yu, Brian L. Phillips, Hongtao Wang, Junfeng Ji, Bingcai Pan, and Wei Li. “Molecular-scale investigation of fluoride sorption mechanism by nanosized hydroxyapatite using ^{19}F solid-state NMR spectroscopy”. In: *Journal of Colloid and Interface Science* 557 (Dec. 1, 2019), pp. 357–366.

[17] Wei Li, Andro-Marc Pierre-Louis, Kideok D. Kwon, James D. Kubicki, Daniel R. Strongin, and Brian L. Phillips. “Molecular level investigations of phosphate sorption on corundum (α-Al_2O_3) by ^{31}P solid state NMR, ATR-FTIR and quantum chemical calculation”. In: *Geochimica et Cosmochimica Acta* 107 (Apr. 15, 2013), pp. 252–266.

[18] Adrian Hühn, Dorothea Wisser, Manuel Corral Valero, Teddy Roy, Mickaël Rivallan, Leonor Catita, Anne Lesage, Carine Michel, and Pascal Raybaud. “Structural Characterization of Phosphate Species Adsorbed on γ-Alumina by Combining

DNP Surface Enhanced NMR Spectroscopy and DFT Calculations". In: *ACS Catalysis* (Aug. 25, 2021), pp. 11278–11292.

[19] C. Wolverton and K. C. Hass. "Phase stability and structure of spinel-based transition aluminas". In: *Physical Review B* 63.2 (Dec. 11, 2000), p. 024102.

[20] Gonzalo Gutiérrez, Adrian Taga, and Börje Johansson. "Theoretical structure determination of γ-Al_2O_3". In: *Physical Review B* 65.1 (Nov. 29, 2001), p. 012101.

[21] V. A. Ushakov and E. M. Moroz. "Structure of low-temperatureγ- andη-Al_2O_3". In: *Reaction Kinetics and Catalysis Letters* 24.1 (Mar. 1, 1984), pp. 113–118.

[22] Xénophon Krokidis, Pascal Raybaud, Anne-Elisabeth Gobichon, Bernadette Rebours, Patrick Euzen, and Hervé Toulhoat. "Theoretical Study of the Dehydration Process of Boehmite to γ-Alumina". In: *The Journal of Physical Chemistry B* 105.22 (June 1, 2001), pp. 5121–5130.

[23] Gianluca Paglia, Andrew L. Rohl, Craig E. Buckley, and Julian D. Gale. "Determination of the structure of gamma-alumina from interatomic potential and first-principles calculations: The requirement of significant numbers of nonspinel positions to achieve an accurate structural model". In: *Physical Review B* 71.22 (June 28, 2005), p. 224115.

[24] M. Digne, P. Sautet, P. Raybaud, P. Euzen, and H. Toulhoat. "Hydroxyl Groups on γ-Alumina Surfaces: A DFT Study". In: *Journal of Catalysis* 211.1 (Oct. 1, 2002), pp. 1–5.

[25] William M. Haynes, ed. *CRC Handbook of Chemistry and Physics, 97th Edition.* 97th edition. CRC Press, June 24, 2016.

[26] Donald E. Canfield, Erik Kristensen, and Bo Thamdrup. "The Phosphorus Cycle". In: *Advances in Marine Biology.* Ed. by Donald E. Canfield, Erik Kristensen, and Bo Thamdrup. Vol. 48. Aquatic Geomicrobiology. Academic Press, Jan. 1, 2005, pp. 419–440.

[27] Publications Office of the European Union. *Report on critical raw materials and the circular economy.* Nov. 5, 2018.

[28] Michael B. Geeson and Christopher C. Cummins. “Let’s Make White Phosphorus Obsolete”. In: *ACS Central Science* 6.6 (June 24, 2020), pp. 848–860.

[29] Edward H. Sandy, Ruth E. Blake, Sae Jung Chang, Yao Jun, and Chan Yu. “Oxygen isotope signature of UV degradation of glyphosate and phosphonoacetate: Tracing sources and cycling of phosphonates”. In: *Journal of Hazardous Materials* 260 (Sept. 15, 2013), pp. 947–954.

[30] Wiem Saidani, Aymen Wahbi, Badreddine Sellami, Mohamed Amine Helali, Abdelhafidh Khazri, Ezzeddine Mahmoudi, Soufiane Touil, Olivier Joubert, and Hamouda Beyrem. “Toxicity assessment of organophosphorus in Ruditapes decussatus via physiological, chemical and biochemical determination: A case study with the compounds γ-oximo- and γ-amino-phosphonates and phosphine oxides”. In: *Marine Pollution Bulletin* 169 (Aug. 1, 2021), p. 112556.

[31] Luis O. Torres-Dorante, Norbert Claassen, Bernd Steingrobe, and Hans-Werner Olfs. “Hydrolysis rates of inorganic polyphosphates in aqueous solution as well as in soils and effects on P availability”. In: *Journal of Plant Nutrition and Soil Science* 168.3 (2005), pp. 352–358.

[32] O. Agstam-Norlin, E. E. Lannergård, E. Rydin, M. N. Futter, and B. J. Huser. “A 25-year retrospective analysis of factors influencing success of aluminum treatment for lake restoration”. In: *Water Research* 200 (July 15, 2021), p. 117267.

[33] I. de Vicente, H. S. Jensen, and F. Ø. Andersen. “Factors affecting phosphate adsorption to aluminum in lake water: Implications for lake restoration”. In: *Science of The Total Environment* 389.1 (Jan. 15, 2008), pp. 29–36.

[34] Inmaculada de Vicente, Ping Huang, Frede Ø. Andersen, and Henning S. Jensen. “Phosphate Adsorption by Fresh and Aged Aluminum Hydroxide. Consequences

for Lake Restoration". In: *Environmental Science & Technology* 42.17 (Sept. 1, 2008), pp. 6650–6655.

[35] Brian J. Huser, Przemyslaw G. Bajer, Steve Kittelson, Scott Christenson, and Kevin Menken. "Changes to water quality and sediment phosphorus forms in a shallow, eutrophic lake after removal of common carp (Cyprinus carpio)". In: *Inland Waters* 0.0 (Mar. 5, 2021), pp. 1–14.

[36] Adam C. Martiny, Michael W. Lomas, Weiwei Fu, Philip W. Boyd, Yuh-ling L. Chen, Gregory A. Cutter, Michael J. Ellwood, Ken Furuya, Fuminori Hashihama, Jota Kanda, David M. Karl, Taketoshi Kodama, Qian P. Li, Jian Ma, Thierry Moutin, E. Malcolm S. Woodward, and J. Keith Moore. "Biogeochemical controls of surface ocean phosphate". In: *Science Advances* 5.8 (Aug. 1, 2019).

[37] Herbert Diskowski and Thomas Hofmann. "Phosphorus". In: *Ullmann's Encyclopedia of Industrial Chemistry.* American Cancer Society, 2000.

[38] Klaus Schrödter, Gerhard Bettermann, Thomas Staffel, Friedrich Wahl, Thomas Klein, and Thomas Hofmann. "Phosphoric Acid and Phosphates". In: *Ullmann's Encyclopedia of Industrial Chemistry.* American Cancer Society, 2008.

[39] F. H. Westheimer. "Why nature chose phosphates". In: *Science* 235.4793 (Mar. 6, 1987), pp. 1173–1178.

[40] Shina C. L. Kamerlin, Pankaz K. Sharma, Ram B. Prasad, and Arieh Warshel. "Why nature really chose phosphate". In: *Quarterly Reviews of Biophysics* 46.1 (Feb. 2013), pp. 1–132.

[41] Thomas Georgelin, Maguy Jaber, Houssein Bazzi, and Jean-François Lambert. "Formation of Activated Biomolecules by Condensation on Mineral Surfaces – A Comparison of Peptide Bond Formation and Phosphate Condensation". In: *Origins of Life and Evolution of Biospheres* 43.4 (Oct. 1, 2013), pp. 429–443.

[42] Terence P. Kee, David E. Bryant, Barry Herschy, Katie E. R. Marriott, Nichola E. Cosgrove, Matthew A. Pasek, Zachary D. Atlas, and Claire R. Cousins. "Phosphate Activation via Reduced Oxidation State Phosphorus (P). Mild Routes to Condensed-P Energy Currency Molecules". In: *Life* 3.3 (Sept. 2013), pp. 386–402.

[43] Erich Thilo. "Die kondensierten Phosphate". In: *Angewandte Chemie* 67.5 (1955), pp. 141–145.

[44] Joan Crowther and A. E. R. Westman. "The Hydrolysis of the Condensed Phosphates: III. Sodium Tetrametaphosphate and Sodium Tetraphosphate". In: *Canadian Journal of Chemistry* (Mar. 12, 1956).

[45] Joan Pedley Crowther and A. E. R. Westman. "The Hydrolysis of the Condensed Phosphates: I. Sodium Pyrophosphate and Sodium Triphosphate". In: *Canadian Journal of Chemistry* (Sept. 28, 1953).

[46] Philip George, Robert J. Witonsky, Mendel Trachtman, Clara Wu, William Dorwart, Linda Richman, William Richman, Fahd Shurayh, and Barry Lentz. ""Squiggle-H2O". An enquiry into the importance of solvation effects in phosphate ester and anhydride reactions". In: *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 223.1 (Nov. 3, 1970), pp. 1–15.

[47] Randy B. Stockbridge and Richard Wolfenden. "Enhancement of the Rate of Pyrophosphate Hydrolysis by Nonenzymatic Catalysts and by Inorganic Pyrophosphatase*". In: *Journal of Biological Chemistry* 286.21 (May 27, 2011), pp. 18538–18546.

[48] Michael E. Colvin, Earl Evleth, and Yamina Akacem. "Quantum chemical studies of pyrophosphate hydrolysis". In: *Journal of the American Chemical Society* 117.15 (Apr. 1, 1995), pp. 4357–4362.

[49] Ravi Tripathi, Rachel Glaves, and Dominik Marx. “The GTPase hGBP1 converts GTP to GMP in two steps via proton shuttle mechanisms”. In: *Chemical Science* 8.1 (Dec. 19, 2016), pp. 371–380.

[50] Rachel Glaves, Gerald Mathias, and Dominik Marx. “Mechanistic Insights into the Hydrolysis of a Nucleoside Triphosphate Model in Neutral and Acidic Solution”. In: *Journal of the American Chemical Society* 134.16 (Apr. 25, 2012), pp. 6995–7000.

[51] B. Ram Prasad, Nikolay V. Plotnikov, and Arieh Warshel. “Addressing Open Questions about Phosphate Hydrolysis Pathways by Careful Free Energy Mapping”. In: *The Journal of Physical Chemistry B* 117.1 (Jan. 10, 2013), pp. 153–163.

[52] Regina Luschtinetz, Augusto F. Oliveira, Johannes Frenzel, Jan-Ole Joswig, Gotthard Seifert, and Helio A. Duarte. “Adsorption of phosphonic and ethylphosphonic acid on aluminum oxide surfaces”. In: *Surface Science* 602.7 (Apr. 1, 2008), pp. 1347–1359.

[53] Sarah Blanck, Sophie Loehlé, Stephan N. Steinmann, and Carine Michel. “Adhesion of lubricant on aluminium through adsorption of additive head-groups on γ-alumina: A DFT study”. In: *Tribology International* 145 (May 1, 2020), p. 106140.

[54] I. Milošev, D. Zimerl, Ch Carriére, S. Zanna, A. Seyeux, J. Iskra, S. Stavber, F. Chiter, M. Poberžnik, D. Costa, A. Kokalj, and P. Marcus. “Editors’ Choice—The Effect of Anchor Group and Alkyl Backbone Chain on Performance of Organic Compounds as Corrosion Inhibitors for Aluminum Investigated Using an Integrative Experimental-Modeling Approach”. In: *Journal of The Electrochemical Society* 167.6 (Apr. 2020), p. 061509.

[55] K. C. Kumara Swamy and N. Satish Kumar. “New Features in Pentacoordinate Phosphorus Chemistry”. In: *Accounts of Chemical Research* 39.5 (May 1, 2006), pp. 324–333.

[56] Samuel Constant and Jerome Lacour. “New Trends in Hexacoordinated Phosphorus Chemistry”. In: *New Aspects in Phosphorus Chemistry V: -/-*. Ed. by Jean-Pierre Majoral. Topics in Current Chemistry. Berlin, Heidelberg: Springer, 2005, pp. 1–41.

[57] Norimasa Yoza, Nobuyuki Ueda, and Sachiko Nakashima. “pH-dependence of ^{31}P-NMR spectroscopic parameters of monofluorophosphate, phosphate, hypophosphate, phosphonate, phosphinate and their dimers and trimers”. In: *Fresenius' Journal of Analytical Chemistry* 348.10 (Oct. 1, 1994), pp. 633–638.

[58] William F. Bleam, Philip E. Pfeffer, and James S. Frye. “^{31}P solid-state nuclear magnetic resonance spectroscopy of aluminum phosphate minerals”. In: *Physics and Chemistry of Minerals* 16.5 (Apr. 1, 1989), pp. 455–464.

[59] E. R. H. Van Eck, A. P. M. Kentgens, H. Kraus, and R. Prins. “A Solid State Double Resonance NMR Investigation of Phosphorus-Impregnated γ-Al_2O_3”. In: *The Journal of Physical Chemistry* 99.43 (Oct. 1, 1995), pp. 16080–16086.

[60] Kan-Nian Hu, Hsiao-hua Yu, Timothy M. Swager, and Robert G. Griffin. “Dynamic Nuclear Polarization with Biradicals”. In: *Journal of the American Chemical Society* 126.35 (Sept. 1, 2004), pp. 10844–10845.

[61] Qing Zhe Ni, Eugenio Daviso, Thach V. Can, Evgeny Markhasin, Sudheer K. Jawla, Timothy M. Swager, Richard J. Temkin, Judith Herzfeld, and Robert G. Griffin. “High Frequency Dynamic Nuclear Polarization”. In: *Accounts of Chemical Research* 46.9 (Sept. 17, 2013), pp. 1933–1941.

[62] Anne Lesage, Moreno Lelli, David Gajan, Marc A. Caporini, Veronika Vitzthum, Pascal Miéville, Johan Alauzun, Arthur Roussey, Chloé Thieuleux, Ahmad Mehdi, Geoffrey Bodenhausen, Christophe Copéret, and Lyndon Emsley. “Surface enhanced NMR spectroscopy by dynamic nuclear polarization”. In: *Journal of the American Chemical Society* 132.44 (Nov. 10, 2010), pp. 15459–15461.

[63] Aaron J. Rossini, Alexandre Zagdoun, Moreno Lelli, Anne Lesage, Christophe Copéret, and Lyndon Emsley. “Dynamic Nuclear Polarization Surface Enhanced NMR Spectroscopy”. In: *Accounts of Chemical Research* 46.9 (Sept. 17, 2013), pp. 1942–1951.

[64] Maxence Valla, Aaron J. Rossini, Maxime Caillot, Céline Chizallet, Pascal Raybaud, Mathieu Digne, Alexandra Chaumonnot, Anne Lesage, Lyndon Emsley, Jeroen A. van Bokhoven, and Christophe Copéret. “Atomic Description of the Interface between Silica and Alumina in Aluminosilicates through Dynamic Nuclear Polarization Surface-Enhanced NMR Spectroscopy and First-Principles Calculations”. In: *Journal of the American Chemical Society* 137.33 (Aug. 26, 2015), pp. 10710–10719.

[65] Ana T. F. Batista, Dorothea Wisser, Thomas Pigeon, David Gajan, Fabrice Diehl, Mickael Rivallan, Leonor Catita, Anne-Sophie Gay, Anne Lesage, Céline Chizallet, and Pascal Raybaud. “Beyond γ-Al_2O_3 crystallite surfaces: The hidden features of edges revealed by solid-state ^1H NMR and DFT calculations”. In: *Journal of Catalysis* 378 (Oct. 1, 2019), pp. 140–143.

[66] Chris J. Pickard and Francesco Mauri. “All-electron magnetic response with pseudopotentials: NMR chemical shifts”. In: *Physical Review B* 63.24 (May 10, 2001), p. 245101.

[67] Thibault Charpentier. “The PAW/GIPAW approach for computing NMR parameters: A new dimension added to NMR study of solids”. In: *Solid State Nuclear Magnetic Resonance* 40.1 (July 1, 2011), pp. 1–20.

[68] Sharon E. Ashbrook, Marica Cutajar, Chris J. Pickard, Richard I. Walton, and Stephen Wimperis. “Structure and NMR assignment in calcined and as-synthesized forms of AlPO-14: a combined study by first-principles calculations and high-

resolution ^{27}Al–^{31}P MAS NMR correlation". In: *Physical Chemistry Chemical Physics* 10.37 (Oct. 9, 2008), pp. 5754–5764.

[69] Daniel M. Dawson, Valerie R. Seymour, and Sharon E. Ashbrook. "Effects of Extraframework Species on the Structure-Based Prediction of ^{31}P Isotropic Chemical Shifts of Aluminophosphates". In: *The Journal of Physical Chemistry C* 121.50 (Dec. 21, 2017), pp. 28065–28076.

[70] Frédérique Pourpoint, Adi Kolassiba, Christel Gervais, Thierry Azaïs, Laure Bonhomme-Coury, Christian Bonhomme, and Francesco Mauri. "First Principles Calculations of NMR Parameters in Biocompatible Materials Science: The Case Study of Calcium Phosphates, β- and γ-$Ca(PO_3)_2$. Combination with MAS-J Experiments". In: *Chemistry of Materials* 19.26 (Dec. 1, 2007), pp. 6367–6369.

[71] F. Pourpoint, C. Gervais, L. Bonhomme-Coury, T. Azaïs, C. Coelho, F. Mauri, B. Alonso, F. Babonneau, and C. Bonhomme. "Calcium Phosphates and Hydroxyapatite: Solid-State NMR Experiments and First-Principles Calculations". In: *Applied Magnetic Resonance* 32.4 (Dec. 1, 2007), pp. 435–457.

[72] Wei Li, Xionghan Feng, Yupeng Yan, Donald L. Sparks, and Brian L. Phillips. "Solid-State NMR Spectroscopic Study of Phosphate Sorption Mechanisms on Aluminum (Hydr)oxides". In: 47 (July 25, 2013), pp. 8308–8315.

[73] Frederik Tielens, Christel Gervais, Geraldine Deroy, Maguy Jaber, Lorenzo Stievano, Cristina Coelho Diogo, and Jean-François Lambert. "Characterization of Phosphate Species on Hydrated Anatase TiO_2 Surfaces". In: *Langmuir* 32.4 (Feb. 2, 2016), pp. 997–1008.

[74] Daniel M. Dawson and Sharon E. Ashbrook. "Investigating Relationships between the Crystal Structure and ^{31}P Isotropic Chemical Shifts in Calcined Aluminophosphates". In: *The Journal of Physical Chemistry C* 118.40 (Oct. 9, 2014), pp. 23285–23296.

[75] Daniel M. Dawson, John M. Griffin, Valerie R. Seymour, Paul S. Wheatley, Mahrez Amri, Teresa Kurkiewicz, Nathalie Guillou, Stephen Wimperis, Richard I. Walton, and Sharon E. Ashbrook. "A Multinuclear NMR Study of Six Forms of AlPO-34: Structure and Motional Broadening". In: *The Journal of Physical Chemistry C* 121.3 (Jan. 26, 2017), pp. 1781–1793.

[76] Geoffrey M. J. West. "Predicting phosphorus NMR shifts using neural networks". In: *Journal of Chemical Information and Computer Sciences* 33.4 (July 1, 1993), pp. 577–589.

[77] Ziyad Chaker, Mathieu Salanne, Jean-Marc Delaye, and Thibault Charpentier. "NMR shifts in aluminosilicate glasses via machine learning". In: *Physical Chemistry Chemical Physics* 21.39 (Oct. 9, 2019), pp. 21709–21725.

[78] Patrick Euzen, Pascal Raybaud, Xénophon Krokidis, Hervé Toulhoat, Jean-Luc Le Loarer, Jean-Pierre Jolivet, and Cédric Froidefond. "Alumina". In: *Handbook of Porous Solids*. Wiley Online Books. Apr. 25, 2008.

[79] Guido Busca. "The surface of transitional aluminas: A critical review". In: *Catalysis Today*. Acid-Base Catalysis Advanced Sciences and Spreading Applications to Solutions of Environmental, Resources and Energy Issues: ABC-7, 7th International Symposium on Acid-Base Catalysis, Tokyo, May 12-15, 2013 226 (May 1, 2014), pp. 2–13.

[80] Terry Threlfall. "Structural and Thermodynamic Explanations of Ostwald's Rule". In: *Organic Process Research & Development* 7.6 (Nov. 1, 2003), pp. 1017–1027.

[81] R. A. Van Santen. "The Ostwald step rule". In: *The Journal of Physical Chemistry* 88.24 (Nov. 1, 1984), pp. 5768–5769.

[82] Roel Prins. "On the structure of γ-Al_2O_3". In: *Journal of Catalysis* 392 (Dec. 1, 2020), pp. 336–346.

[83] B. C. Lippens and J. H. De Boer. “Study of phase transformations during calcination of aluminum hydroxides by selected area electron diffraction”. In: *Acta Crystallographica* 17.10 (1964), pp. 1312–1321.

[84] L. A. O’Dell, S. L. P. Savin, A. V. Chadwick, and M. E. Smith. “A ^{27}Al MAS NMR study of a sol–gel produced alumina: Identification of the NMR parameters of the θ-Al_2O_3 transition alumina phase”. In: *Solid State Nuclear Magnetic Resonance* 31.4 (July 1, 2007), pp. 169–173.

[85] Karl Sohlberg, Stephen J. Pennycook, and Sokrates T. Pantelides. “Hydrogen and the Structure of the Transition Aluminas”. In: *Journal of the American Chemical Society* 121.33 (Aug. 1, 1999), pp. 7493–7499.

[86] Ary R. Ferreira, Emine Küçükbenli, Alexandre A. Leitão, and Stefano de Gironcoli. “Ab initio ^{27}Al NMR chemical shifts and quadrupolar parameters for Al_2O_3 phases and their precursors”. In: *Physical Review B* 84.23 (Dec. 6, 2011), p. 235119.

[87] Gianluca Paglia, C. E. Buckley, and Andrew L. Rohl. “Comment on “Examination of Spinel and Nonspinel Structural Models for γ-Al_2O_3 by DFT and Rietveld Refinement Simulations””. In: *The Journal of Physical Chemistry B* 110.41 (Oct. 1, 2006), pp. 20721–20723.

[88] Zbigniew Łodziana, Nan-Yu Topsøe, and Jens K. Nørskov. “A negative surface energy for alumina”. In: *Nature Materials* 3.5 (May 2004), pp. 289–293.

[89] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. “A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu”. In: *The Journal of Chemical Physics* 132.15 (Apr. 16, 2010), p. 154104.

[90] Pascal Raybaud, Céline Chizallet, Christophe Mager-Maury, Mathieu Digne, Hervé Toulhoat, and Philippe Sautet. “From γ-alumina to supported platinum nanoclusters in reforming conditions: 10 years of DFT modeling and beyond”. In:

Journal of Catalysis. 50th Anniversary Special Issue 308 (Dec. 1, 2013), pp. 328–340.

[91] Kim Larmier, Céline Chizallet, and Pascal Raybaud. “Tuning the Metal–Support Interaction by Structural Recognition of Cobalt-Based Catalyst Precursors”. In: *Angewandte Chemie International Edition* 54.23 (2015), pp. 6824–6827.

[92] Raphael Wischert, Pierre Laurent, Christophe Copéret, Françoise Delbecq, and Philippe Sautet. “γ-Alumina: The Essential and Unexpected Role of Water for the Structure, Stability, and Reactivity of “Defect” Sites”. In: *Journal of the American Chemical Society* 134.35 (Sept. 5, 2012), pp. 14430–14449.

[93] Zhongfan Zhang, Long Li, and Judith C. Yang. “γ-Al_2O_3 thin film formation via oxidation of β-NiAl(110)”. In: *Acta Materialia* 59.15 (Sept. 1, 2011), pp. 5905–5916.

[94] Libor Kovarik, Mark Bowden, and János Szanyi. “High temperature transition aluminas in δ-Al_2O_3/θ-Al_2O_3 stability range: Review”. In: *Journal of Catalysis* 393 (Jan. 1, 2021), pp. 357–368.

[95] M Digne, P Sautet, P Raybaud, P Euzen, and H Toulhoat. “Use of DFT to achieve a rational understanding of acid–basic properties of γ-alumina surfaces”. In: *Journal of Catalysis* 226.1 (Aug. 15, 2004), pp. 54–68.

[96] Romain Réocreux, Tao Jiang, Marcella Iannuzzi, Carine Michel, and Philippe Sautet. “Structuration and Dynamics of Interfacial Liquid Water at Hydrated γ-Alumina Determined by ab Initio Molecular Simulations: Implications for Nanoparticle Stability”. In: *ACS Applied Nano Materials* 1.1 (Jan. 26, 2018), pp. 191–199.

[97] Matthieu Lagauche, Kim Larmier, Elsa Jolimaitre, Karin Barthelet, Céline Chizallet, Loïc Favergeon, and Michèle Pijolat. “Thermodynamic Characterization of the Hydroxyl Group on the γ-Alumina Surface by the Energy Distribution Function”. In: *The Journal of Physical Chemistry C* 121.31 (Aug. 10, 2017), pp. 16770–16782.

[98] Henry P. Pinto, R. M. Nieminen, and Simon D. Elliott. “Ab initio study of gamma-Al_2O_3 surfaces”. In: *Physical Review B* 70.12 (Sept. 3, 2004), p. 125402.

[99] Thomas Pigeon, Celine Chizallet, and Pascal Raybaud. “in preparation”. In: (2021).

[100] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. “Generalized Gradient Approximation Made Simple”. In: *Physical Review Letters* 77.18 (Oct. 28, 1996), pp. 3865–3868.

[101] P. E. Blöchl. “Projector augmented-wave method”. In: *Physical Review B* 50.24 (Dec. 15, 1994), pp. 17953–17979.

[102] G. Kresse and D. Joubert. “From ultrasoft pseudopotentials to the projector augmented-wave method”. In: *Physical Review B* 59.3 (Jan. 15, 1999), pp. 1758–1775.

[103] G. Kresse and J. Furthmüller. “Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set”. In: *Physical Review B* 54.16 (Oct. 15, 1996), pp. 11169–11186.

[104] Thomas D. Kühne, Marcella Iannuzzi, Mauro Del Ben, Vladimir V. Rybkin, Patrick Seewald, Frederick Stein, Teodoro Laino, Rustam Z. Khaliullin, Ole Schütt, Florian Schiffmann, Dorothea Golze, Jan Wilhelm, Sergey Chulkov, Mohammad Hossein Bani-Hashemian, Valéry Weber, Urban Borštnik, Mathieu Taillefumier, Alice Shoshana Jakobovits, Alfio Lazzaro, Hans Pabst, Tiziano Müller, Robert Schade, Manuel Guidon, Samuel Andermatt, Nico Holmberg, Gregory K. Schenter, Anna Hehn, Augustin Bussy, Fabian Belleflamme, Gloria Tabacchi, Andreas Glöß, Michael Lass, Iain Bethune, Christopher J. Mundy, Christian Plessl, Matt Watkins, Joost VandeVondele, Matthias Krack, and Jürg Hutter. “CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations”. In: *The Journal of Chemical Physics* 152.19 (May 21, 2020), p. 194103.

[105] S. Goedecker, M. Teter, and J. Hutter. “Separable dual-space Gaussian pseudopotentials”. In: *Physical Review B* 54.3 (July 15, 1996), pp. 1703–1710.

[106] Jonathan R. Yates, Chris J. Pickard, and Francesco Mauri. “Calculation of NMR chemical shifts for extended systems using ultrasoft pseudopotentials”. In: *Physical Review B* 76.2 (July 2, 2007), p. 024401.

[107] D. M. Deaven and K. M. Ho. “Molecular Geometry Optimization with a Genetic Algorithm”. In: *Physical Review Letters* 75.2 (July 10, 1995), pp. 288–291.

[108] Andriy O. Lyakhov, Artem R. Oganov, Harold T. Stokes, and Qiang Zhu. “New developments in evolutionary structure prediction algorithm USPEX”. In: *Computer Physics Communications* 184.4 (Apr. 1, 2013), pp. 1172–1182.

[109] Erik Bitzek, Pekka Koskinen, Franz Gähler, Michael Moseler, and Peter Gumbsch. “Structural Relaxation Made Simple”. In: *Physical Review Letters* 97.17 (Oct. 27, 2006), p. 170201.

[110] Graeme Henkelman and Hannes Jónsson. “Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points”. In: *The Journal of Chemical Physics* 113.22 (Dec. 8, 2000), pp. 9978–9985.

[111] Péter Pulay. “Convergence acceleration of iterative sequences. the case of scf iteration”. In: *Chemical Physics Letters* 73.2 (July 15, 1980), pp. 393–398.

[112] Donald A. McQuarrie. *Statistical Mechanics*. 1st edition. Sausalito, Calif: University Science Books, Feb. 1, 2000.

[113] Giovanni Bussi, Davide Donadio, and Michele Parrinello. “Canonical sampling through velocity rescaling”. In: *The Journal of Chemical Physics* 126.1 (Jan. 7, 2007), p. 014101.

[114] Massimiliano Bonomi, Davide Branduardi, Giovanni Bussi, Carlo Camilloni, Davide Provasi, Paolo Raiteri, Davide Donadio, Fabrizio Marinelli, Fabio Pietrucci, Ricardo A. Broglia, and Michele Parrinello. “PLUMED: A portable plugin for

free-energy calculations with molecular dynamics". In: *Computer Physics Communications* 180.10 (Oct. 1, 2009), pp. 1961–1972.

[115] Gareth A. Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. "PLUMED 2: New feathers for an old bird". In: *Computer Physics Communications* 185.2 (Feb. 2014), pp. 604–613. arXiv: `1310.0980`.

[116] Massimiliano Bonomi, Giovanni Bussi, Carlo Camilloni, Gareth A. Tribello, Pavel Banáš, Alessandro Barducci, Mattia Bernetti, Peter G. Bolhuis, Sandro Bottaro, Davide Branduardi, Riccardo Capelli, Paolo Carloni, Michele Ceriotti, Andrea Cesari, Haochuan Chen, Wei Chen, Francesco Colizzi, Sandip De, Marco De La Pierre, Davide Donadio, Viktor Drobot, Bernd Ensing, Andrew L. Ferguson, Marta Filizola, James S. Fraser, Haohao Fu, Piero Gasparotto, Francesco Luigi Gervasio, Federico Giberti, Alejandro Gil-Ley, Toni Giorgino, Gabriella T. Heller, Glen M. Hocky, Marcella Iannuzzi, Michele Invernizzi, Kim E. Jelfs, Alexander Jussupow, Evgeny Kirilin, Alessandro Laio, Vittorio Limongelli, Kresten Lindorff-Larsen, Thomas Löhr, Fabrizio Marinelli, Layla Martin-Samos, Matteo Masetti, Ralf Meyer, Angelos Michaelides, Carla Molteni, Tetsuya Morishita, Marco Nava, Cristina Paissoni, Elena Papaleo, Michele Parrinello, Jim Pfaendtner, Pablo Piaggi, GiovanniMaria Piccini, Adriana Pietropaolo, Fabio Pietrucci, Silvio Pipolo, Davide Provasi, David Quigley, Paolo Raiteri, Stefano Raniolo, Jakub Rydzewski, Matteo Salvalaglio, Gabriele Cesare Sosso, Vojtěch Spiwok, Jiří Šponer, David W. H. Swenson, Pratyush Tiwary, Omar Valsson, Michele Vendruscolo, Gregory A. Voth, Andrew White, and The PLUMED consortium. "Promoting transparency and reproducibility in enhanced molecular simulations". In: *Nature Methods* 16.8 (Aug. 2019), pp. 670–673.

[117] Alessandro Laio and Michele Parrinello. "Escaping free-energy minima". In: *Proceedings of the National Academy of Sciences* 99.20 (Oct. 1, 2002), pp. 12562–12566.

[118] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. “Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method”. In: *Physical Review Letters* 100.2 (Jan. 18, 2008), p. 020603.

[119] James F. Dama, Michele Parrinello, and Gregory A. Voth. “Well-Tempered Metadynamics Converges Asymptotically”. In: *Physical Review Letters* 112.24 (June 18, 2014), p. 240602.

[120] Michele Invernizzi and Michele Parrinello. “Rethinking Metadynamics: From Bias Potentials to Probability Distributions”. In: *The Journal of Physical Chemistry Letters* (Mar. 19, 2020).

[121] Raphael Wischert, Christophe Copéret, Françoise Delbecq, and Philippe Sautet. “Dinitrogen: a selective probe for tri-coordinate Al “defect” sites on alumina”. In: *Chemical Communications* 47.17 (May 7, 2011), pp. 4890–4892.

[122] Raphael Wischert, Pierre Florian, Christophe Copéret, Dominique Massiot, and Philippe Sautet. “Visibility of Al Surface Sites of γ-Alumina: A Combined Computational and Experimental Point of View”. In: *The Journal of Physical Chemistry C* 118.28 (July 17, 2014), pp. 15292–15299.

[123] Wei Li, Jian Feng, Kideok D. Kwon, James D. Kubicki, and Brian L. Phillips. “Surface Speciation of Phosphate on Boehmite (γ-AlOOH) Determined from NMR Spectroscopy”. In: *Langmuir* 26.7 (Apr. 6, 2010), pp. 4753–4761.

[124] Erich Thilo and Rudolf Sauer. “Zur Chemie der kondensierten Phosphate und Arsenate. XVII. Der Verlauf und die Produkte der Entwässerung der Monophosphorsäure H_3PO_4”. In: *Journal für Praktische Chemie* 4.5 (1957), pp. 324–348.

[125] Teddy Roy, Dorothea Wisser, Mickaël Rivallan, Manuel Corral Valero, Thibaut Corre, Olivier Delpoux, Gerhard D. Pirngruber, and Grégory Lefèvre. “Phosphate Adsorption on γ-Alumina: A Surface Complex Model Based on Surface Character-

ization and Zeta Potential Measurements". In: *The Journal of Physical Chemistry C* 125.20 (May 27, 2021), pp. 10909–10918.

[126] Abraham Yosipof, Oren E. Nahum, Assaf Y. Anderson, Hannah-Noa Barad, Arie Zaban, and Hanoch Senderowitz. "Data Mining and Machine Learning Tools for Combinatorial Material Science of All-Oxide Photovoltaic Cells". In: *Molecular Informatics* 34.6 (2015), pp. 367–379.

[127] Jonathan Schmidt, Mário R. G. Marques, Silvana Botti, and Miguel A. L. Marques. "Recent advances and applications of machine learning in solid-state materials science". In: *npj Computational Materials* 5.1 (Aug. 8, 2019), pp. 1–36.

[128] Bhavya Kailkhura, Brian Gallagher, Sookyung Kim, Anna Hiszpanski, and T. Yong-Jin Han. "Reliable and explainable machine-learning methods for accelerated material discovery". In: *npj Computational Materials* 5.1 (Nov. 14, 2019), pp. 1–9.

[129] Philomena Schlexer Lamoureux, Kirsten T. Winther, Jose Antonio Garrido Torres, Verena Streibel, Meng Zhao, Michal Bajdich, Frank Abild-Pedersen, and Thomas Bligaard. "Machine Learning for Computational Heterogeneous Catalysis". In: *ChemCatChem* 11.16 (2019), pp. 3581–3601.

[130] D. Müller, E. Jahn, G. Ladwig, and U. Haubenreisser. "High-resolution solid-state ^{27}Al and ^{31}P NMR: correlation between chemical shift and mean Al-O-P angle in AlPO4 polymorphs". In: *Chemical Physics Letters* 109.4 (Aug. 24, 1984), pp. 332–336.

[131] Viviane Campomar. "Rmn en rotation a l'angle magique du 31p. Application a l'etude de quelques phosphates". These de doctorat. Paris 6, Jan. 1, 1990.

[132] Eric Jonas and Stefan Kuhn. "Rapid prediction of NMR spectral properties with quantified uncertainty". In: *Journal of Cheminformatics* 11.1 (Aug. 6, 2019), p. 50.

[133] W. Bremser. "Hose — a novel substructure code". In: *Analytica Chimica Acta* 103.4 (Dec. 15, 1978), pp. 355–365.

[134] Isao Tanaka, ed. *Nanoinformatics*. Springer Singapore, 2018.

[135] Felix Faber, Alexander Lindmaa, O. Anatole von Lilienfeld, and Rickard Armiento. “Crystal structure representations for machine learning models of formation energies”. In: *International Journal of Quantum Chemistry* 115.16 (2015), pp. 1094–1101.

[136] Lauri Himanen, Marc O. J. Jäger, Eiaki V. Morooka, Filippo Federici Canova, Yashasvi S. Ranawat, David Z. Gao, Patrick Rinke, and Adam S. Foster. “DScribe: Library of descriptors for machine learning in materials science”. In: 1 (Oct. 2, 2019).

[137] Katja Hansen, Franziska Biegler, Raghunathan Ramakrishnan, Wiktor Pronobis, O. Anatole von Lilienfeld, Klaus-Robert Müller, and Alexandre Tkatchenko. “Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space”. In: *The Journal of Physical Chemistry Letters* 6.12 (June 18, 2015), pp. 2326–2331.

[138] Albert P. Bartók, Risi Kondor, and Gábor Csányi. “On representing chemical environments”. In: *Physical Review B* 87.18 (May 28, 2013), p. 184115.

[139] Jörg Behler. “Atom-centered symmetry functions for constructing high-dimensional neural network potentials”. In: *The Journal of Chemical Physics* 134.7 (Feb. 21, 2011), p. 074106.

[140] Nongnuch Artrith, Alexander Urban, and Gerbrand Ceder. “Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species”. In: *Physical Review B* 96.1 (July 21, 2017), p. 014112.

[141] Donald E. Hilt and Donald W. Seegrist. *Ridge, a computer program for calculating ridge regression estimates.* 1977.

[142] Simon Haykin. *Neural Networks and Learning Machines.* 3rd edition. New York: Pearson, Nov. 18, 2008.

[143] Tianqi Chen and Carlos Guestrin. “XGBoost: A Scalable Tree Boosting System”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Aug. 13, 2016), pp. 785–794. arXiv: `1603.02754`.

[144] Frank Campbell and Gary Smith. “A Critical Analysis of Ridge Regression”. In: (Feb. 1, 1975).

[145] Marvin Gruber. *Improving Efficiency by Shrinkage: The James–Stein and Ridge Regression Estimators*. CRC Press, Feb. 26, 1998.

[146] Haoyan Huo and Matthias Rupp. “Unified Representation of Molecules and Crystals for Machine Learning”. In: *arXiv:1704.06439v3 [physics.chem-ph]* (Jan. 2, 2018). arXiv: `1704.06439`.

[147] E. Schrödinger. “An Undulatory Theory of the Mechanics of Atoms and Molecules”. In: *Physical Review* 28.6 (Dec. 1, 1926), pp. 1049–1070.

[148] Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Reprint Edition. Mineola, N.Y: Dover Publications, July 2, 1996.

[149] M. Born and R. Oppenheimer. “Zur Quantentheorie der Molekeln”. In: *Annalen der Physik* 389.20 (1927), pp. 457–484.

[150] D. R. Hartree. “The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and Methods”. In: *Proceedings of the Cambridge Philosophical Society* 24.1 (1928), pp. 89–110.

[151] J. C. Slater. “The Theory of Complex Spectra”. In: *Physical Review* 34.10 (Nov. 15, 1929), pp. 1293–1322.

[152] Chr. Møller and M. S. Plesset. “Note on an Approximation Treatment for Many-Electron Systems”. In: *Physical Review* 46.7 (Oct. 1, 1934), pp. 618–622.

[153] C. C. J. Roothaan. "New Developments in Molecular Orbital Theory". In: *Reviews of Modern Physics* 23.2 (Apr. 1, 1951), pp. 69–89.

[154] P. Hohenberg and W. Kohn. "Inhomogeneous Electron Gas". In: *Physical Review* 136.3 (Nov. 9, 1964), B864–B871.

[155] Louis Garrigue. "Hohenberg-Kohn theorems for interactions, spin and temperature". In: *Journal of Statistical Physics* 177.3 (Nov. 2019), pp. 415–437. arXiv: `1906.03191`.

[156] R. Gaudoin and K. Burke. "Lack of Hohenberg-Kohn Theorem for Excited States". In: *Physical Review Letters* 93.17 (Oct. 18, 2004), p. 173001.

[157] W. Kohn and L. J. Sham. "Self-Consistent Equations Including Exchange and Correlation Effects". In: *Physical Review* 140.4 (Nov. 15, 1965), A1133–A1138.

[158] L. H. Thomas. "The calculation of atomic fields". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 23.5 (Jan. 1927), pp. 542–548.

[159] Elliott H Lieb and Barry Simon. "The Thomas-Fermi theory of atoms, molecules and solids". In: *Advances in Mathematics* 23.1 (Jan. 1, 1977), pp. 22–116.

[160] C. F. v. Weizsäcker. "Zur Theorie der Kernmassen". In: *Zeitschrift für Physik* 96.7 (July 1, 1935), pp. 431–458.

[161] John C. Snyder, Matthias Rupp, Katja Hansen, Klaus-Robert Müller, and Kieron Burke. "Finding Density Functionals with Machine Learning". In: *Physical Review Letters* 108.25 (June 19, 2012), p. 253002.

[162] Kun Yao and John Parkhill. "Kinetic Energy of Hydrocarbons as a Function of Electron Density and Convolutional Neural Networks". In: *Journal of Chemical Theory and Computation* 12.3 (Mar. 8, 2016), pp. 1139–1147.

[163] Ralf Meyer, Manuel Weichselbaum, and Andreas W. Hauser. "Machine Learning Approaches toward Orbital-free Density Functional Theory: Simultaneous Training on the Kinetic Energy Density Functional and Its Functional Derivative". In: *Journal of Chemical Theory and Computation* 16.9 (Sept. 8, 2020), pp. 5685–5694.

[164] John P. Perdew and Karla Schmidt. "Jacob's ladder of density functional approximations for the exchange-correlation energy". In: *AIP Conference Proceedings* 577.1 (July 6, 2001), pp. 1–20.

[165] Stephan N. Steinmann and Clemence Corminboeuf. "Comprehensive Benchmarking of a Density-Dependent Dispersion Correction". In: *Journal of Chemical Theory and Computation* 7.11 (Nov. 8, 2011), pp. 3567–3577.

[166] Jiří Klimeš, David R. Bowler, and Angelos Michaelides. "Chemical accuracy for the van der Waals density functional". In: *Journal of Physics: Condensed Matter* 22.2 (2010), p. 022201.

[167] D. R. Hamann, M. Schlüter, and C. Chiang. "Norm-Conserving Pseudopotentials". In: *Physical Review Letters* 43.20 (Nov. 12, 1979), pp. 1494–1497.

[168] David Vanderbilt. "Soft self-consistent pseudopotentials in a generalized eigenvalue formalism". In: *Physical Review B* 41.11 (Apr. 15, 1990), pp. 7892–7895.

[169] Aline Goiffon, Jean-Claude Jumas, Maurice Maurin, and Etienne Philippot. "Etude comparée à diverses températures (173, 293 et 373°K) des structures de type quartz α des phases $M^{III}X^{V}O_4$ (M^{III} = Al, Ga et X^{V} = P, As)". In: *Journal of Solid State Chemistry* 61.3 (Mar. 1, 1986), pp. 384–396.

[170] R. Kniep, D. Mootz, and A. Vegas. "Variscite". In: *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry* 33.1 (Jan. 15, 1977), pp. 263–265.

[171] Takaharu Araki and Tibor Zoltai. "The crystal structure of wavellite". In: *Zeitschrift für Kristallographie,* 127 (Dec. 14, 1967), pp. 21–33.

[172] Andrew D. White and Gregory A. Voth. “Efficient and Minimal Method to Bias Molecular Simulations with Experimental Data”. In: *Journal of Chemical Theory and Computation* 10.8 (Aug. 12, 2014), pp. 3023–3030.

Theoretical Studies on the Adsorption of Phosphoric Acid on γ-Al_2O_3 at the Solid-Liquid Interface

Abstract

Phosphate adsorption on oxides is of significant interest in heterogeneous catalysis. The CoMoS/γ-Al_2O_3 hydrodesulfuration catalyst which is used at the scale of petroleum refinement, uses phosphate additives for preparation. Its preparation follows the stages of wet incipient impregnation, drying and activation by thermal treatment. Understanding the chemistry involved in these processes requires models of the oxide-water and oxide-air interfaces at different temperatures, which are also of interest in different fields. Oxide surfaces like γ-Al_2O_3 are by themselves very complex, which renders the characterization of surface phosphate species even more intricate.

Yet by today, improved NMR spectroscopic methods complemented with proven computational models allowed us now to mold an atomistic model of phosphate speciation. The present work starts with the creation of a systematic database by exploring more than 1000 geometries at different adsorption modes, adsorption sites, degrees of coverage and an optimized hydrogen bond environment for two polyphosphate species. For the most stable results magnetic shielding was calculated at DFT level. Good agreement was achieved with ^{31}P NMR experiments at drying conditions. Liquid conditions were simulated using metadynamics and showed a different set of most stable species. The explored reaction energies and kinetic barriers consistently show that elevated temperatures are needed to form the experimentally observed species. A ^{31}P NMR prediction model based on machine learning of chemical shifts from the local structure was developed to access chemical shifts in dynamic simulations.

UNIVERSITÉ DE LYON
ENS DE LYON