UC Santa Barbara

UC Santa Barbara Electronic Theses and Dissertations

Title

Modeling the preparation and reactivity of atomically dispersed catalysts on amorphous supports: method development and applications

Permalink

https://escholarship.org/uc/item/4888t7bz

Author

Khan, Salman Ahmad

Publication Date

2021

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

University of California
Santa Barbara

Modeling the preparation and reactivity of atomically dispersed catalysts on amorphous supports: method development and applications

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Chemical Engineering

by

Salman Ahmad Khan

Committee in charge:

Professor Baron G. Peters, Co-Chair
Professor Susannah L. Scott, Co-Chair
Professor Phillip Christopher
Professor Linda Petzold

September 2021

The Dissertation of Salman Ahmad Khan is approved.

Professor Phillip Christopher

Professor Linda Petzold

Professor Susannah L. Scott, Committee Co-Chair

Professor Baron G. Peters, Committee Co-Chair

August 2021

Acknowledgements

I am grateful to my advisors, Professor Baron Peters and Professor Susannah Scott, for being excellent mentors and for providing me the opportunity to work on a wide range of projects. I have greatly benefited from the breadth and depth of Baron's expertise in different areas of catalysis, statistical physics, rare events, and probability theory. He has shaped much of my computational thinking and approach to problems as a theorist. I am thankful to him for providing encouragement and the freedom to explore topics outside of my main project and for being receptive to many of my ideas. I thank Susannah for her guidance and feedback in keeping my modeling efforts closely connected to experiments and her approachability despite her extremely busy schedule. Many of the projects presented in this thesis were conceived as a result of attempting to answer questions posed by her.

I am grateful to my committee members: Professor Phillip Christopher and Professor Linda Petzold, for their insightful comments and their evaluation of my work through the years.

I thank Christian Leitold, Tanmoy Sanyal, and Kartik Kamat for discussions on statistical physics and thermodynamics and Xijun Wang for helping me get started with electronic structure calculations. I am grateful to the members of the Peters group: Craig Vandervelden, Armin Shayesteh Zadeh, Ziqiu Chen, Vivekananda Bal, Vikram Khanna, and Ryan Yappert, who were kind enough on many occasions to discuss my projects and provide critical feedback. I also thank members of the Scott and Shell groups at UCSB and our collaborators at KU.

I will be remiss if I do not thank, Rohit Reddy, Arnab Pal, Pradeep Shekhar, Hyunjin Moon, Sravan Chalasani, and Sameer Kalavalapalli, who made my time at UCSB memorable.

I will forever be grateful to my parents for their motivation and unconditional support throughout my life. I am greatly indebted to my mother for instilling in me a love for mathematics and science from a young age and for being a great teacher throughout my life.

Curriculum Vitæ

Salman Ahmad Khan

Education

2021	Ph.D. in Chemical Engineering, University of California, Santa Barbara.
2016	M.tech in Chemical Engineering, Indian Institute of Technology, Kanpur.
2015	B.tech in Chemical Engineering, Indian Institute of Technology, Kanpur.

Publications

- **Salman A. Khan**, Pubudu Wimalasiri, Sahan Godahewa, Ward Thompson, Susannah L. Scott, and Baron Peters, *Grafting $TiCl_4$ onto amorphous silica: modeling effects of silanol heterogeneity*, manuscript in preparation.
- **Salman A. Khan**, Eric W. Deguns, Rosemary E. White, Ziyad A. Taha, Baron Peters, and Susannah L. Scott, *Quantifying abundances of grafted $[\equiv(SiO)_{4\text{-}x}TiCl_x]$ sites on amorphous silica*, manuscript in preparation.
- **Salman A. Khan**, Susannah L. Scott, and Baron Peters, *On the various computational models of silica: a statistical analysis of structural differences* , manuscript in preparation.
- Craig A. Vandervelden, **Salman A. Khan**, and Baron Peters, *Importance learning estimator for the site-averaged turnover frequency of a disordered solid catalyst*, **J. Chem. Phys.**, 2020, 150 (24), 244120.
- **Salman A. Khan**, Bradley M. Dickson, Baron Peters, *How fluxional reactants limit the accuracy/efficiency of infrequent metadynamics*, **J. Chem. Phys.**, 2020, 153 (5), 054125.
- Craig A. Vandervelden,* **Salman A. Khan**,* Susannah L. Scott, and Baron Peters, *Site-averaged kinetics for catalysts on amorphous supports: an importance learning algorithm*, **React. Chem. Eng.**, 2020, 5 (1), 77-86.
- **Salman A. Khan**,* Craig A. Vandervelden,* Susannah L. Scott, and Baron Peters, *Grafting metal complexes onto amorphous supports: from elementary steps to catalyst site populations via kernel regression*, **React. Chem. Eng.**, 2020, 5 (1), 66-76.

* indicates equal contribution

Presentations

- CECAM Activated Events Online Conference, **2021** (**poster presentation**).
- AIChE Midwest Regional Conference, Chicago, IL, **2020** (**oral presentation**).
- Seminar at the University of Kansas, KS, **2019** (**oral presentation**).
- AIChE Annual Meeting, Orlando, FL, **2019** (**oral presentation**).
- Summer school, Rare Events: Applications, Computation, and Theory, Indian Institute of Science, Bangalore, IN, **2019** (**participant**).
- Midwest Thermodynamics and Statistical Mechanics Conference, Urbana-Champaign, IL, **2019 (participant)**.
- Machine Learning in Science and Engineering, Atlanta, GA, **2019 (oral presentation)**.
- Spring Symposium of the Catalysis Club of Chicago, Chicago, IL, **2019** (**poster presentation**).
- 11th Annual Amgen-Clorox Graduate Student Symposium, UCSB, Santa Barbara, CA, **2018** (**poster presentation**).

Honors and Awards

- **Travel Award**, Machine Learning in Science and Engineering Symposium, Atlanta (2019)
- **Academic Excellence** award 2015-16, IIT Kanpur, awarded to 2 students in the dual degree Chemical Engineering class (2016)
- **Academic Excellence** award 2013-14, IIT Kanpur, awarded to top 7% students in the institute (2015)
- Intern project at Whirlpool Corporation awarded the **2nd** prize out of 21 projects (2014)
- Qualified written test and interview to receive the **Kishore Vaigyanik Protsahan Yojana (KVPY) scholarship**, offered by the Department Of Science and Technology India to around 500 students all over the country to pursue undergraduate studies in Physics, Chemistry, or Mathematics (2011)

Abstract

Modeling the preparation and reactivity of atomically dispersed catalysts on amorphous supports: method development and applications

by

Salman Ahmad Khan

Several industrially important catalysts are single metals dispersed on amorphous supports. For example, Cr dispersed on amorphous SiO_2 is used to catalyze ethene polymerization and W dispersed on amorphous Al_2O_3 and SiO_2 is used to catalyze olefin metathesis. Despite their extensive use in the industry, these catalysts have largely been intractable to both, experimental and modeling investigations. In particular, they present the following challenges: (i) an unknown quenched disordered structure of the amorphous support, (ii) metal atoms attach to various surface grafting sites with different rates and have different activation and catalytic reaction kinetics, and (iii) only a small fraction of the sites are active. These challenges particularly render *ab initio* computational tools, routinely applied to study homogeneous and ordered heterogeneous catalysts, inefficient and impractical. The overarching goal of this thesis is developing computational tools to efficiently model the synthesis and reactivity of atomically dispersed catalysts on amorphous supports.

Atomically dispersed amorphous catalysts are synthesized by grafting organometallic complexes onto amorphous supports. We develop a machine learning (ML) parametrized population balance model to predict the evolving population of active sites during catalyst synthesis. We apply the population balance modeling framework to model the grafting of $TiCl_4$ onto amorphous silica. Equilibrium predictions of the model agree with experimentally determined populations of grafted Ti sites. Additionally, we develop an

Importance Learning (IL) algorithm to efficiently calculate the site-averaged activation barrier for amorphous catalysts. IL uses a combination of ML and importance sampling to discover rare and active catalytic sites which dominate kinetics.

Different studies have generated atomistic models of amorphous silica using different simulation protocols. Most studies have claimed that their models are representative of real silicas. Using statistical hypothesis testing, we show that different protocols lead to models with different structural features. We discuss the effects of these structural variations on the grafted catalyst and outline experimental metrics that can be used to validate models in future studies.

Finally, we present a site balance algebra to quantifying amounts of different species generated in grafting experiments. We quantify the amounts of [$\equiv$SiOTiCl$_3$] and [$(\equiv$SiO$)_2$TiCl$_2$] sites obtained on grafting TiCl$_4$ onto amorphous silica.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The North American Catalysis Society estimates that catalysis, directly or indirectly, contributes to about 35% of the world's GDP.[1] Catalysts are critical in the production of materials and chemicals important in several industries including food,[2] automotive,[3] energy,[4,5] textile,[6] etc. Understanding catalytic mechanisms and the effect of catalyst structure and composition on properties (activity, selectivity, stability, etc.) is critical in the development of efficient catalysts. One of the first systematic investigations of catalysis, describing the conversion of alcohols to ethylene over silica, is from 1796.[7] However, catalysis remained a mystery for much of the 18^{th} and 19^{th} centuries. It was only by the early/mid-20^{th} century that some rational theories of catalysis were proposed. For example, Langmuir successfully described the mechanism of catalysis on metals through adsorption of reactants on surfaces in the 1910s.[8–10] H.S. Taylor postulated the role of surface metal atoms in catalyzing reactions in 1925.[11]

Still, without much understanding of the atomic and electronic structure of materials, most of the progress was based on trial-and-error.[12] It was only after the advent of

spectroscopic methods, like UV-Vis,[13] Raman,[14] Mössbauer,[15] IR[16] etc., that catalysis could be investigated at an atomic scale. Spectroscopic methods enabled the investigation of catalyst structure/composition effects on properties. Spectroscopic methods have been successful in elucidating reaction mechanisms for several catalysts and in turn improving their activity and selectivity.[17]

Catalysis involves multiple processes operating at different length and timescales, like adsorption, diffusion, reactions, restructuring of the catalyst sites while in operation, etc. Spectroscopic studies provide space and time averaged spectra and deconvoluting the effects of different processes is not always straightforward. Computational models can supplement experiments by providing tools to systematically investigate the effect of different processes on catalyst properties and mechanisms.[18] They can also be used to calculate experimental observables, like activation barriers and turnover frequencies, and thus provide a framework to test mechanistic hypotheses.[18,19]

The development of quantum mechanics (QM) in the early 20$^{\text{th}}$ century provided,[20] for the first time, an *ab initio* framework to model reactions of molecules. However, it was only towards the end of the 20$^{\text{th}}$ century that scalable and efficient methods to solve the Schrödinger Equation for multi-electron systems were developed.[21] These developments made it possible to investigate system sizes relevant to catalysis. Density functional theory (DFT) is a method to approximate the solution to the Schrödinger equation and it offers reasonable accuracy with a modest computational cost.[22,23] Hence, it has become one of the most commonly used QM methods in catalysis.[24] DFT can be used to calculate the potential energy surface (PES) of a system (potential energy as a function of atomic coordinates). In principle, it is possible to sample the PES using molecular dynamics or Monte Carlo methods.[25] The trajectories obtained from these methods can be used to calculate abundances of different reaction intermediates and rates of elementary reactions using statistical mechanics.[26] However, in practice

it is infeasible to sample the PES for many systems because of the prohibitive cost of even the fastest QM methods. Instead, the PES is estimated by locating positions of minima (corresponding to reactants, intermediates, and products) and saddle points (corresponding to transition states connecting minima) on it.[27] Following this, the PES is approximated using a harmonic approximation (other approximations have also been used) around these critical points. Free energies of different species and transition states are calculated assuming a Boltzmann distribution of energies. A more detailed description of methods to explore the PES and calculate free energies can be found elsewhere.[27]

Free energy differences can be used to estimate abundances of different intermediates in a catalytic pathway. Free energy barriers can be used to calculate rate constants of elementary reactions using transition state theory (TST).[19] Rate constants can be used as an input in a microkinetic model[28] or a kinetic Monte Carlo simulation[29] to simulate long time dynamics of catalytic systems and calculate experimental observables like turnover frequencies and activation barriers, thus providing a direct link between structure and properties of catalysts. Computational studies have been instrumental in not only supplementing experimental methods to understand catalytic mechanisms, but also in discovering new catalytic materials. For example, high throughput computations were used to discovery a BiPt alloy active for the hydrogen evolution reaction (HER) by screening 700 binary alloys. The discovered material was later shown to be more active than Pt (the most active metal for HER).[30]

However, most computational advancements have been limited to homogeneous catalysts[31–33] and ordered heterogeneous catalysts,[31,32] like zeolites,[34,35] metal oxide nanoparticles,[36,37] and metal nanoparticles.[38,39] These catalysts have in common the advantage of possessing only a limited variety of sites. On the other hand, amorphous catalysts present a quenched distribution of sites and still remain elusive on several fronts.[40,41] The presence of multiple active sites with different structural environments is challeng-

ing to model using expensive *ab initio* methods. Moreover, they lack long-range order and hence, their spectra do not have sharp features, unlike their ordered counterparts.[42,43] This complicates their characterization and consequently, determining the exact structure of amorphous catalysts is not straightforward.

Atomically dispersed metals on amorphous supports are a class of amorphous catalysts with several applications in the industry. For example, Cr and Ti dispersed on amorphous SiO_2 are active for ethene polymerization[44] and ethene epoxidation,[45] respectively and W dispersed on amorphous Al_2O_3 and SiO_2 is active for metathesis.[46] In principle, these catalysts offer maximum site dispersion and the most efficient use of the active material. However, many of these catalysts, despite being extensively used in the industry, are still poorly understood. For example, the Phillips catalyst (Cr dispsered on SiO_2), discovered in 1953, is currently used to produce about 50% polyethylene in the world.[44] Still, questions about the structure of its active sites and its activation mechanism have not been resolved.[47–50]

This thesis is a method development effort to model the synthesis and reactivity of atomically dispersed catalysts. The following section describes the problem in more detail and outlines the main computational challenges. It also presents current computational methods applied to model these catalysts and highlights their shortcomings.

1.2 Atomically dispersed amorphous catalysts

A complete in silico model of atomically dispersed catalysts on amorphous supports will require the following: 1) atomistic models representative of real amorphous supports, 2) modeling grafting of active metal sites to amorphous supports, and 3) modeling the catalytic activity of the grafted catalyst. Each of these steps are discussed in detail below.

1.2.1 Amorphous supports: silica

Silica,[44] silica-alumina,[51] and alumina[52] are some examples of amorphous materials used as supports for atomically dispersed catalysts. Here, we limit our discussion of supports to amorphous silica. Challenges presented in developing atomistic models of amorphous silica also apply to other amorphous supports.

Silica is one of the most commonly used supports for atomically dispersed catalysts. Because of the absence of significant Lewis and Brønsted acidity silica does not take part in side-reactions.[43] Furthermore, it offers a large surface area and high mechanical strength.[43] The surface of amorphous silica is terminated by silanol groups (≡SiOH) and siloxanes (≡Si-O-Si≡). Silanols are often classified based on their connectivity to other silanols. For example, hydroxyl groups attached to the same Si atom are called geminal silanols and two silanols connected via a Si-O-Si bridge are called vicinal silanols.[43] Vicinal silanol pairs can be H-bonded or not. Other distance-based classifications have also been proposed: silanols separated by less than 4 Å are called nearly free silanols and silanols separated by greater than 6 Å are called completely free silanols.[53] Apart from these classifications into discrete categories, the amorphous nature of the support leads to different local structural environments, with different bond lengths and bond angles. These local structural variations influence the reactivity of silanols with other molecules.

Amorphous silicas can be non-porous or mesoporous. Aerosil, an example of a non-porous silica support, is synthesized via flame pyrolysis of $SiCl_4$.[54] SBA-15[55] and MCM-41,[56,57] examples of mesoporous silica supports, are synthesized by creating an ordered template with parallel cylindrical micelles via the self-assembly of surfactant molecules. Following this, a silica precursor, for example tetraethoxysilane (TEOS), is added to the ordered template. Finally, the material is calcined to remove the template and obtain the mesoporous amorphous support. Silicas synthesized via these methods can be

further calcined to remove adsorbed water and to condense neighboring silanol groups to modulate the silanol density.[43] These calcination and self-assembly processes are often performed under non-equilibrium conditions.[58]

Molecular simulations to model these non-equilibrium synthesis processes with multiple activated events are challenging.[19] Some studies have attempted to model silica supports using alchemical procedures.[59–62] These methods generally start by melting models of crystalline silica at a high temperature (ca. 3000-7000K). The melt is then quenched to obtain bulk amorphous silica. The surface of the amorphous bulk is cleaved, and unsaturated O atoms are capped with H atoms and unsaturated Si atoms are capped with OH groups. Nearby silanol pairs can be condensed to tune the surface silanol density of these models. However, many ad hoc assumptions are made in generating these models. Different studies use different force fields/*ab initio* methods, melting temperatures, and quench rates. Furthermore, different cleaving, capping, and condensation protocols are employed. These differences lead to different structural features of the produced silica models, which can influence the subsequent grafting and catalytic steps.

Amorphous silica exhibits short-range and limited medium-range order.[42,63] However, as described in the introduction, the lack long-range order precludes its precise characterization.[43] Therefore, the structures obtained from molecular simulations cannot be directly compared to experimentally synthesized silicas. Some studies have attempted to qualitatively reproduce the experimental IR spectra of silicas and the experimentally measured silanol density vs. calcination temperature curves.[59–62] However, several questions still remain unanswered. For example, are models generated using these alchemical melt-quench methods accurate? Is qualitatively reproducing the IR spectrum and silanol density vs. calcination temperature curves enough to establish the validity of silica models? In addition, these models are small with tens of silanols. It is unlikely that such small models can represent real silica materials with ca. 10^{18} silanols (silica samples used

in typical catalyst experiments have ca. 10^{18} silanols).[64]

1.2.2 Grafting and catalysis

Atomically dispersed catalysts can be synthesized by grafting organometallic precursors onto amorphous silica. Precursors can graft as monopodal or bipodal species via reaction with one or two silanol groups, respectively.[44,65,66] Some studies have also reported the formation of tripodal species.[66–68] Some examples of organometallic precursors used to graft metals onto amorphous supports are $TiCl_4$,[69] $VOCl_3$,[70], CrO_2Cl_2,[71,72] and $GaMe_3$.[73] Eq. 1.1 shows an example grafting reaction. An organometallic precursor ML_4 (a metal M bonded to four ligands L) reacts with a pair of silanols and grafts as a bipodal species following the elimination of 2 HL molecules.

$$2(\equiv \mathrm{SiOH}) + \mathrm{ML_4} \rightleftarrows (\equiv \mathrm{SiO})_2\mathrm{ML_2} + 2\mathrm{HL}. \tag{1.1}$$

The nature of grafted species has been inferred using a variety of spectroscopic methods, including, EXAFS,[74,75] XAS,[76,77] IR,[78,79] etc. But precise structural information about the distribution of site environments is not possible because of the absence of long-range order in these materials as described in the last section. Differences in the local structure of the silanol sites can influence the kinetics and thermodynamics of grafting.[66,80] Investigating grafting kinetics can be important in scenarios where a fast-forming species is the dominant product instead of a slow forming thermodynamically stable species. It has been shown experimentally, in some cases, that the nature of the grafted species can depend on the grafting reaction time.[81] But experimental kinetic studies of grafting have been rare. The few studies which have investigated grafting kinetics have used single-exponent models to fit kinetic data.[82–84] This clearly does not take into account the possibility of different sites grafting at different rates.

Computational studies of grafting metal complexes onto amorphous supports have also been scarce. A few studies have investigated grafting using single-site models.[85–87] These models are selected based on ad hoc assumptions about the silica site and do not account for the diversity of structural environments on amorphous silica surfaces. Some studies have used amorphous support models with a limited number of sites to model grafting.[88–90] However, most such efforts have been limited to investigating the thermodynamics of the grafting process without considering the kinetics.

The final step is to study the activtiy of the synthesized catalyst. The grafted metal centers can turnover at different rates depending on their local structural environment.[91–93] In some cases, reactions can even proceed via different catalytic mechanisms on different sites.[94] Furthermore, only a small fraction, ca. 1%-10%, of the sites are active in such catalysts as estimated by active site counting experiments.[95–97] This further complicates their characterization as spectroscopic measurements represent features of common but inactive sites. Similar to kinetic studies of grafting, experimental studies investigating catalytic kinetics use single-exponent models to fit kinetic data, thus not considering the distribution of sites on these catalysts.[65,98,99]

For these catalysts, most computational investigations of catalytic mechanisms have used single-site models.[49,100,101] Similar to grafting, ad hoc assumptions about the structure of these sites are made. And the diversity of sites is not considered. Modeling the catalytic activity is further complicated because only a small fraction of sites are active. Randomly sampling sites to calculate site-averaged kinetic properties will primarily select inactive sites for expensive *ab initio* calculations, making traditional computational approaches impractical. Hence, we need new computational methods to efficiently model the grafting and the catalytic steps for these catalysts.

1.3 Outline

The challenges outlined in the previous section cannot be satisfactorily answered using established computational methods. This thesis is a collection of method development efforts towards modeling the synthesis and reactivity of atomically dispersed catalysts on amorphous supports. Specifically, we develop tools to model the grafting of organometallic precursors to amorphous supports and the catalytic activity of the grafted catalysts, i.e., steps 2 and 3 described in section 1.2.

Chapter 2 develops a machine learning (ML) parametrized population balance model to predict the evolving distribution of sites during grafting. In chapter 3, we present an importance learning algorithm, to discover rare and active sites on atomically dispersed catalysts. Importance learning uses a combination of importance sampling and machine learning to efficiently calculate the site-averaged activation barrier of these catalysts.

Chapter 4 applies the population balance framework (developed in chapter 2) to model the kinetics of $TiCl_4$ grafting onto amorphous silica. The predicted equilibrium populations of the grafted sites are compared with the outcomes of $TiCl_4$ grafting experiments. The effect of reaction conditions on the distribution of grafted species is discussed.

Chapter 5 compares atomistic models of amorphous silica generated using different simulations protocols. Similarities and differences between different models are discussed and experimental metrics that future studies can use to validate their models are presented.

Chapter 6 describes the development of a site balance algebra to quantify the amounts of monopodal [$\equiv SiOTiCl_3$] and bipodal [$(\equiv SiO)_2TiCl_2$] sites produced in $TiCl_4$ grafting experiments. The use of such a site balance algebra as a consistency checking tool for grafting experiments is discussed.

Chapter 7 summarizes the work presented in chapters 2 through 6 and provides some

broad conclusions. Additionally, it outlines challenges that still remain to be solved and describes a few directions that can immediately be pursued to extend the work presented in this thesis.

Bibliography

1. Ma, Z. & Zaera, F. Heterogeneous catalysis by metals. *Encyclopedia of Inorganic Chemistry* (2006).

2. Faria, J. A. Renaissance of ammonia synthesis for sustainable production of energy and fertilizers. *Current Opinion in Green and Sustainable Chemistry* **29,** 100466. ISSN: 2452-2236. https://doi.org/10.1016/j.cogsc.2021.100466 (2021).

3. Catalytic converters: state of the art and perspectives. *Catalysis Today* **51,** 351–360. ISSN: 0920-5861. https://doi.org/10.1016/S0920-5861(99)00024-3 (1999).

4. Carrette, L., Friedrich, K. A. & Stimming, U. Fuel Cells: Principles, Types, Fuels, and Applications. *ChemPhysChem* **1,** 162–193. https://doi.org/10.1002/1439-7641(20001215)1:4%3C162::AID-CPHC162%3E3.0.CO;2-Z (2000).

5. Rostrup-Nielsen, J. R. Catalysis and large-scale conversion of natural gas. *Catalysis Today* **21,** 257–267. ISSN: 0920-5861. https://doi.org/10.1016/0920-5861(94)80147-9 (1994).

6. Khouni, I., Marrot, B., Moulin, P. & Ben Amar, R. Decolourization of the reconstituted textile effluent by different process treatments: Enzymatic catalysis, coagulation/flocculation and nanofiltration processes. *Desalination* **268,** 27–37. ISSN: 0011-9164. https://doi.org/10.1016/j.desal.2010.09.046 (2011).

7. Wisniak, J. The History of Catalysis. From the Beginning to Nobel Prizes. *Educación Química* **21,** 60–69. ISSN: 0187-893X. https://doi.org/10.1016/S0187-893X(18)30074-0 (2010).

8. Langmuir, I. THE CONSTITUTION AND FUNDAMENTAL PROPERTIES OF SOLIDS AND LIQUIDS. PART I. SOLIDS. *Journal of the American Chemical Society* **38,** 2221–2295. https://doi.org/10.1021/ja02268a002 (1916).

9. Langmuir, I. THE CONSTITUTION AND FUNDAMENTAL PROPERTIES OF SOLIDS AND LIQUIDS. II. LIQUIDS.1. *Journal of the American Chemical Society* **39,** 1848–1906. https://doi.org/10.1021/ja02254a006 (1917).

10. Langmuir, I. THE ADSORPTION OF GASES ON PLANE SURFACES OF GLASS, MICA AND PLATINUM. *Journal of the American Chemical Society* **40,** 1361–1403. https://doi.org/10.1021/ja02242a004 (1918).

11. Taylor, H. S. A theory of the catalytic surface. *Proc. R. Soc. London, Ser. A* **108,** 105. https://doi.org/10.1098/rspa.1925.0061 (1925).

12. Ertl, G., Knözinger, H., Weitkamp, J., *et al. Handbook of heterogeneous catalysis* (VCH Weinheim, 1997).

13. Schoonheydt, R. A. UV-VIS-NIR spectroscopy and microscopy of heterogeneous catalysts. *Chem. Soc. Rev.* **39,** 5051–5066. http://dx.doi.org/10.1039/C0CS00080A (12 2010).

14. Stair, P. C. Advances in Raman spectroscopy methods for catalysis research. *Current Opinion in Solid State and Materials Science* **5,** 365–369. ISSN: 1359-0286. https://doi.org/10.1016/S1359-0286(01)00034-1 (2001).

15. Dumesic, J. A. & Topsøe, H. in (eds Eley, D., Pines, H. & Weisz, P. B.) 121–246 (Academic Press, 1977). https://doi.org/10.1016/S0360-0564(08)60071-1.

16. Ryczkowski, J. IR spectroscopy in catalysis. *Catalysis Today* **68.** IR Spectroscopy in Catalysis, 263–381. ISSN: 0920-5861. https://doi.org/10.1016/S0920-5861(01)00334-0 (2001).

17. Hunger, M. & Weitkamp, J. In situ IR, NMR, EPR, and UV/Vis Spectroscopy: Tools for New Insight into the Mechanisms of Heterogeneous Catalysis. *Angewandte Chemie International Edition* **40,** 2954–2971. https://doi.org/10.1002/1521-3773(20010817)40:16%3C2954::AID-ANIE2954%3E3.0.CO;2-%5C# (2001).

18. Van Santen, R. A. & Neurock, M. *Molecular heterogeneous catalysis: a conceptual and computational approach* (John Wiley & Sons, 2009).

19. Peters, B. *Reaction Rate Theory and Rare Events* (Elsevier, Amsterdam ; Cambrige, MA, 2017).

20. Schrödinger, E. An Undulatory Theory of the Mechanics of Atoms and Molecules. *Phys. Rev.* **28,** 1049–1070. https://doi.org/10.1103/PhysRev.28.1049 (6 1926).

21. Hehre, W. J., Ditchfield, R. & Pople, J. A. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *The Journal of Chemical Physics* **56,** 2257–2261. https://doi.org/10.1063/1.1677527 (1972).

22. Parr, R. G. in *Horizons of quantum chemistry* 5–15 (Springer, 1980).

23. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77,** 3865–3868. https://doi.org/10.1103/PhysRevLett.77.3865 (18 1996).

24. Greeley, J., Nørskov, J. K. & Mavrikakis, M. ELECTRONIC STRUCTURE AND CATALYSIS ON METAL SURFACES. *Annual Review of Physical Chemistry* **53.** PMID: 11972011, 319–348. https://doi.org/10.1146/annurev.physchem.53.100301.131630 (2002).

25. Marx, D. & Hutter, J. Ab initio molecular dynamics: Theory and implementation. *Modern methods and algorithms of quantum chemistry* **1,** 141 (2000).

26. Frenkel, D. & Smit, B. *Understanding molecular simulation: from algorithms to applications* (Elsevier, 2001).

27. Jensen, F. *Introduction to computational chemistry* (John wiley & sons, 2017).

28. Ovesen, C. *et al.* A Microkinetic Analysis of the Water–Gas Shift Reaction under Industrial Conditions. *Journal of Catalysis* **158,** 170–180. ISSN: 0021-9517. https://doi.org/10.1006/jcat.1996.0016 (1996).

29. Reuter, K. & Scheffler, M. First-principles kinetic Monte Carlo simulations for heterogeneous catalysis: Application to the CO oxidation at $RuO_2(110)$. *Phys. Rev. B* **73,** 045433. https://doi.org/10.1103/PhysRevB.73.045433 (4 2006).

30. Greeley, J., Jaramillo, T. F., Bonde, J., Chorkendorff, I. & Nørskov, J. K. Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. *Nature materials* **5,** 909–913. https://doi.org/10.1038/nmat1752 (2006).

31. Morokuma, K. & Musaev, D. G. *Computational Modeling for Homogeneous and Enzymatic Catalysis: A Knowledge-Base for Designing Efficient Catalysis* (John Wiley & Sons, 2008).

32. Sperger, T., Sanhueza, I. A., Kalvet, I. & Schoenebeck, F. Computational Studies of Synthetically Relevant Homogeneous Organometallic Catalysis Involving Ni, Pd, Ir, and Rh: An Overview of Commonly Employed DFT Methods and Mechanistic Insights. *Chem. Rev.* **115,** 9532–9586. https://www.doi.org/10.1021/acs.chemrev.5b00163 (2015).

33. Goldsmith, B. R., Hwang, T., Seritan, S., Peters, B. & Scott, S. L. Rate-Enhancing Roles of Water Molecules in Methyltrioxorhenium-Catalyzed Olefin Epoxidation by Hydrogen Peroxide. *J. Am. Chem. Soc.* **137,** 9604–9616. https://www.doi.org/10.1021/jacs.5b03750 (2015).

34. Blaszkowski, S. R. & van Santen, R. A. Theoretical study of the mechanism of surface methoxy and dimethyl ether formation from methanol catalyzed by zeolitic protons. *The Journal of Physical Chemistry B* **101,** 2292–2305. https://doi.org/10.1021/jp962006 (1997).

35. Li, Y.-P., Head-Gordon, M. & Bell, A. T. Analysis of the Reaction Mechanism and Catalytic Activity of Metal-Substituted Beta Zeolite for the Isomerization of Glucose to Fructose. *ACS Catalysis* **4,** 1537–1545. https://www.doi.org/10.1021/cs401054f (2014).

36. McFarland, E. W. & Metiu, H. Catalysis by Doped Oxides. *Chemical Reviews* **113,** 4391–4427. https://www.doi.org/10.1021/cr300418s (2013).

37. Le Bahers, T., Rérat, M. & Sautet, P. Semiconductors Used in Photovoltaic and Photocatalytic Devices: Assessing Fundamental Properties from DFT. *The Journal of Physical Chemistry C* **118,** 5997–6008. https://www.doi.org/10.1021/jp409724c (2014).

38. Hammer, B. & Nørskov, J. K. in, 71–129 (Academic Press, 2000). https://www.doi.org/10.1016/S0360-0564(02)45013-4.

39. Greeley, J. Theoretical Heterogeneous Catalysis: Scaling Relationships and Computational Catalyst Design. *Annu Rev Chem Biomol* **7,** 605–635. https://www.doi.org/10.1146/annurev-chembioeng-080615-034413 (2016).

40. Yoon, C. & Cocke, D. L. Potential of amorphous materials as catalysts. *Journal of Non-Crystalline Solids* **79,** 217–245. ISSN: 0022-3093. https://doi.org/10.1016/0022-3093(86)90224-3 (1986).

41. Goldsmith, B. R., Peters, B., Johnson, J. K., Gates, B. C. & Scott, S. L. Beyond Ordered Materials: Understanding Catalytic Sites on Amorphous Solids. *ACS Catal.* **7,** 7543–7757. https://www.doi.org/10.1021/acscatal.7b01767 (2017).

42. Elliott, S. R. Medium-range structural order in covalent amorphous solids. *Nature* **354,** 445–452. ISSN: 1476-4687. https://doi.org/10.1038/354445a0 (1991).

43. Vansant, E. E., Van Der Voort, P. & Vrancken, K. C. *Characterization and Chemical Modification of the Silica Surface* (1995).

44. McDaniel, M. P. A Review of the Phillips Supported Chromium Catalyst and Its Commercial Use for Ethylene Polymerization. English. *Adv. Catal.* **53,** 123–606. https://www.doi.org/https://doi.org/10.1016/S0360-0564(10)53003-7 (2010).

45. Cativiela, C., Fraile, J., García, J. & Mayoral, J. A new titanium-silica catalyst for the epoxidation of alkenes. *Journal of Molecular Catalysis A: Chemical* **112,** 259–267. ISSN: 1381-1169. https://doi.org/10.1016/1381-1169(96)00131-8 (1996).

46. Rhers, B. *et al.* A Well-Defined, Silica-Supported Tungsten Imido Alkylidene Olefin Metathesis Catalyst. *Organometallics* **25,** 3554–3557. ISSN: 0276-7333. https://doi.org/10.1021/om060279d (2006).

47. McDaniel, M. P. & Welch, M. B. The activation of the phillips polymerization catalyst: I. Influence of the hydroxyl population. *J. Catal.* **82,** 98–109. https://www.doi.org/10.1016/0021-9517(83)90121-5 (1983).

48. Fong, A., Vandervelden, C., Scott, S. L. & Peters, B. Computational Support for Phillips Catalyst Initiation via Cr-C Bond Homolysis in a Chromacyclopentane Site. *ACS Catal.* **8,** 1728–1733. https://doi.org/10.1021/acscatal.7b03724 (2018).

49. Delley, M. F. *et al.* Proton Transfers Are Key Elementary Steps in Ethylene Polymerization on Isolated Chromium(III) Silicates. *Proc. Natl. Acad. Sci.* **111,** 11624–11629. https://www.doi.org/10.1073/pnas.1405314111 (2014).

50. Peters, B., Scott, S. L., Fong, A., Wang, Y. & Stiegman, A. E. Reexamining the Evidence for Proton Transfers in Ethylene Polymerization. *Proc. Natl. Acad. Sci. USA* **112,** E4160–E4161. https://www.doi.org/10.1073/pnas.1422589112 (2015).

51. Ali, M., Tatsumi, T. & Masuda, T. Development of heavy oil hydrocracking catalysts using amorphous silica-alumina and zeolites as catalyst supports. *Applied Catalysis A: General* **233,** 77–90. ISSN: 0926-860X. https://doi.org/10.1016/S0926-860X(02)00121-7 (2002).

52. Tang, N. *et al.* Coordinatively Unsaturated Al3+ Sites Anchored Subnanometric Ruthenium Catalyst for Hydrogenation of Aromatics. *ACS Catalysis* **7,** 5987–5991. https://doi.org/10.1021/acscatal.7b01816 (2017).

53. Pavan, C. *et al.* Nearly free surface silanols are the critical molecular moieties that initiate the toxicity of silica particles. *Proceedings of the National Academy of Sciences* **117,** 27836. https://doi.org/10.1073/pnas.2008006117 (2020).

54. Mathias, J. & Wannemacher, G. Basic characteristics and applications of aerosil: 30. The chemistry and physics of the aerosil Surface. *Journal of Colloid and Interface Science* **125,** 61–68. ISSN: 0021-9797. https://doi.org/10.1016/0021-9797(88)90054-9 (1988).

55. Zhao, D. *et al.* Triblock Copolymer Syntheses of Mesoporous Silica with Periodic 50 to 300 Angstrom Pores. *Science* **279,** 548–552. ISSN: 0036-8075. https://doi.org/10.1126/science.279.5350.548 (1998).

56. Beck, J. S. *et al.* A new family of mesoporous molecular sieves prepared with liquid crystal templates. *Journal of the American Chemical Society* **114,** 10834–10843. https://doi.org/10.1021/ja00053a020 (1992).

57. Kresge, a. C., Leonowicz, M., Roth, W. J., Vartuli, J. & Beck, J. Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism. *nature* **359,** 710–712. https://doi.org/10.1038/359710a0 (1992).

58. Mann, S. Self-assembly and transformation of hybrid nano-objects and nanostructures under equilibrium and non-equilibrium conditions. *Nature materials* **8,** 781–792. https://doi.org/10.1038/nmat2496 (2009).

59. Tielens, F., Gervais, C., Lambert, J. F., Mauri, F. & Costa, D. Ab initio study of the hydroxylated surface of amorphous silica: A representative model. *Chem. Mater.* **20,** 3336–3344. https://www.doi.org/10.1021/cm8001173 (2008).

60. Ugliengo, P. *et al.* Realistic models of hydroxylated amorphous silica surfaces and MCM- 41 mesoporous material simulated by large-scale periodic B3LYP calculations. *Adv. Mater.* **20,** 4579–4583. https://www.doi.org/10.1002/adma.200801489 (2008).

61. Ewing, C. S., Bhavsar, S., Veser, G., McCarthy, J. J. & Johnson, J. K. Accurate amorphous silica surface models from first-principles thermodynamics of surface dehydroxylation. *Langmuir* **30,** 5133–5141. https://doi.org/10.1021/la500422p (2014).

62. Comas-Vives, A. Amorphous SiO_2 surface models: Energetics of the dehydroxylation process, strain, ab initio atomistic thermodynamics and IR spectroscopic signatures. *Phys. Chem. Chem. Phys.* **18,** 7475–7482. https://www.doi.org/10.1039/c6cp00602g (2016).

63. Sørensen, S. S., Biscio, C. A. N., Bauchy, M., Fajstrup, L. & Smedskjaer, M. M. Revealing hidden medium-range order in amorphous materials using topological data analysis. *Science Advances* **6,** eabc2320. https://doi.org/10.1126/sciadv.abc2320 (2020).

64. Khan, S. A., Vandervelden, C. A., Scott, S. L. & Peters, B. Grafting metal complexes onto amorphous supports: From elementary steps to catalyst site populations: Via kernel regression. *React. Chem. Eng.* **5,** 66–76. https://www.doi.org/10.1039/c9re00357f (2020).

65. Zhong, L. *et al.* Spectroscopic and structural characterization of Cr(II)/SiO_2 active site precursors in model Phillips polymerization catalysts. *J. Catal.* **293,** 1–12. https://www.doi.org/10.1016/j.jcat.2012.05.014 (2012).

66. Mania, P., Verel, R., Jenny, F., Hammond, C. & Hermans, I. Thermal Restructuring of Silica-Grafted TiClx Species and Consequences for Epoxidation Catalysis. *Chemistry - A European Journal* **19,** 9849–9858. ISSN: 0947-6539. https://doi.org/10.1002/chem.201300842 (2013).

67. Iwai, T., Murakami, R., Harada, T., Kawamorita, S. & Sawamura, M. Silica-Supported Tripod Triarylphosphane: Application to Transition MetalCatalyzed

C(sp3)H Borylations. *Advanced Synthesis & Catalysis* **356,** 1563–1570. https://doi.org/10.1002/adsc.201301147 (2014).

68. Coperet, C. *et al.* Surface Organometallic and Coordination Chemistry toward Single-Site Heterogeneous Catalysts: Strategies, Methods, Structures, and Activities. *Chem. Rev.* **116,** 323–421. https://www.doi.org/10.1021/acs.chemrev.5b00373 (2016).

69. Kytökivi, A. & Haukka, S. Reactions of HMDS, $TiCl_4$, $ZrCl_4$, and $AlCl_4$ with Silica As Interpreted from Low-Frequency Diffuse Reflectance Infrared Spectra. *The Journal of Physical Chemistry B* **101,** 10365–10372. https://doi.org/10.1021/jp971822u (1997).

70. Zhu, H. *et al.* VO_x/SiO_2 Catalyst Prepared by Grafting $VOCl_3$ on Silica for Oxidative Dehydrogenation of Propane. *ChemCatChem* **7,** 3332–3339. https://www.doi.org/10.1002/cctc.201500607 (2015).

71. McDaniel, M. The state of Cr(VI) on the Phillips polymerization catalyst: II. The reaction between silica and CrO_2Cl_2. *Journal of Catalysis* **76,** 17–28. ISSN: 0021-9517. https://doi.org/10.1016/0021-9517(82)90232-9 (1982).

72. Demmelmaier, C. A., White, R. E., van Bokhoven, J. A. & Scott, S. L. Nature of $\equiv SiOCrO_2Cl$ and $(\equiv SiO)_2CrO_2$ Sites Prepared by Grafting CrO_2Cl_2 onto Silica. *The Journal of Physical Chemistry C* **112,** 6439–6449. https://doi.org/10.1021/jp7119153 (2008).

73. Taha, Z. A., Deguns, E. W., Chattopadhyay, S. & Scott, S. L. Formation of Digallium Sites in the Reaction of Trimethylgallium with Silica. *Organometallics* **25,** 1891–1899. ISSN: 0276-7333. https://doi.org/10.1021/om051034o (2006).

74. Roberts, D. R., Ford, R. G. & Sparks, D. L. Kinetics and mechanisms of Zn complexation on metal oxides using EXAFS spectroscopy. *Journal of Colloid and Interface Science* **263,** 364–376. ISSN: 0021-9797. https://doi.org/10.1016/S0021-9797(03)00281-9 (2003).

75. östhols, E., Manceau, A., Farges, F. & Charlet, L. Adsorption of Thorium on Amorphous Silica: An EXAFS Study. *Journal of Colloid and Interface Science* **194,** 10–21. ISSN: 0021-9797. https://doi.org/10.1006/jcis.1997.5065 (1997).

76. Groppo, E. *et al.* In situ, Cr K-edge XAS study on the Phillips catalyst: activation and ethylene polymerization. *Journal of Catalysis* **230,** 98–108. ISSN: 0021-9517. https://doi.org/10.1016/j.jcat.2004.11.017 (2005).

77. van der Meer, Y., Hensen, E., van Veen, J. & van der Kraan, A. Characterization and thiophene hydrodesulfurization activity of amorphous-silica–alumina-supported NiW catalysts. *Journal of Catalysis* **228,** 433–446. ISSN: 0021-9517. https://doi.org/10.1016/j.jcat.2004.09.019 (2004).

78. Trombetta, M. *et al.* FT-IR Studies on Light Olefin Skeletal Isomerization Catalysis: III. Surface Acidity and Activity of Amorphous and Crystalline Catalysts Belonging to the SiO_2–Al_2O_3System. *Journal of Catalysis* **179,** 581–596. ISSN: 0021-9517. https://doi.org/10.1006/jcat.1998.2251 (1998).

79. Lamberti, C., Zecchina, A., Groppo, E. & Bordiga, S. Probing the surfaces of heterogeneous catalysts by in situ IR spectroscopy. *Chem. Soc. Rev.* **39,** 4951–5001. http://dx.doi.org/10.1039/C0CS00117A (12 2010).

80. Fraile, J. M., García, J. I., Mayoral, J. A. & Vispe, E. Catalytic sites in silica-supported titanium catalysts: silsesquioxane complexes as models. *Journal of Catal-*

ysis **233,** 90–99. ISSN: 0021-9517. https://doi.org/10.1016/j.jcat.2005.04.018 (2005).

81. Zapilko, C. *et al.* Advanced Surface Functionalization of Periodic Mesoporous Silica: Kinetic Control by Trisilazane Reagents. *Journal of the American Chemical Society* **128,** 16266–16276. ISSN: 0002-7863. https://doi.org/10.1021/ja065444v (2006).

82. Cullen, R. J. *et al.* Spontaneous Grafting of Nitrophenyl Groups on Amorphous Carbon Thin Films: A Structure–Reactivity Investigation. *Chemistry of Materials* **24,** 1031–1040. ISSN: 0897-4756. https://doi.org/10.1021/cm2030262 (2012).

83. Sha, X., Xu, X., Sohlberg, K., Loll, P. J. & Penn, L. S. Evidence that three-regime kinetics is inherent to formation of a polymer brush by a grafting-to approach. *RSC Advances* **4,** 42122–42128. http://dx.doi.org/10.1039/C4RA05663A (2014).

84. Chaudhari, C. V. *et al.* Ethylene vinyl acetate based radiation grafted hydrophilic matrices: Process parameter standardization, grafting kinetics and characterization. *Radiation Physics and Chemistry* **125,** 213–219. ISSN: 0969-806X. https://doi.org/10.1016/j.radphyschem.2016.04.017 (2016).

85. Hu, Z. & Turner, C. H. Initial Surface Reactions of TiO_2 Atomic Layer Deposition onto SiO_2 Surfaces: Density Functional Theory Calculations. *The Journal of Physical Chemistry B* **110,** 8337–8347. ISSN: 1520-6106. https://doi.org/10.1021/jp060367b (2006).

86. Hu, Z. & Turner, C. H. Atomic Layer Deposition of TiO_2 from TiI_4 and H_2O onto SiO_2 Surfaces: Ab Initio Calculations of the Initial Reaction Mechanisms. *Journal of the American Chemical Society* **129,** 3863–3878. ISSN: 0002-7863. https://doi.org/10.1021/ja066529z (2007).

87. Cheng, R., Liu, X., Fang, Y., Terano, M. & Liu, B. High-resolution 29Si CP/MAS solid state NMR spectroscopy and DFT investigation on the role of geminal and single silanols in grafting chromium species over Phillips Cr/silica catalyst. *Applied Catalysis A: General* **543,** 26–33. ISSN: 0926-860X. https://doi.org/10.1016/j.apcata.2017.05.011 (2017).

88. Guesmi, H. & Tielens, F. Chromium oxide species supported on silica: A representative periodic DFT model. *J Phys Chem C* **116,** 994–1001. https://doi.org/10.1021/jp209680r (2012).

89. Guesmi, H., Gryboś, R., Handzlik, J. & Tielens, F. Characterization of molybdenum monomeric oxide species supported on hydroxylated silica: a DFT study. *Physical Chemistry Chemical Physics* **16,** 18253–18260. ISSN: 1463-9076. http://dx.doi.org/10.1039/C4CP02296C (2014).

90. Gueddida, S., Lebégue, S. & Badawi, M. Interaction between transition metals (Co, Ni, and Cu) systems and amorphous silica surfaces: A DFT investigation. *Applied Surface Science* **533,** 147422. ISSN: 0169-4332. https://doi.org/10.1016/j.apsusc.2020.147422 (2020).

91. Goldsmith, B. R., Sanderson, E. D., Bean, D. & Peters, B. Isolated Catalyst Sites on Amorphous Supports: A Systematic Algorithm for Understanding Heterogeneities in Structure and Reactivity. *J. Chem. Phys.* **138,** 204105. https://www.doi.org/10.1063/1.4807384 (2013).

92. Goldsmith, B. R., Fong, A. & Peters, B. in *Reaction Rate Constant Computations: Theories and Applications* 213–232 (The Royal Society of Chemistry, 2013). ISBN: 978-1-84973-650-3. http://dx.doi.org/10.1039/9781849737753-00213.

93. Floryan, L., Borosy, A. P., Nunez-Zarur, F., Comas-Vives, A. & Coperet, C. Strain Effect and Dual Initiation Pathway in Cr(III)/SiO_2 Polymerization Catalysts from

Amorphous Periodic Models. English. *J. Catal.* **346,** 50–56. https://www.doi.org/10.1016/j.jcat.2016.11.037 (2017).

94. Khan, S. W., Deguns, E. W., Peters, B. & Scott, S. L. Quantifing abundances of grafted [(SiO)4-xTiClx] sites on amorphous silica. *in preparation.*

95. Clark, A. Olefin polymerization on supported chromium oxide catalysts. *Catal. Rev.* **3,** 145. https://doi.org/10.1080/01614947008076858 (1970).

96. Gillespie, R. D., Burwell, R. L. & Marks, T. J. Organo-f-element-based heterogeneous catalysts. Kinetics and mechanism of olefin hydrogenation by supported pentamethylcyclopentadienyl actinide complexes. *Langmuir* **6,** 1465 (1990).

97. Amakawa, K. *et al.* In situ generation of active sites in olefin metathesis. *J. Am. Chem. Soc.* **134,** 11473. https://www.doi.org/10.1021/ja3011989 (2012).

98. Woo, T. W. & Woo, S. I. Ethylene polymerization with phillips catalyst Co-catalyzed with Al(i-Bu)$_3$. *Journal of Catalysis* **123,** 215–227. ISSN: 0021-9517. https://doi.org/10.1016/0021-9517(90)90170-O (1990).

99. Corma, A., Grande, M., Gonzalez-Alfaro, V. & Orchilles, A. Cracking Activity and Hydrothermal Stability of MCM-41 and Its Comparison with Amorphous Silica-Alumina and a USY Zeolite. *Journal of Catalysis* **159,** 375–382. ISSN: 0021-9517. https://doi.org/10.1006/jcat.1996.0100 (1996).

100. Fong, A., Yuan, Y., Ivry, S. L., Scott, S. L. & Peters, B. Computational Kinetic Discrimination of Ethylene Polymerization Mechanisms for the Phillips (Cr/SiO_2) Catalyst. English. *ACS Catal.* **5,** 3360–3374. https://www.doi.org/10.1021/acscatal.5b00016 (2015).

101. Handzlik, J. & Ogonowski, J. Structure of Isolated Molybdenum(VI) and Molybdenum(IV) Oxide Species on Silica: Periodic and Cluster DFT Studies. *The Journal of Physical Chemistry C* **116,** 5571–5584. https://www.doi.org/10.1021/jp207385h (2012).

Chapter 2

Grafting metal complexes onto amorphous supports: from elementary steps to catalyst site populations via kernel regression

Reproduced in part with permission from: Khan, S. A.; Vandervelden, C. A.; Scott, S. L.; Peters, B. Grafting metal complexes onto amorphous supports: from elementary steps to catalyst site populations *via* kernel regression. https://doi.org/10.1039/C9RE00357F

2.1 Introduction

Most *ab initio* computational catalysis studies focus on homogeneous catalysts,[1–3] enzymes,[4–6] or heterogeneous catalysts with ordered structures such as metals,[7–11] zeolites,[12–14] and crystalline metal oxides.[15–17] All of these materials have in common the advantage that many features of the catalyst structure are known. Even for molecular

catalysts and enzymes, where the active site resides within a fluctuating environment, there are systematic computational frameworks for averaging over the fluctuations.[18–20] In contrast, amorphous catalysts cannot be modelled with small, periodically repeating solid structures, nor by sampling a well-defined ensemble for liquid phase disorder. Instead, the quenched disorder in an amorphous heterogeneous catalyst[21,22] is a permanent signature of its non-equilibrium preparation history. Examples within this family include the Phillips catalyst (Cr/SiO_2) for ethylene polymerization,[23] molybdenum (Mo/SiO_2) and tungsten (W/SiO_2) catalysts for olefin metathesis,[24] and titanium catalysts (Ti/SiO_2) for alkene epoxidation.[25]

Because of these difficulties, amorphous catalysts have mostly been avoided in *ab initio* computational studies. Those exceptions in which calculations on amorphous catalysts were attempted were forced to rely on questionable assumptions.[22,26–34] For example, are the model sites representative of the real material? Do the models accurately represent the most active sites? Can reliable conclusions about the reaction kinetics be drawn from a single-site computational model? At present, none of these questions can be satisfactorily answered with *ab initio* calculations.

These questions are addressed in two papers, this one and a companion. They provide a computational framework that combines machine learning, statistical importance sampling, and population balance modeling techniques. To illustrate the concepts and methods, we begin with a model for an atomically-dispersed catalyst on an amorphous support. The essential features of the model are a quenched disordered support scaffold (to represent an amorphous silica matrix), surface silanol sites where precursors can be attached (to represent surface hydroxyl groups), and a microkinetic model for grafting at each silanol site. These microkinetic models have rate parameters that depend on the individual grafting site characteristics. The rate parameters at each grafting site will be determined, much like in a real *ab initio* calculation, by structural optimization of the

intermediates using a simple force field.

This first paper deals with how the active sites are generated during catalyst preparation. In particular, we show how the populations of both grafted sites and the unreacted grafting sites evolve during an idealized grafting process. If the surface reactions are irreversible, the final metal site distribution will be determined by those surface grafting sites with the fastest grafting kinetics. However, if the surface reactions are reversible, the final grafted site distribution will favor grafted sites that lead to the most stable grafted species. To enable *ab initio* studies in the future, the algorithm must efficiently predict the characteristics of the most reactive grafting sites and their abundances, without performing exhaustive *ab initio* calculations for many thousands of grafting sites. We demonstrate how kernel regression can learn to anticipate the outcomes of these optimizations. Then, by applying the kernel regression model to thousands of grafting sites, we can construct a population balance model for the grafting process. The simplicity of our model system allows us to test the accelerated predictions against an exhaustive parameterization from structure optimizations at thousands of grafting sites.

The companion paper uses a grafted site population that reflects both the disordered support and the superimposed grafting kinetics to predict site-averaged kinetics. Because turnover frequencies at individual grafting sites depend exponentially on their activation energies, site-averaged kinetics are difficult to converge without rare events sampling methods. The second paper deals with averaging over the non-uniform distribution of grafted sites to predict the overall kinetics.

The remainder of this paper is as follows. First, we introduce simple models for the amorphous support and grafting kinetics. Next, we use kernel regression tools to predict the grafting thermodynamics and kinetics based on a concise list of grafting site characteristics. Finally, we use the kernel regression results and kinetic models to parameterize the population balance model for grafting.

2.2 Amorphous silica

Amorphous silica is a commonly used catalyst support because of its thermal and mechanical stability, large surface area, and its chemical inertness. The surface of silica is terminated by silanol groups which may be categorized as isolated, geminal, vicinal etc. Real amorphous silicas are created via sol-gel synthesis, spray drying, pyrolysis, or precipitation methods. Silica can be calcined to increase its mechanical strength and to remove adsorbed water.[35–40] The calcination temperature also determines the residual surface silanol density, which in turn influences the activity of the supported catalyst.[23,41,42]

Many studies have used spectroscopic techniques like IR, NMR, and EPR to investigate the populations of different silanol types.[43,44] However, in contrast to crystalline materials, the absence of long-range order results in broad peaks that complicate the precise characterization of silica.

Silicas of different types exhibit different ring size distributions,[45–47] and silanols of the same type can have different bond angles and different dihedral angles.[35,44] These subtle structural differences between silanols and their environments are likely to influence their reactivity. Many investigators have grafted metal atoms to silica via reactions between silanols and molecular complexes like $AlCl_3$,[48] GaR_3,[49,50] $TiCl_4$,[51–53] and $VOCl_3$.[54,55] These grafting reactions are useful both as probes of local structure and as routes to supported organometallic catalysts.

In a typical grafting experiment, a fluid phase molecular precursor reacts with amorphous silica.[49,55,56] A protonolysis reaction between the precursor and surface silanols results in a metal atom grafted to the silica surface with one, two, or three M-O-Si linkages, sometimes called monopodal, bipodal, or tripodal species, fig. 2.1.[24,57–60]

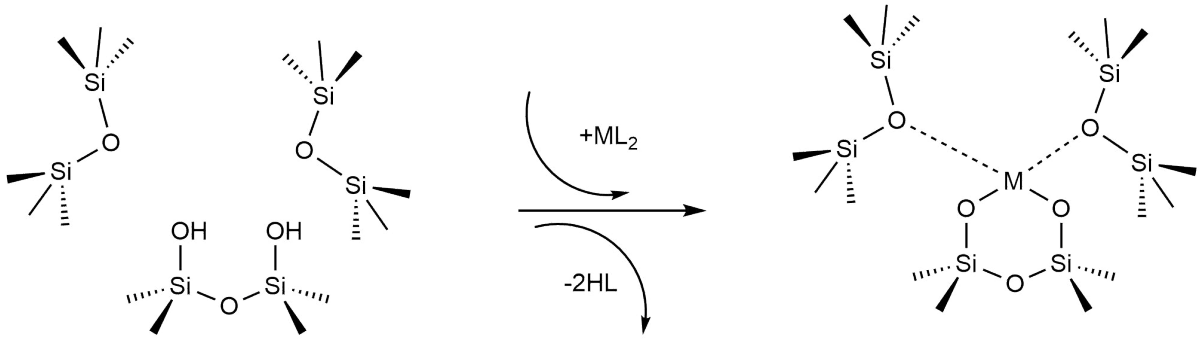


Figure 2.1: Scheme showing the grafting of a molecular $\mathbf{ML_2}$ complex to a vicinal silanol pair. The metal forms two bonds to the silanolate oxygens while two **HL** molecules are eliminated. The metal may also coordinate to nearby siloxane oxygens.

Computational studies of atomically-dispersed metals on silica often use cluster models terminated by hydroxyl groups or hydrogen atoms. These models generally range in sizes from a few to tens of silicon atoms.[28,29,34,61–67] The cluster models are often carved from crystalline materials like zeolites[68] or β-cristobalite.[66,69] In such clusters, the peripheral atoms are fixed at positions characteristic of the crystalline material. The de facto assumption is that larger cluster models are more representative of the real amorphous catalyst. Indeed large cluster models more accurately account for elasticity of the silica matrix and for dispersion interactions between adsorbates and the support.[33,70] However, each layer of silica requires additional and unjustified assumptions about the environment. In this sense, large cluster models are overly specific, while small cluster models are amenable to systematic investigation of the effects of local grafting site geometry.[22]

In the past decade, some computational studies generated amorphous silica surfaces that attempt to reproduce experimental observables like surface silanol density and the IR spectrum.[71–74] Typically, such surfaces are prepared by molecular dynamics simulations, in which crystalline models are heated to high temperatures followed by rapid quenching to generate disordered structures. Then the bulk amorphous structure is cleaved to create the surface. Unsaturated oxygens are capped with hydrogen atoms, and unsaturated

silicons are capped with hydroxyl groups. Finally, pairs of proximal silanols are condensed to achieve the correct surface silanol density. These methods generate atomistic amorphous models of silica with a non-uniform structural distribution of surface silanols. However, such in silico preparation routes for amorphous silica do not correspond to experimental synthesis procedures. In particular, the high surface area of a real silica does not result from cleavage and subsequent functionalization. In addition, the system sizes modelled are typically quite small (100-200 silanols). For comparison, a 10 mg sample of silica with area 350 m^2g^{-1} and 1.0 silanols/nm^2 contains about 10^{18} silanols.

2.2.1 A simple model for amorphous silica

The mechanisms of grafting, activation, and catalytic reactions are still debated for many amorphous catalysts.[26,30,50,58,69,75–77] To resolve the outstanding questions, we need methods that can predict the kinetics at each grafting site and estimate proper site-averaged kinetic properties. Then a given support model (if large enough) and proposed mechanism will yield well-defined, site-averaged predictions to be tested against experiments. To develop such methods, we selected a simple example system for which benchmark calculations can be performed exhaustively, for the full ensemble of non-uniform sites. In this section, we propose a simple abstract model of the amorphous support.

We model the amorphous support as a 2D lattice with quenched disorder, fig. 2.2. Note the loose similarity to the qualitative model of Peri and Hensley.[48] First, a uniform lattice is created in which nearest neighbours are separated by a unit (dimensionless) distance. Each site is randomly displaced (δ) from the uniform lattice by random displacements along the x and y directions to create a disordered lattice. Displacements are drawn from an isotropic 2-D Gaussian distribution (described in the SI). The lattice is then "functionalized" with hydroxyl (-OH), siloxane (≡SiOSi≡), and empty sites with

probabilities p_{OH}, p_{siloxane}, and p_{empty}.

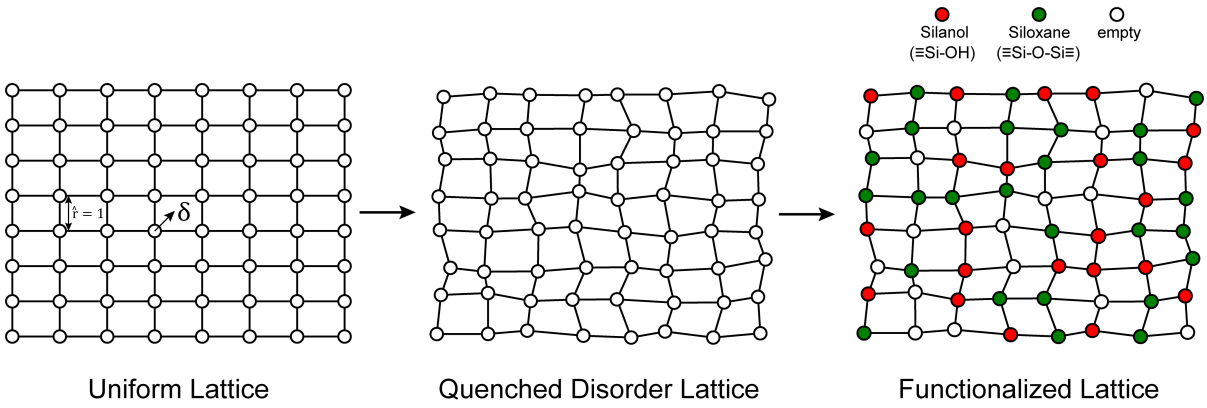


Figure 2.2: Steps to form a functionalized, quenched disorder lattice.

2.2.2 Grafting molecular metal complexes: a simple model

Grafting sites in our model are empty sites surrounded by a pair of vicinal hydroxyls on one axis and a pair of siloxanes on the other axis. Fig. 2.3 shows a grafting site located between vicinal silanols $(\equiv\mathrm{SiOH})_2$ and two siloxanes ($\equiv$SiOSi$\equiv$). The precursor $\mathbf{ML}_2$, a molecular complex, in our model has two displaceable ligands. A real catalyst precursor may have additional ligands like chloride, oxo, or methyl groups that remain bonded to the metal **M** after grafting. The metal is grafted as a bipodal species ($\equiv$SiO**M**OSi$\equiv$) upon reaction of $\mathbf{ML}_2$ with the vicinal hydroxyls to eliminate two **HL** molecules. The metal may also interact with neighbouring siloxanes to form $\mathbf{M}\cdots\mathrm{O}(\mathrm{Si}\equiv)_2$ bonds. The strengths of the $\equiv$SiO-**M** and $\mathbf{M}\cdots\mathrm{O}(\mathrm{Si}\equiv)_2$ bonds depend on the local geometry near the grafting site.

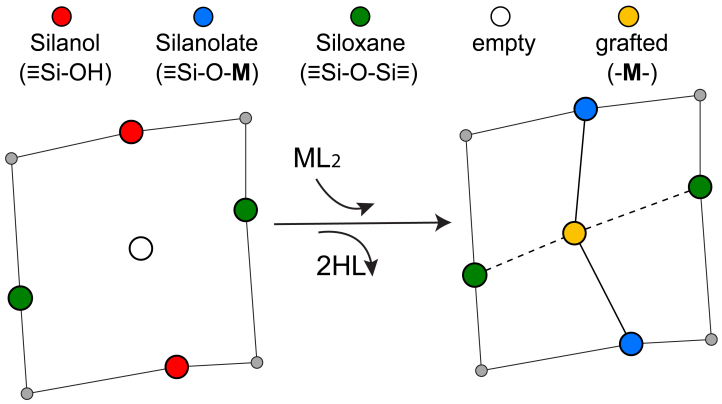


Figure 2.3: Grafting sites on the amorphous 2-D lattice model. One set of opposite nearest neighbour sites are hydroxyl groups, while the other set is siloxanes. $\mathbf{ML_2}$ reacts with two hydroxyls and interacts with the siloxanes to create a grafted **M** atom as shown.

2.2.3 Computing grafting rates on the amorphous silica model

To model grafting kinetics of vicinal silanol sites on the amorphous 2D lattice, we consider the grafting mechanism outlined in Sec. 2.2.2. The grafting process at each vicinal silanol site is assumed to be irreversible with the following rate law:

$$\mathbf{r}(\mathbf{x}) = k(\mathbf{x})[\mathbf{ML_2}]. \tag{2.1}$$

Here $[\mathbf{ML_2}]$ is the gas phase concentration of $\mathbf{ML_2}$, $\mathbf{x}$ represents the local environment of the vicinal silanol site, and $k(\mathbf{x})$ is a site-dependent rate constant.[78] We use concentration to construct rate laws in this work. One can instead use the precursor partial pressure, but note that one must beware of the resulting complications in extracting activation energies. For example, when precursor pressure is set by its T-dependent vapor pressure, as in CrO_2Cl_2 grafting,[58] the pressure and temperature cannot be separately controlled. We use transition state theory (TST) to model the temperature and site-geometry dependence of the grafting rate constant. TST rate constants are widely used

to predict and interpret activation barriers and kinetics across a wide range of catalysis applications.[79–83] TST rate constants are now readily computed from electronic structure calculations.[84] The TST rate constant is:

$$k(\mathbf{x}) = \frac{k_B T}{h} \hat{V}_0 exp[-\beta \Delta G^{\ddagger}(\mathbf{x})]. \tag{2.2}$$

Here $\Delta G^{\ddagger}(\mathbf{x})$ is the grafting barrier as computed with $[\mathbf{ML_2}]$ at the reference volume ($\hat{V}_0$) per particle. Next, we use a Linear Free Energy Relationship (LFER) to model the grafting free energy barrier. Specifically, we assume that the free energy of grafting is linearly related to the activation barrier for grafting:[78]

$$\Delta G^{\ddagger}(\mathbf{x}) = \Delta G^{\ddagger}_{ref} + \alpha \Delta G^{o}(\mathbf{x}). \tag{2.3}$$

Here α $(0 < \alpha < 1)$ is the Brϕnsted coefficient and $\Delta G^{\ddagger}_{ref}$ is the grafting barrier for a reference grafting site with a thermoneutral grafting free energy ($\Delta G^{o}(\mathbf{x}) = 0$). The value of α indicates the position of the transition state between the reactant and product states. Small values of α (near 0) indicate an early transition state that resembles the reactants. Large values of α (near 1) indicate a transition state that resembles the products. In practice, intermediate values of α are common, so we have chosen $\alpha = 1/2$.[85] The value of $\Delta G^{\ddagger}_{ref}$ determines the time scale for grafting, but it will have no bearing on results after non-dimensionalization. Thus, to complete the kinetic model, including the effects of non-uniform grafting sites, we only need a model for $\Delta G^{o}(\mathbf{x})$. The energy to graft the precursor at an empty site is

$$\Delta E(\mathbf{x}) = 2\epsilon_{\mathbf{HF}} + V_{\mathbf{M*}}(\mathbf{x}) - V_{*} - 2\epsilon_{\mathbf{ML}}. \tag{2.4}$$

Here $V_{\mathbf{M}*}(\mathbf{x})$ is the energy of the grafted metal site, V_* is the energy of the unreacted silica site, $\epsilon_{\mathbf{ML}}$ is the energy of the M-L bond, and $\epsilon_{\mathbf{HL}}$ is the energy of the **H-L** bond. V_* is twice the O-H bond energy,

$$V_* = 2\epsilon_{OH} \tag{2.5}$$

Here ϵ_{OH} is the O-H bond energy. To compute $V_{\mathbf{M}*}(\mathbf{x})$, the **M**-OSi≡ bond energy and **M**...O(Si≡)$_2$ bond energy are modelled as Morse potentials:

$$\epsilon_i(r) = D_i(1 - exp[-a_i(r - r_{i,rq})])^2 - D_i. \tag{2.6}$$

Here i is the interaction type (**M**-OSi≡ or **M**$\cdots$O(Si≡)$_2$), Di is the equilibrium energy of the interaction, a_i is related to the width of the potential well, $r_{i,eq}$ is the equilibrium distance, and r is the metal-oxygen bond length. All constants defined in this section are shown in Table 2.1. $V_{\mathbf{M}*}(\mathbf{x})$ is computed by optimizing the position of the metal with surrounding hydroxyl and siloxane positions fixed:

$$V_{\mathbf{M}*}(\mathbf{x}) = \underset{\mathbf{x_M}}{min}(\epsilon_{\mathbf{M}-O}(r_{\mathbf{M}-O_1}) + \epsilon_{\mathbf{M}-O}(r_{\mathbf{M}-O_2}) + \epsilon_{\mathbf{M}-O}(r_{\mathbf{M}\cdots O'_1}) + \epsilon_{\mathbf{M}-O}(r_{\mathbf{M}\cdots O'_2})). \tag{2.7}$$

Here, $r_{\mathbf{M}-O_i}$ is a metal-oxygen bond distance, and $r_{\mathbf{M}\cdots O'_i}$ is a metal-siloxane coordination distance, as shown in fig. 2.4a. The bond lengths are functions of the (variable) metal atom position $\mathbf{x_M}$ and the (quenched/fixed) peripheral siloxane and silanol locations in x. The optimization indicated in eq. 2.7 therefore involves optimization of the metal atom position within the fixed peripheral environment.

Finally, the free energy of grafting is computed using

$$\Delta G^o(\mathbf{x}) = \Delta E(\mathbf{x}) + \Delta PV - T\Delta S^o. \tag{2.8}$$

Table 2.1: Constants used in computing grafting barriers and defining the quenched disorder lattice (see A.1 of the appendix for further explanations)

Parameter	**Value**
T	298.15K
$r_{\mathbf{M}-O,eq}$	1.0
$r_{\mathbf{M}\cdots O,eq}$	1.16
$\sigma^2_{lattice}$	0.00022
p_{OH}	0.3
$p_{siloxane}$	0.3
p_{empty}	0.4
$D_{\mathbf{M}-O}$	524.4 kJ mol^{-1}
$a_{\mathbf{M}-O}$	1.9
$D_{\mathbf{M}\cdots O}$	120.0 kJ mol^{-1}
$a_{\mathbf{M}\cdots O}$	2.3
$2\epsilon_{\mathbf{HL}} - (V_* + 2\epsilon_{\mathbf{ML}}) + \Delta PV + \Delta S^o$	1229.56 kJ mol^{-1}
M	0.026
α	0.5
$\Delta G^o_{unperturbed}$	-30 kJ mol^{-1}
$\Delta G^{\ddagger}_{ref}$	131.3 kJ mol^{-1}

Here ΔS^o is the entropy of the grafting reaction and $\Delta E(\mathbf{x}) + \Delta PV$ is the enthalpy. The entropy changes are predominantly from site-independent contributions like translational and rotational degrees of freedom of the $\mathbf{ML}_2$ and **HL** species. As noted for the parameter $\Delta G^{\ddagger}_{ref}$, the site-independent terms in eq. 2.8 have no bearing on the results after non-dimensionalization.

2.3 Kernel regression model for grafting barriers

Because *ab initio* calculations are costly, computational studies of catalyst grafting have been based on single sites, or at most a few sites. Ultimately, one hopes to make predictions about grafting across the entire distribution of non-uniform sites. In this section, we propose a machine learning method (kernel regression) to learn structure-property relations from a modest number of training calculations.[86–88] Kernel regression

was chosen because it is a non-parametric method; hence it does not need a predefined form for the fitting function. Specifically, we will use calculations at a small collection of grafting sites to predict barriers and kinetics for all grafting sites.

The training data includes a collection of computed barriers, $\Delta\hat{G}^{\ddagger}(\mathbf{x}_1)$, $\Delta\hat{G}^{\ddagger}(\mathbf{x}_2)$, $\Delta\hat{G}^{\ddagger}(\mathbf{x}_3)$, etc. The estimated barrier for a new peripheral environment $\mathbf{x}$ is a kernel-weighted average of the training data:

$$\Delta\hat{G}^{\ddagger} = \sum_{i=1}^{N_{train}} w(\mathbf{x}, \mathbf{x}_i)\Delta G^{\ddagger}(\mathbf{x}_i). \tag{2.9}$$

Here, $\Delta\hat{G}^{\ddagger}(\mathbf{x})$ is the prediction for a grafting site with local geometry $\mathbf{x}$, $\Delta\hat{G}^{\ddagger}(\mathbf{x}_i)$ values represent the barriers of grafting sites in the training set, N_{train} is the number of training examples, and $w(\mathbf{x}, \mathbf{x}_i)$ are the weights. The weights are represented using a Gaussian kernel:[89]

$$w(\mathbf{x}, \mathbf{x}_i) = \frac{exp[-d^2(\mathbf{x}, \mathbf{x}_i)]}{\sum_{i=1}^{N_{train}} exp[-d^2(\mathbf{x}, \mathbf{x}_i)]}. \tag{2.10}$$

Here $d^2(\mathbf{x}, \mathbf{x}')$ is a squared non-Euclidean Mahalanobis distance between structures $\mathbf{x}$ and $\mathbf{x}$'

$$d^2(\mathbf{x}, \mathbf{x}') = (\mathbf{x} - \mathbf{x}')^T \mathbf{S} (\mathbf{x} - \mathbf{x}'). \tag{2.11}$$

$\mathbf{S}$ is a square, symmetric, and positive definite matrix. To ensure that $\mathbf{S}$ remains positive definite while being optimized/learned, we write $\mathbf{S}$ as

$$\mathbf{S} = \mathbf{A}\mathbf{A}^T. \tag{2.12}$$

Here $\mathbf{A}$ is a lower triangular matrix.[90] Matrix A should be optimized so that eq. 2.9 accurately predicts $\Delta\hat{G}^{\ddagger}(\mathbf{x})$ at new grafting sites. The training data from optimization

of a small collection of grafting sites is used in a leave-one-out objective function

$$L = \sum_{i=1}^{N_{train}} (\Delta G^{\ddagger}(\mathbf{x}_i) - \Delta \hat{G}^{\ddagger}(\mathbf{x}_i))^2. \tag{2.13}$$

to determine $\mathbf{A}$. In L, $\Delta \hat{G}^{\ddagger}(\mathbf{x}_i)$ is a weighted average of all data points in the training set excluding itself:

$$\Delta \hat{G}^{\ddagger}(\mathbf{x}_i) = \sum_{\substack{j=1 \\ j \neq i}}^{N_{train}} \Delta G^{\ddagger}(\mathbf{x}_j) w(\mathbf{x}_i, \mathbf{x}_j). \tag{2.14}$$

The Gaussian kernel in eq. 2.10 generates a continuous and differentiable model of $\Delta \hat{G}^{\ddagger}(\mathbf{x})$, so the leave-one-out error function is easily minimized with conjugate gradient methods or other superlinear minimization schemes.[91] We use kernel regression as implemented in the metric-learn Python library.[92] The library minimizes L using the conjugate gradient method with analytical derivatives of L.

2.4 Local coordinates

The Gaussian kernel function in eq. 2.9 can be constructed from the complete set of internal coordinates for the local environment. However, a subset of the internal coordinates will usually be sufficient to predict the activation barriers. We do not know a priori which coordinates are most important, but these can be identified as illustrated below.

The local environment of silanol and siloxane groups in our model is specified by five coordinates (2 dimensions $\times$ 4 "atoms" - 1 rotation - 2 centre-of-mass translations). We use three of the five coordinates to construct the kernel regression model: (1) distance between OH groups (d1), (2) distance between siloxane groups (d2), and (3) angle between the OH-siloxane groups (θ), fig. 2.4.

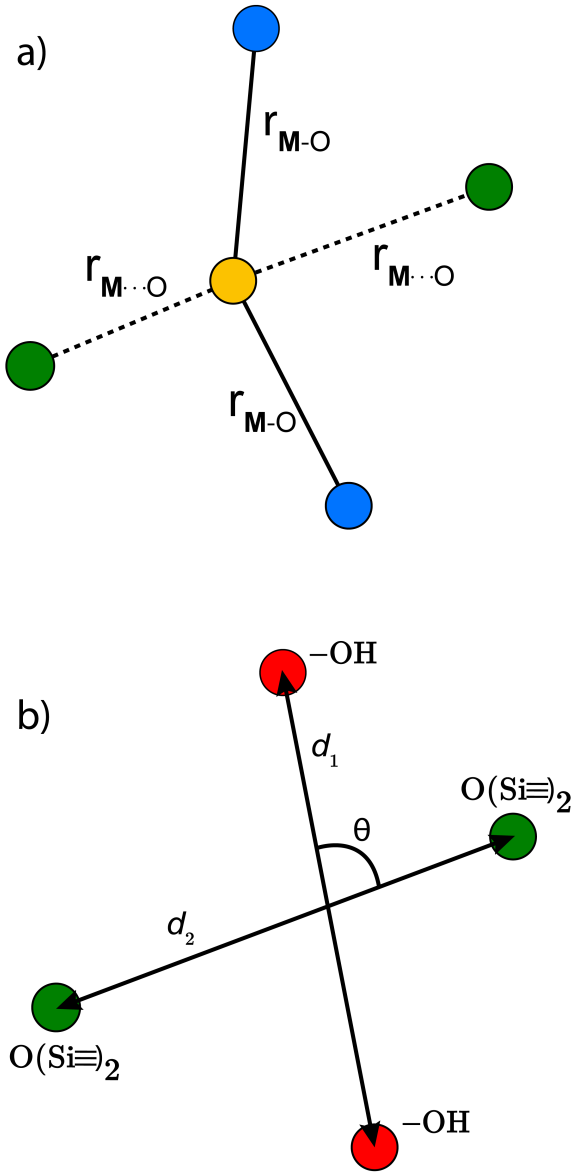


Figure 2.4: Bond lengths in the force field and in the optimization of the **M**-atom position. (b) Coordinates for describing the local environment around the grafting site. We have used three of the five ($2 \times 4 - 1$ (rotation) $- 2$ (translations)) peripheral environment coordinates in the initial kernel regression model.

2.5 Sites with non-uniform grafting barriers: a population balance perspective

As described in Sec. 2.2, an amorphous support will have a distribution of grafting sites with different grafting rates. As time progresses, the most reactive grafting sites will be consumed, while grafting sites with higher reaction barriers remain unreacted and reduce the rate of further grafting. This situation can be modelled using the following population balance scheme:

$$\frac{d\rho(\Delta G^{\ddagger},t)}{dt} = -\mathbf{r}(\Delta G^{\ddagger},m)\rho(\Delta G^{\ddagger},t). \tag{2.15}$$

Here $\rho(\Delta\hat{G}^{\ddagger},t)$ is the population of unreacted vicinal silanol sites at time t with a barrier of $\Delta\hat{G}^{\ddagger}$, $\mathbf{r}(\Delta\hat{G}^{\ddagger},m)$ is the rate at which the sites react (eq. 2.1), and $m = [\mathbf{ML_2}]/\hat{V}_0^{-1}$ is the ratio of the concentration of the precursor $\mathbf{ML_2}$ in the gas phase to the reference concentration ($\hat{V}_0^{-1}$) at which $\Delta\hat{G}^{\ddagger}$ is computed. The rate of change of m is

$$\frac{dm}{dt} = -\int d\Delta G^{\ddagger}\rho(\Delta G^{\ddagger},t)k(\Delta G^{\ddagger},m) + m_G. \tag{2.16}$$

Here, the first term on the right-hand side is rate of consumption of $\mathbf{ML_2}$ due to the grafting reaction, and m_G is the rate at which $\mathbf{ML_2}$ is fed to the reactor. In some grafting experiments, the molecular complex is constantly replenished by evaporation from a reservoir, so that its gas phase concentration is always in equilibrium with its liquid reservoir.[50,58] In such cases, the $\mathbf{ML_2}$ concentration remains constant at its vapor pressure as grafting proceeds. Assuming constant m, eq. 2.15 can be integrated to yield

$$\rho(\Delta G^{\ddagger},t) = \rho_o(\Delta G^{\ddagger})exp[-\frac{k_BT}{h}e^{-\beta\Delta G^{\ddagger}}mt]. \tag{2.17}$$

Here $\rho_0(\Delta\hat{G}^{\ddagger})$ is the initial population of vicinal silanol sites.

Defining non-dimensional time as:

$$\tau = \frac{k_B T}{h} exp[-\beta \Delta G^{\ddagger}_{ref} t]. \tag{2.18}$$

leads to the population of unreacted vicinal silanol sites as a function of τ and grafting free energy barrier:

$$\rho(\Delta G^{\ddagger}, \tau) = \rho_0(\Delta G^{\ddagger}) exp[-e^{-\beta\alpha\Delta G^o} m\tau]. \tag{2.19}$$

2.6 Results and discussion

2.6.1 Evolution of grafting site population

A 1500×1500 lattice was randomly perturbed using the procedure outlined in Sec. 2.2.1. A total of 19368 grafting sites were identified. A metal atom was placed in each grafting site, and its position was optimized. The grafting free energy barrier was computed for each grafting site. A histogram of the results was constructed to approximate the initial distribution $\rho_o(\Delta\hat{G}^{\ddagger})$. Note that the horizontal axis depends on the choice of $\Delta\hat{G}^{\ddagger}_{ref}$ and $\Delta S^o/k_B$, i.e., different values of these parameters will shift the distribution left and right along the $\Delta\hat{G}^{\ddagger}$ axis. In a real system, *ab initio* calculations yield $\Delta\hat{G}^{\ddagger}$ and ΔS^o values for all grafting sites with no adjustable parameters, so there would be no arbitrary shift. Following this, eq. 2.19 was used to compute the evolution of the unreacted grafting site population, fig. 2.5a.

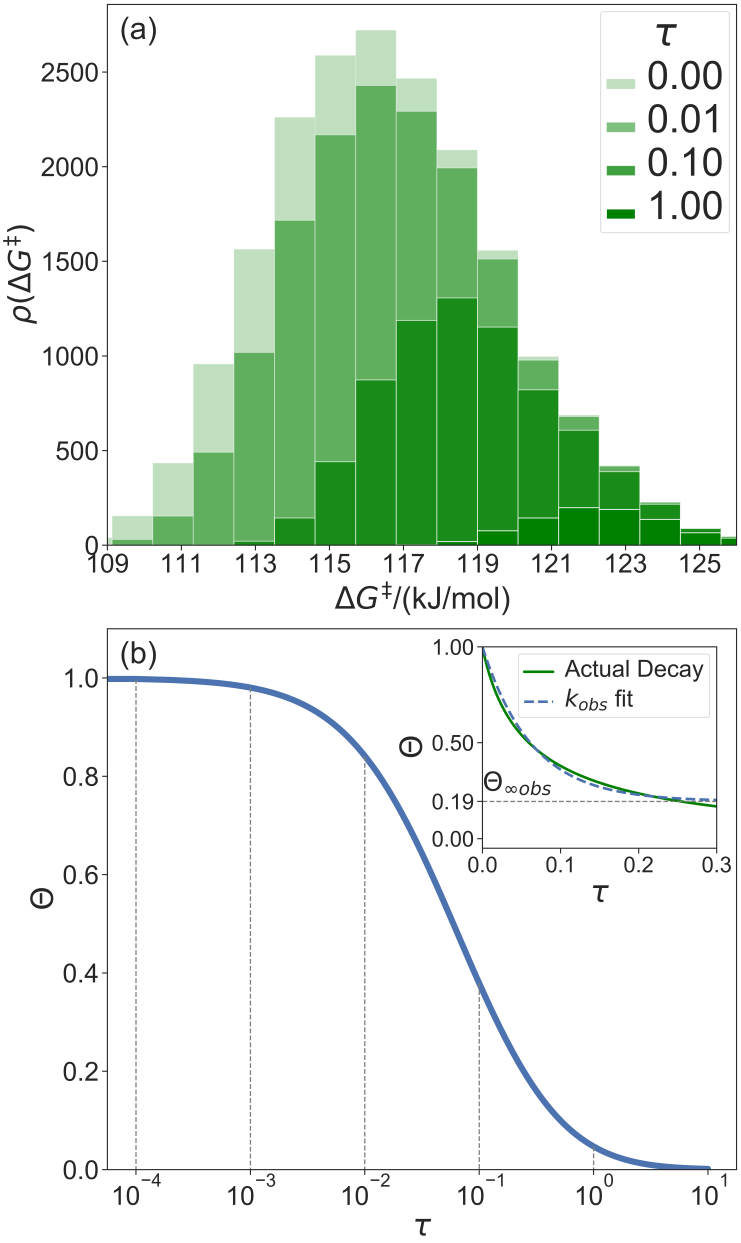


Figure 2.5: (a) Evolution of the unreacted vicinal silanol site population as a function of non-dimensional grafting time. (b) Fraction of unreacted vicinal silanol sites as a function of logarithmic time. The inset shows the evolution as a function of real time in the range $0 < \tau < 0.3$. It also includes an exponential decay model fit to this data.

The initial range of grafting barriers spans 23 kJ mol^{-1}. Grafting sites with the lowest barriers react first, so the distribution shifts to the right as grafting proceeds. The 23

kJ mol^{-1} width of the distribution causes the grafting sites to react at markedly different rates. The fastest grafting sites react in about $10^{-4}\tau$. Grafting is complete in about 10τ.

During a grafting experiment, the total number of grafted sites at any time can be measured, e.g., by monitoring the amount of HL released. The fraction of unreacted vicinal silanol sites (relative to the total number of vicinal silanol sites) is:

$$\Theta = \frac{\int \rho(\Delta G^{\ddagger}, \tau) d\Delta G^{\ddagger}}{\int \rho(\Delta G^{\ddagger}, 0) d\Delta G^{\ddagger}}. \tag{2.20}$$

Fig. 2.5b shows the evolution of the fraction of unreacted vicinal silanols (note the log scale). Grafting progress slows dramatically as the most reactive grafting sites vanish from the distribution. In an experiment, the reaction might seem complete when all the vicinal silanols with low barriers have reacted. The inset of fig. 2.5b shows how the data would appear if the fraction of unreacted silanols were monitored only for time $0 < \tau < 0.3$. The inset also shows a fit to the common pseudo-first-order kinetic model $\Theta = \Theta_{\infty}(1 - \Theta_{\infty})exp(-k_{obs}\tau)$ fraction of unreacted vicinal silanol sites and k_{obs} is an "apparent" grafting rate constant. Over the range $0 < \tau < 0.3$, the data appears to be approximately an exponential decay, thus one might infer that all silanols react with the same rate constant (kobs), and that 19% (from = 0.19) of the vicinal silanol sites are unreactive. However, all of the silanols (in this model) do react at exponentially longer time intervals. The final silanol sites react last because they are different. Therefore, they change the distribution of grafted sites, and may also change the catalytic activity. Hence, it is important to analyse grafting kinetics on a logarithmic time scale.

Predictions about catalytic activity require information about the abundance of grafted sites and their characteristics. Both the grafting kinetics and the catalytic turnover frequency at a particular grafted site depend on the local grafted site environment. However, the most readily grafted sites may not correspond to the most catalytically active sites.

Therefore, predictions of the overall catalyst activity require predictions about grafting propensity and characteristics of the grafted sites. The companion paper develops tools for computing site-averaged kinetics starting from the grafted distribution.[93]

2.6.2 Applying kernel regression to predict grafting barriers

The kernel regression model was trained on grafting barriers for 100 vicinal silanol sites randomly sampled from the set of all 19368 grafting sites using local coordinates described in Sec. 2.4. Justification for choosing a training set size of 100 is provided in section A.2 of the appendix. A parity plot of the true $\Delta\hat{G}^{\ddagger}$ values and kernel regression $\Delta\hat{G}^{\ddagger}$ predictions is shown in fig. 2.6.

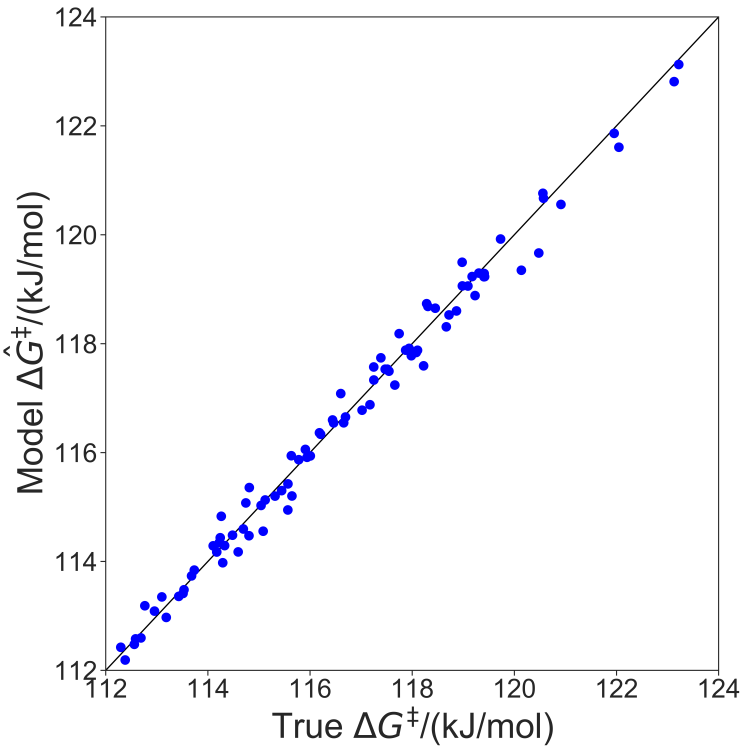


Figure 2.6: Parity plot showing predictions of grafting activation barriers by the kernel regression model trained on 100 grafting sites.

The model trained with 100 $\Delta\hat{G}^{\ddagger}$ calculations was used to predict grafting barriers

for all 19368 grafting sites. After training, the only input information for each grafting site are its values of d_1, d_2, and θ. The residuals of the predictions are plotted as a distribution in fig. 2.7. Nearly all residuals are within ± 1 kJ mol^{-1}, and the standard deviation of the residual distribution is 0.48 kJ mol^{-1}.

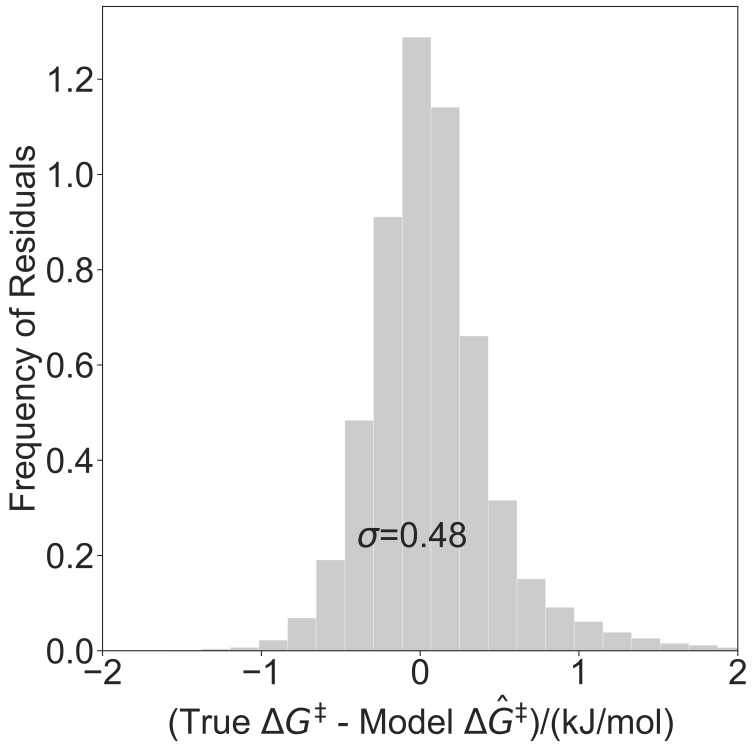


Figure 2.7: Distribution of residuals for a model trained on 100 grafting sites.

2.6.3 Identifying important local coordinates

The results in Sec. 2.6.2 used three of five coordinates to construct the kernel regression model. Three is already a relatively compact structural parameter set, but for this model it can be reduced further. To evaluate the importance of different combinations of local coordinates, the model was retrained by systematically excluding some coordinates. Table 2.2 shows R^2 values for fits with different coordinates. Fig. 2.8 shows parity plots like the one in fig. 2.6, but for a model based only on d1, and for a model based on d_1

Table 2.2: R^2 values of kernel regression models with different combinations of local coordinates)

Coordinates	$\mathbf{R^2}$
θ	-0.02
d_2	0.52
d_1	0.60
d_2, θ	0.52
d_1, θ	0.60
d_1, d_2	0.99
d_1, d_2, θ	0.99

and d_2 (i.e., without θ).

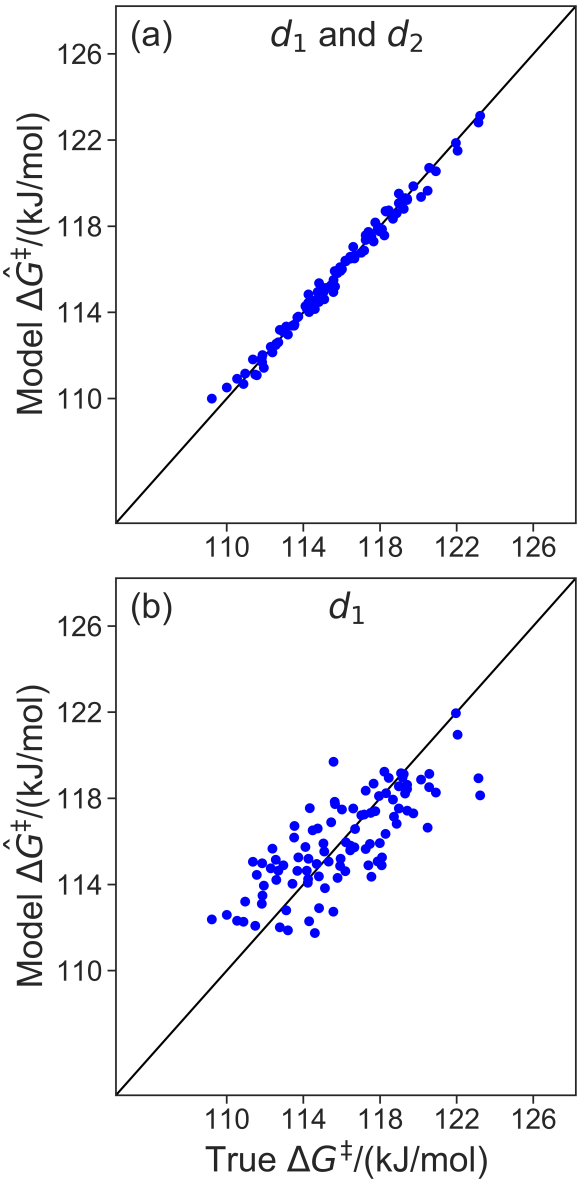


Figure 2.8: (a) Parity plot of the model trained with d_1 and d_2. (b) Parity plot of the model trained with d_1 only.

The model trained using d_1 and d_2 ($R^2 = 0.99$) is comparable in accuracy to the model trained using all coordinates ($R^2 = 0.99$) Clearly, d_1 and d_2 are both important for describing barriers, but θ is inconsequential as its omission does not diminish the accuracy of the kernel regression model. We can also see that the model cannot be further

simplified from d_1 and d_2 dependence. The models trained on only d_1 ($R^2 = 0.60$) and d_2 ($R^2 = 0.52$) have severely diminished accuracy.

Now, using just two coordinates, we can project the grafting free energy barriers onto a 2D plot, fig. 2.9. The barrier decreases monotonically with increasing d_1 or d_2. Therefore, grafting sites with large values of d_1 and d_2 react first, while grafting sites with small values of d_1 and d_2 react more slowly. We emphasize that, even with this simple model, it was not obvious a priori how structural characteristics would influence the grafting kinetics. The procedures in this paper should help to identify features of the most reactive silanol sites.

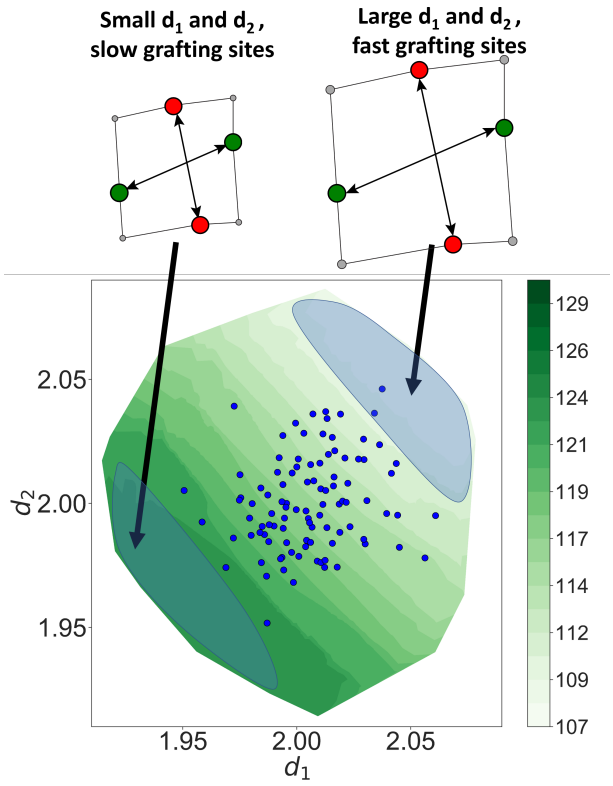


Figure 2.9: Model-predicted barriers as function of d_1 and d_2. Blue dots show training set grafting sites. The figure also shows the structures of active and inactive grafting sites. Grafting sites with smaller values of d_1 and d_2 have larger barriers, while grafting sites with larger values of d_1 and d_2 have smaller barriers.

2.6.4 Predicting the time evolving population of grafting sites

In this section, we recompute results from section 6.1, now using the kernel regression model. We use the model based only on d_1 and d_2 and trained on just 100 randomly sampled grafting sites to predict the evolving population of unreacted silanols. The results are shown in fig. 2.10a.

The training set of 100 grafting sites does not include examples of grafting sites at the extreme fast and slow grafting limits. Accordingly, the trained model does not accurately predict grafting kinetics at the extreme fast and slow limits. Fortunately, the extreme tails account for only a small portion of the total grafting sites, so important properties like the overall grafting progress are still accurately predicted by the model, fig. 2.10b.

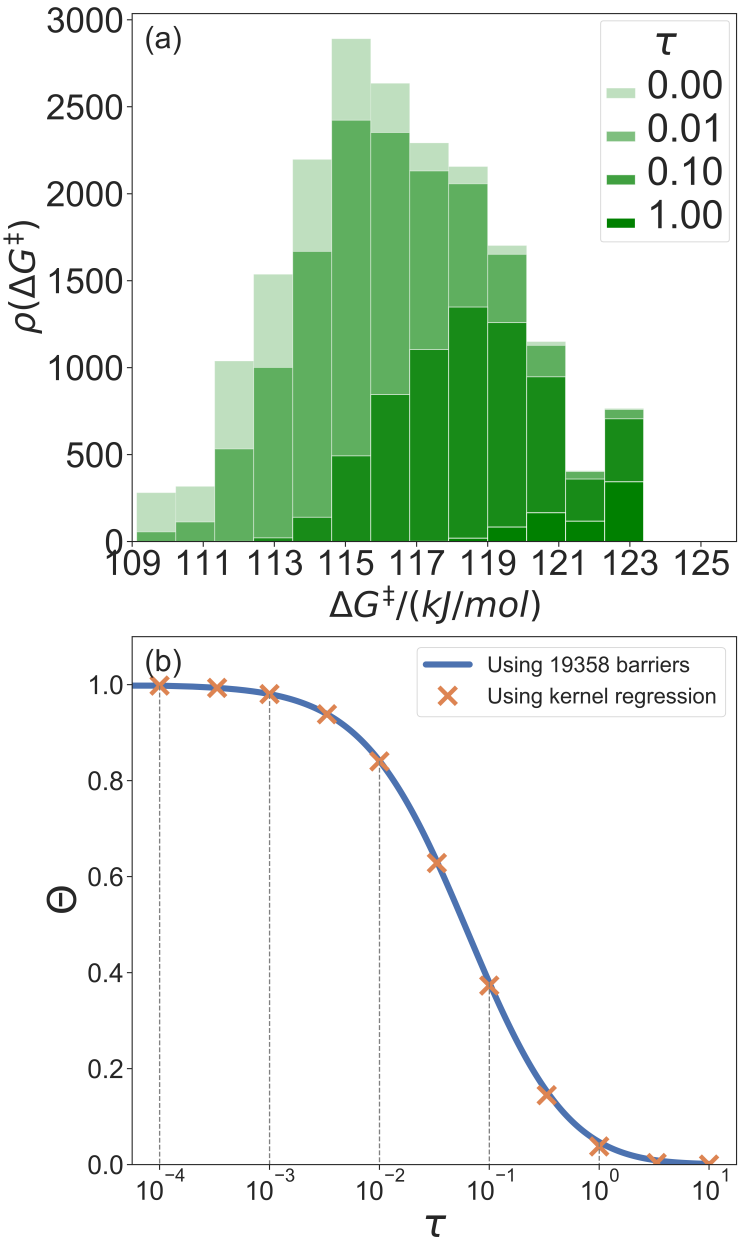


Figure 2.10: (a) Evolving population of grafting sites predicted using a kernel regression model trained on 100 grafting sites. (b) The predicted fraction of unreacted grafting sites as a function of logarithmic time.

2.7 Conclusions

Several factors make *ab initio* rate calculations prohibitively difficult for single-atom catalysts grafted to amorphous supports such as silica. First, the quenched disorder of the support presents an unknown distribution of local environments. Second, grafted site abundances depend on differences in grafting kinetics at different grafting sites. Third, differences between the grafted sites can cause differences in their catalytic activity. Several investigators are working to overcome the first challenge.[71–74] This paper addresses the second challenge by combining transition state theory, kernel regression, and population balance models. A companion paper to this one addresses the third challenge.[93]

To illustrate and test the new methodology, we introduce a simple 2D disordered lattice model of amorphous silica. The model allows us to compute the grafting rate at nearly 20 thousand grafting sites to obtain essentially exact solutions for the evolving grafting/grafted site population during grafting. Then, we trained a kernel regression model to predict grafting rates from a training set of rate calculations at just 100 grafting sites. The regression model predicted barriers with ca. ± 0.5 kJ mol^{-1} accuracy on the test set of about 20 thousand grafting sites. We also showed how the kernel regression results can identify those grafting site characteristics that most strongly influence the grafting kinetics. Finally, the trained kernel regression model was used to predict the evolving population of unreacted silanols.

In future work, we will use this framework with *ab initio* calculations and more realistic silica models to predict grafting rates and active site abundances during the preparation of real single-site catalysts on amorphous silica. Given a model of amorphous silica, the new algorithm should enable quantitative predictions about the grafting process and the grafted site distribution without assuming the characteristics of the most active or most abundant sites.

Bibliography

1. Morokuma, K. & Musaev, D. G. *Computational Modeling for Homogeneous and Enzymatic Catalysis: A Knowledge-Base for Designing Efficient Catalysis* (John Wiley & Sons, 2008).

2. Sperger, T., Sanhueza, I. A., Kalvet, I. & Schoenebeck, F. Computational Studies of Synthetically Relevant Homogeneous Organometallic Catalysis Involving Ni, Pd, Ir, and Rh: An Overview of Commonly Employed DFT Methods and Mechanistic Insights. *Chem. Rev.* **115,** 9532–9586. https://www.doi.org/10.1021/acs.chemrev.5b00163 (2015).

3. Goldsmith, B. R., Hwang, T., Seritan, S., Peters, B. & Scott, S. L. Rate-Enhancing Roles of Water Molecules in Methyltrioxorhenium-Catalyzed Olefin Epoxidation by Hydrogen Peroxide. *J. Am. Chem. Soc.* **137,** 9604–9616. https://www.doi.org/10.1021/jacs.5b03750 (2015).

4. Voth, G. *Computational approaches for studying enzyme mechanism* (Academic Press, 2016).

5. Garcia-Viloca, M., Gao, J., Karplus, M. & Truhlar, D. G. How Enzymes Work: Analysis by Modern Rate Theory and Computer Simulations. *Science* **303,** 186. https://www.doi.org/10.1126/science.1088172 (2004).

6. Tuñón, I. & Moliner, V. *Simulating enzyme reactivity: computational methods in enzyme catalysis* (Royal Society of Chemistry, 2016).

7. Hammer, B. & Nørskov, J. K. in, 71–129 (Academic Press, 2000). https://www.doi.org/10.1016/S0360-0564(02)45013-4.

8. Getman, R. B., Schneider, W. F., Smeltz, A. D., Delgass, W. N. & Ribeiro, F. H. Oxygen-Coverage Effects on Molecular Dissociations at a Pt Metal Surface. *Phys. Rev. Lett.* **102,** 076101. https://www.doi.org/10.1103/PhysRevLett.102.076101 (2009).

9. Greeley, J. Theoretical Heterogeneous Catalysis: Scaling Relationships and Computational Catalyst Design. *Annu Rev Chem Biomol* **7,** 605–635. https://www.doi.org/10.1146/annurev-chembioeng-080615-034413 (2016).

10. Saleheen, M. & Heyden, A. Liquid-Phase Modeling in Heterogeneous Catalysis. *ACS Catal.* **8,** 2188–2194. https://www.doi.org/10.1021/acscatal.7b04367 (2018).

11. Salciccioli, M., Stamatakis, M., Caratzoulas, S. & Vlachos, D. G. A review of multiscale modeling of metal-catalyzed reactions: Mechanism development for complexity and emergent behavior. *Chem. Eng. Sci.* **66,** 4319–4355. https://www.doi.org/10.1016/j.ces.2011.05.050 (2011).

12. Blaszkowski, S. R. & van Santen, R. A. Theoretical study of the mechanism of surface methoxy and dimethyl ether formation from methanol catalyzed by zeolitic protons. *The Journal of Physical Chemistry B* **101,** 2292–2305. https://doi.org/10.1021/jp962006 (1997).

13. Li, Y.-P., Head-Gordon, M. & Bell, A. T. Analysis of the Reaction Mechanism and Catalytic Activity of Metal-Substituted Beta Zeolite for the Isomerization of

Glucose to Fructose. *ACS Catalysis* **4,** 1537–1545. https://www.doi.org/10.1021/cs401054f (2014).

14. Newsome, D. & Coppens, M.-O. Molecular dynamics as a tool to study heterogeneity in zeolites - Effect of Na+ cations on diffusion of CO_2 and N_2 in Na-ZSM-5. *Chemical Engineering Science* **121,** 300–312. https://www.doi.org/10.1016/j.ces.2014.09.024 (2015).

15. McFarland, E. W. & Metiu, H. Catalysis by Doped Oxides. *Chemical Reviews* **113,** 4391–4427. https://www.doi.org/10.1021/cr300418s (2013).

16. Le Bahers, T., Rérat, M. & Sautet, P. Semiconductors Used in Photovoltaic and Photocatalytic Devices: Assessing Fundamental Properties from DFT. *The Journal of Physical Chemistry C* **118,** 5997–6008. https://www.doi.org/10.1021/jp409724c (2014).

17. Li, F., Luo, S., Sun, Z., Bao, X. & Fan, L.-S. Role of metal oxide support in redox reactions of iron oxide for chemical looping applications: experiments and density functional theory calculations. *Energy & Environmental Science* **4,** 3661–3667. https://www.doi.org/10.1039/C1EE01325D (2011).

18. Knott, B. C. *et al.* The Mechanism of Cellulose Hydrolysis by a Two-Step, Retaining Cellobiohydrolase Elucidated by Structural and Transition Path Sampling Studies. *J. Am. Chem. Soc.* **136,** 321–329. https://www.doi.org/10.1021/ja410291u (2014).

19. Masgrau, L. & Truhlar, D. G. The Importance of Ensemble Averaging in Enzyme Kinetics. *Acc. Chem. Res.* **48,** 431–438. https://www.doi.org/10.1021/ar500319e (2015).

20. Brunk, E., Arey, J. S. & Rothlisberger, U. Role of Environment for Catalysis of the DNA Repair Enzyme MutY. *J. Am. Chem. Soc.* **134,** 8608–8616. https://www.doi.org/10.1021/ja301714j (2012).

21. Peters, B. & Scott, S. L. Single Atom Catalysts on Amorphous Supports: A Quenched Disorder Perspective. *J. Chem. Phys.* **142,** 104708. https://www.doi.org/10.1063/1.4914145 (2015).

22. Goldsmith, B. R., Sanderson, E. D., Bean, D. & Peters, B. Isolated Catalyst Sites on Amorphous Supports: A Systematic Algorithm for Understanding Heterogeneities in Structure and Reactivity. *J. Chem. Phys.* **138,** 204105. https://www.doi.org/10.1063/1.4807384 (2013).

23. McDaniel, M. P. A Review of the Phillips Supported Chromium Catalyst and Its Commercial Use for Ethylene Polymerization. English. *Adv. Catal.* **53,** 123–606. https://www.doi.org/https://doi.org/10.1016/S0360-0564(10)53003-7 (2010).

24. Coperet, C. *et al.* Surface Organometallic and Coordination Chemistry toward Single-Site Heterogeneous Catalysts: Strategies, Methods, Structures, and Activities. *Chem. Rev.* **116,** 323–421. https://www.doi.org/10.1021/acs.chemrev.5b00373 (2016).

25. Buijink, J. K. F., van Vlaanderen, J. J. M., Crocker, M. & Niele, F. G. M. Propylene epoxidation over titanium-on-silica catalyst–the heart of the SMPO process. *Catal. Today* **93-95,** 199–204. https://www.doi.org/10.1016/j.cattod.2004.06.041 (2004).

26. Delley, M. F. *et al.* Proton Transfers Are Key Elementary Steps in Ethylene Polymerization on Isolated Chromium(III) Silicates. *Proc. Natl. Acad. Sci.* **111,** 11624–11629. https://www.doi.org/10.1073/pnas.1405314111 (2014).

27. Floryan, L., Borosy, A. P., Nunez-Zarur, F., Comas-Vives, A. & Coperet, C. Strain Effect and Dual Initiation Pathway in Cr(III)/SiO_2 Polymerization Catalysts from Amorphous Periodic Models. English. *J. Catal.* **346,** 50–56. https://www.doi.org/10.1016/j.jcat.2016.11.037 (2017).

28. Espelid, Ø. & Børve, K. J. Molecular-Level Insight into Cr/Silica Phillips-Type Catalysts: Polymerization-Active Dinuclear Chromium Sites. *J. Catal.* **206,** 331–338. https://www.doi.org/10.1006/jcat.2001.3499 (2002).

29. Fong, A., Vandervelden, C., Scott, S. L. & Peters, B. Computational Support for Phillips Catalyst Initiation via Cr-C Bond Homolysis in a Chromacyclopentane Site. *ACS Catal.* **8,** 1728–1733. https://doi.org/10.1021/acscatal.7b03724 (2018).

30. Fong, A., Yuan, Y., Ivry, S. L., Scott, S. L. & Peters, B. Computational Kinetic Discrimination of Ethylene Polymerization Mechanisms for the Phillips (Cr/SiO_2) Catalyst. English. *ACS Catal.* **5,** 3360–3374. https://www.doi.org/10.1021/acscatal.5b00016 (2015).

31. Gierada, M. & Handzlik, J. Computational insights into reduction of the Phillips CrO_x/SiO_2 catalyst by ethylene and CO. *J. Catal.* **359,** 261–271. https://www.doi.org/10.1016/j.jcat.2018.01.014 (2018).

32. Guesmi, H. & Tielens, F. Chromium oxide species supported on silica: A representative periodic DFT model. *J Phys Chem C* **116,** 994–1001. https://doi.org/10.1021/jp209680r (2012).

33. Ewing, C. S. *et al.* Impact of Support Interactions for Single-Atom Molybdenum Catalysts on Amorphous Silica. *Ind Eng Chem Res* **55,** 12350–12357. https://www.doi.org/10.1021/acs.iecr.6b03558 (2016).

34. Jystad, A. M., Biancardi, A. & Caricato, M. Simulations of Ammonia Adsorption for the Characterization of Acid Sites in Metal-Doped Amorphous Silicates. *J. Phys. Chem. C* **121,** 22258–22267. https://www.doi.org/10.1021/acs.jpcc.7b08113 (2017).

35. Vansant, E. E., Van Der Voort, P. & Vrancken, K. C. *Characterization and Chemical Modification of the Silica Surface* (1995).

36. Brinker, C. J. & Scherer, G. W. *Sol-gel science: the physics and chemistry of sol-gel processing* (Academic press, 1990).

37. Bergna, H. E. & Roberts, W. O. *Colloidal silica: fundamentals and applications* (CRC Press, 2005).

38. Brückner, R. Properties and structure of vitreous silica. I. *J. Non-Cryst. Solids* **5,** 123–175. https://www.doi.org/10.1016/0022-3093(70)90190-0 (1970).

39. Brückner, R. Properties and structure of vitreous silica. II. *J. Non-Cryst. Solids* **5,** 177–216. https://www.doi.org/10.1016/0022-3093(71)90032-9 (1971).

40. Flörke, O. W. *et al.* in *Ullmann's Encyclopedia of Industrial Chemistry* (John Wiley & Sons, 2000).

41. Atiqullah, M. *et al. Influence of silica calcination temperature on the performance of supported catalyst* $SiO_2-nBuSnCl_3/MAO/(nBuCp)_2ZrCl_2$ *polymerizing ethylene without separately feeding the MAO cocatalyst* 2007. https://www.doi.org/10.1016/j.apcata.2007.01.023.

42. Oschatz, M. *et al.* Effects of calcination and activation conditions on ordered mesoporous carbon supported iron catalysts for production of lower olefins from synthesis gas. *Catal Sci Technol* **6,** 8464–8473. https://www.doi.org/10.1039/C6CY01251E (2016).

43. Rimola, A., Costa, D., Sodupe, M., Lambert, J.-F. & Ugliengo, P. Silica Surface Features and Their Role in the Adsorption of Biomolecules: Computational Modeling and Experiments. *Chem. Rev.* **113,** 4216–4313. https://www.doi.org/10.1021/cr3003054 (2013).

44. Zhuravlev, L. T. The surface chemistry of amorphous silica. Zhuravlev model. *Colloids Surf., A* **173,** 1–38. https://www.doi.org/10.1016/S0927-7757(00)00556-2 (2000).

45. Sharma, S. K., Mammone, J. F. & Nicol, M. F. Raman investigation of ring configurations in vitreous silica. *Nature* **292,** 140–141. https://www.doi.org/10.1038/292140a0 (1981).

46. Brinker, C. J., Brow, R. K., Tallant, D. R. & Kirkpatrick, R. J. Surface structure and chemistry of high surface area silica gels. *J. Non-Cryst. Solids* **120,** 26–33. https://www.doi.org/10.1016/0022-3093(90)90187-Q (1990).

47. Humbert, B., Burneau, A., Gallas, J. P. & Lavalley, J. C. Origin of the Raman bands, D1 and D2, in high surface area and vitreous silicas. *J. Non-Cryst. Solids* **143,** 75–83. https://www.doi.org/10.1016/S0022-3093(05)80555-1 (1992).

48. Peri, J. B. & Hensley, A. L. The surface structure of silica gel. *The Journal of Physical Chemistry* **72,** 2926–2933. https://www.doi.org/10.1021/j100854a041 (1968).

49. Taha, Z. A., Deguns, E. W., Chattopadhyay, S. & Scott, S. L. Formation of Digallium Sites in the Reaction of Trimethylgallium with Silica. *Organometallics* **25,** 1891–1899. ISSN: 0276-7333. https://doi.org/10.1021/om051034o (2006).

50. Fleischman, S. D. & Scott, S. L. Evidence for the Pairwise Disposition of Grafting Sites on Highly Dehydroxylated Silicas via Their Reactions with $Ga(CH_3)_3$. *J. Am. Chem. Soc.* **133,** 4847–4855. https://www.doi.org/10.1021/ja108905p (2011).

51. Abdillah, B., Rice, G. L. & Scott, S. L. *Mono- and Dinuclear Silica-Supported Titanium(IV) Complexes and the Effect of TiOTi Connectivity on Reactivity* 1999. https://www.doi.org/10.1021/JA9829160.

52. Gao, X., Bare, S. R., Fierro, J. L. G., Banares, M. A. & Wachs, I. E. Preparation and in-Situ Spectroscopic Characterization of Molecularly Dispersed Titanium Oxide on Silica. *The Journal of Physical Chemistry B* **102,** 5653–5666. https://www.doi.org/10.1021/jp981423e (1998).

53. Kytökivi, A. & Haukka, S. Reactions of HMDS, $TiCl_4$, $ZrCl_4$, and $AlCl_3$ with Silica As Interpreted from Low-Frequency Diffuse Reflectance Infrared Spectra. *The Journal of Physical Chemistry B* **101,** 10365–10372. https://www.doi.org/10.1021/jp971822u (1997).

54. Deguns, E. W., Taha, Z., Meitzner, G. D. & Scott, S. L. An X-ray Absorption Study of Two VOCl3-Modified Silicas: Evidence for Chloride-Silica Interactions. *The Journal of Physical Chemistry B* **109,** 5005–5011. https://www.doi.org/10.1021/jp045886y (2005).

55. Zhu, H. *et al.* VO_x/SiO_2 Catalyst Prepared by Grafting $VOCl_3$ on Silica for Oxidative Dehydrogenation of Propane. *ChemCatChem* **7,** 3332–3339. https://www.doi.org/10.1002/cctc.201500607 (2015).

56. Jarupatrakorn, J. & Tilley, T. D. Silica-Supported, Single-Site Titanium Catalysts for Olefin Epoxidation. A Molecular Precursor Strategy for Control of Catalyst Structure. *J. Am. Chem. Soc.* **124,** 8380–8388. https://www.doi.org/10.1021/ja0202208 (2002).

57. Fraile, J. M., García, J. I., Mayoral, J. A. & Vispe, E. Catalytic sites in silica-supported titanium catalysts: silsesquioxane complexes as models. *J. Catal.* **233,** 90–99. https://www.doi.org/10.1016/j.jcat.2005.04.018 (2005).

58. Zhong, L. *et al.* Spectroscopic and structural characterization of Cr(II)/SiO_2 active site precursors in model Phillips polymerization catalysts. *J. Catal.* **293,** 1–12. https://www.doi.org/10.1016/j.jcat.2012.05.014 (2012).

59. Fukudome, K., Ikenaga, N.-o., Miyake, T. & Suzuki, T. Oxidative dehydrogenation of propane using lattice oxygen of vanadium oxides on silica. *Catal Sci Technol* **1,** 987–998. https://www.doi.org/10.1039/C1CY00115A (2011).

60. Samantaray, M. K. *et al.* Surface organometallic chemistry in heterogeneous catalysis. *Chem. Soc. Rev.* **47,** 8403–8437. https://www.doi.org/10.1039/C8CS00356D (2018).

61. Handzlik, J., Grybos, R. & Tielens, F. Structure of Monomeric Chromium(VI) Oxide Species Supported on Silica: Periodic and Cluster DFT Studies. *The Journal of Physical Chemistry C* **117,** 8138–8149. https://www.doi.org/10.1021/jp3103035 (2013).

62. Cavalleri, M. *et al.* Analysis of silica-supported vanadia by X-ray absorption spectroscopy: Combined theoretical and experimental studies. *J. Catal.* **262,** 215–223. https://www.doi.org/10.1016/j.jcat.2008.12.013 (2009).

63. Khaliullin, R. Z. & Bell, A. T. A Density Functional Theory Study of the Oxidation of Methanol to Formaldehyde over Vanadia Supported on Silica, Titania, and Zirconia. *The Journal of Physical Chemistry B* **106,** 7832–7838. https://www.doi.org/10.1021/jp014695h (2002).

64. Wang, T.-H. *et al.* The surprising oxidation state of fumed silica and the nature of water binding to silicon oxides and hydroxides. *Chem. Phys. Lett.* **501,** 159–165. https://www.doi.org/10.1016/j.cplett.2010.11.013 (2011).

65. Jaegers, N. R. *et al.* Investigation of Silica-Supported Vanadium Oxide Catalysts by High-Field 51V Magic-Angle Spinning NMR. *The Journal of Physical Chemistry C* **121,** 6246–6254. https://www.doi.org/10.1021/acs.jpcc.7b01658 (2017).

66. Handzlik, J. Properties and metathesis activity of monomeric and dimeric Mo centres variously located on γ-alumina - A DFT study. *Surf. Sci.* **601,** 2054–2065. https://www.doi.org/10.1016/J.SUSC.2007.03.002 (2007).

67. Chung, L. W. *et al.* The ONIOM Method and Its Applications. *Chem. Rev.* **115,** 5678–5796. https://www.doi.org/10.1021/cr5004419 (2015).

68. Goumans, T. P. M., Catlow, C. R. A. & Brown, W. A. Hydrogenation of CO on a silica surface: An embedded cluster approach. *J. Chem. Phys.* **128,** 134709. https://www.doi.org/10.1063/1.2888933 (2008).

69. Handzlik, J. & Ogonowski, J. Structure of Isolated Molybdenum(VI) and Molybdenum(IV) Oxide Species on Silica: Periodic and Cluster DFT Studies. *The Journal of Physical Chemistry C* **116,** 5571–5584. https://www.doi.org/10.1021/jp207385h (2012).

70. Gomes, J., Zimmerman, P. M., Head-Gordon, M. & Bell, A. T. Accurate Prediction of Hydrocarbon Interactions with Zeolites Utilizing Improved Exchange-Correlation Functionals and QM/MM Methods: Benchmark Calculations of Adsorption Enthalpies and Application to Ethene Methylation by Methanol. *The Journal of Physical Chemistry C* **116,** 15406–15414. https://www.doi.org/10.1021/jp303321s (2012).

71. Ewing, C. S., Bhavsar, S., Veser, G., McCarthy, J. J. & Johnson, J. K. Accurate amorphous silica surface models from first-principles thermodynamics of surface dehydroxylation. *Langmuir* **30,** 5133–5141. https://doi.org/10.1021/la500422p (2014).

72. Ugliengo, P. *et al.* Realistic models of hydroxylated amorphous silica surfaces and MCM- 41 mesoporous material simulated by large-scale periodic B3LYP calculations. *Adv. Mater.* **20,** 4579–4583. https://www.doi.org/10.1002/adma.200801489 (2008).

73. Comas-Vives, A. Amorphous SiO_2 surface models: Energetics of the dehydroxylation process, strain, ab initio atomistic thermodynamics and IR spectroscopic signatures. *Phys. Chem. Chem. Phys.* **18,** 7475–7482. https://www.doi.org/10.1039/c6cp00602g (2016).

74. Tielens, F., Gervais, C., Lambert, J. F., Mauri, F. & Costa, D. Ab initio study of the hydroxylated surface of amorphous silica: A representative model. *Chem. Mater.* **20,** 3336–3344. https://www.doi.org/10.1021/cm8001173 (2008).

75. Peters, B., Scott, S. L., Fong, A., Wang, Y. & Stiegman, A. E. Reexamining the Evidence for Proton Transfers in Ethylene Polymerization. *Proc. Natl. Acad. Sci. USA* **112,** E4160–E4161. https://www.doi.org/10.1073/pnas.1422589112 (2015).

76. Peek, N. M. *et al.* Reassessment of the Electronic Structure of Cr(VI) Sites Supported on Amorphous Silica and Implications for Cr Coordination Number. *The Journal of Physical Chemistry C* **122,** 4349–4358. https://www.doi.org/10.1021/acs.jpcc.7b12079 (2018).

77. Lwin, S., Li, Y., Frenkel, A. I. & Wachs, I. E. Nature of WOx Sites on SiO_2 and Their Molecular Structure-Reactivity/Selectivity Relationships for Propylene

Metathesis. *ACS Catal.* **6,** 3061–3071. https://www.doi.org/10.1021/acscatal.6b00389 (2016).

78. Peters, B. *Reaction Rate Theory and Rare Events* (Elsevier, Amsterdam ; Cambrige, MA, 2017).

79. Nørskov, J. K., Studt, F., Abild-Pedersen, F. & Bligaard, T. *Fundamental Concepts in Heterogeneous Catalysis* (Wiley, 2014).

80. Peters, B. Common Features of Extraordinary Rate Theories. *The Journal of Physical Chemistry B* **119,** 6349–6356. https://www.doi.org/10.1021/acs.jpcb.5b02547 (2015).

81. Chorkendorff, I. & Niemantsverdriet, J. W. *Concepts of modern catalysis and kinetics* (John Wiley & Sons, 2017).

82. Boudart, M. & Djega-Mariadassou, G. *Kinetics of Heterogeneous Catalytic Reactions* (Princeton University Press, Princeton, 2014).

83. Van Santen, R. A. & Neurock, M. *Molecular heterogeneous catalysis: a conceptual and computational approach* (John Wiley & Sons, 2009).

84. Jensen, F. *Introduction to computational chemistry* (John wiley & sons, 2017).

85. Leffler, J. E. Parameters for the Description of Transition States. *Science* **117,** 340–341. https://www.doi.org/10.1126/science.117.3039.340 (1953).

86. Mahalanobis, P. C. On the generalized distance in statistics. *Proc. Natl. Inst. Sci. India* **2,** 49–55. https://doi.org/10.1007/s13171-019-00164-5 (1936).

87. Hofmann, T., Schölkopf, B. & Smola, A. J. Kernel methods in machine learning. *Ann. Stat.* **36,** 1171–1220. https://www.doi.org/10.1214/009053607000000677 (2008).

88. Mohri, M., Rostamizadeh, A. & Talwalkar, A. *Foundations of machine learning* (MIT press, 2018).

89. Weinberger, K. Q. & Tesauro, G. Metric Learning for Kernel Regression. *J. Mach. Learn. Res.,* 8. http://proceedings.mlr.press/v2/weinberger07a.html (2007).

90. Strang, G. & Strang, S. *Linear Algebra and Its Applications* (Thomson, Brooks/Cole, 2006).

91. Nocedal, J. & Wright, S. *Numerical optimization* (Springer Science & Business Media, 2006).

92. Carey, C. J. & Tang, Y. *metric-learn.* https://metric-learn.github.io/metric-learn/ (2015).

93. Vandervelden, C. A., Khan, S. A., Scott, S. L. & Peters, B. Site-averaged kinetics for catalysts on amorphous supports: an importance learning algorithm. *React. Chem. Eng.* **5,** 77–86. https://www.doi.org/10.1039/C9RE00356H (2020).

Chapter 3

Site-averaged kinetics for catalysts on amorphous supports: An importance learning algorithm

Reproduced in part with permission from: Vandervelden, C. A.; Khan, S. A.; Scott, S. L.; Peters, B. B. Siteaveraged kinetics for catalysts on amorphous supports: an importance learning algorithm. https://doi.org/10.1039/C9RE00356H

3.1 Introduction

A recent surge of interest in atomically-dispersed "single atom" catalysts is driven by their unique and potentially selective reactivity,[1–3] and by sustainability efforts that seek to minimize use of scarce elements and maximize atom economy.[4–6] Among single atom catalysts, those which are chemically bonded to a thermally robust oxide support like silica are especially resistant to deactivation by sintering.[7,8] Moreover, grafting strategies that promote selective reaction of the catalyst precursor at specific surface sites may

help to minimize differences between grafted metal sites. Well-studied catalysts that are comprised of single metal atoms grafted onto amorphous silica include chromocenes or chromates for olefin polymerization,[9–11] titanium and tantalum complexes for olefin epoxidation,[12] molybdates for methanol dehydration,[13] and vanadates for partial oxidation of methanol.[14]

Investigators have occasionally drawn comparisons between the metal atoms present in the active sites in enzymes, and metal atoms grafted onto silica surfaces.[12] There are similarities, but there are also important differences. Each enzyme molecule of a given type is the same, while each metal atom on amorphous silica resides in a unique ligand environment. These non-uniform environments can result in metal atoms with non-uniform catalytic properties, including a range of activities, selectivities, adsorption constants, and even different spectroscopic features. When the sites have variable activities, a minority of the sites may contribute most of the overall catalyst activity. Indeed, active site counting experiments confirm that only a small fraction of sites in a heterogeneous catalyst is typically active.[15–18] This poses an extraordinary difficulty in experimental as well as theoretical studies of these catalysts. Powerful characterization tools (NMR, EXAFS, IR, Raman, etc.) generally provide the strongest signals for the most common sites, and these are likely inactive.[11]

If we could understand the mechanisms of these catalysts, we might systematically work to improve them.[19] In some applications like olefin polymerization, where the catalysts are not recovered from the polymer product, one might even use mechanistic understanding to design catalysts with a desired activity distribution capable of generating polymer with a desired molecular weight distribution.

Our first paper introduced a method to predict the distribution of sites that emerges from grafting a precursor onto an amorphous support.[20] The simple model system consisted of a quenched-disordered lattice (to represent the amorphous silica support), sur-

face functional groups (representing pairs or nests of hydroxyl groups) to which a metal complex can be grafted, and a microkinetic model for each grafted site with rate parameters that depend on the site characteristics. Much like in an *ab initio* study, computing activation barriers for the model system requires geometry optimizations of intermediates. Our realistic but simple model allowed us to focus on developing the importance sampling and machine learning tools, without being distracted by controversies about the mechanisms about the mechanisms of these catalysts.

Starting from the simple model and the grafted site population described in our first paper,[20] this second paper aims to compute an average over sites to predict the overall kinetics. Since the turnover frequencies at individual sites vary exponentially with the activation energy, even a small variance in the activation energy leads to an enormous variance in site-specific activities. Such exponential averages are notoriously difficult to converge with standard sampling tools,[21–23] but importance sampling methods can dramatically accelerate convergence. The ideal importance sampling algorithm[24] requires activation energies for each site, but these activation energies are not known a priori. Each activation energy must be obtained through costly ab initio calculations. Because of this limitation, typical approaches calculate just one[25,26] or a small handful of sites[27–34] – far too few to converge site-averaged predictions of kinetic properties.[35] Kernel regression tools can use a modest set of *ab initio* calculations to predict activation energies that have not actually been computed. This paper shows how importance sampling and machine learning can be combined to generate site-averaged predictions efficiently.

In the remainder of this paper, we discuss model elementary steps and a rate law for a catalytic reaction with our simple model system. We briefly review the kernel regression tools (from our first paper) that predict activation energies. We combine the importance sampling and kernel regression tools into an "importance learning algorithm". We then use the new algorithm to identify characteristics of highly active sites and to estimate site-

averaged activation energies. Finally, we compare the efficiency of importance sampling estimates to straightforward sampling.

3.2 Model for amorphous support and grafted sites

Our previous paper described the creation of a disordered, functionalized lattice model to approximate the non-uniform silanol sites and siloxane environments on the surface of amorphous SiO_2. That paper also considered a kinetic model for grafting of metal atoms onto the silanol sites. All sites with two silanol neighbors and two siloxane neighbors on opposite sides were eligible grafting sites in the model. A schematic of the simple model is shown in fig. 3.1.

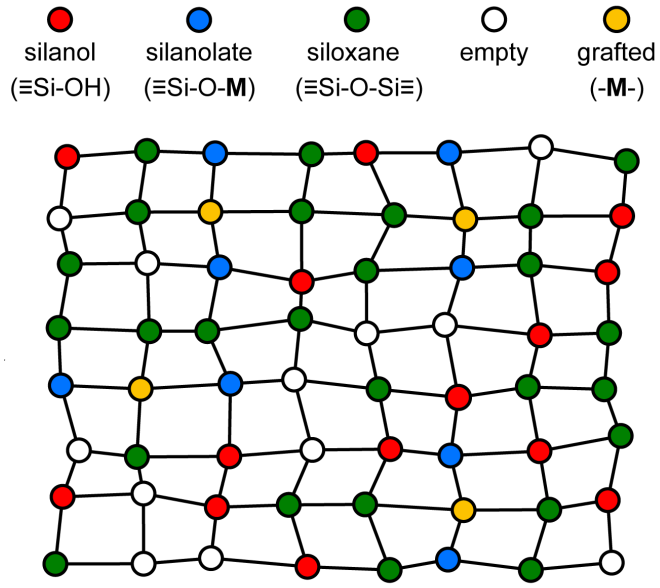


Figure 3.1: Quenched disorder lattice model. Sites with a grafted metal center are shown in gold.

This paper uses the distribution of non-uniform grafted sites, like those shown in fig. 3.1, as its starting point. We assume that grafting has occurred at all eligible sites, but

one could modify the starting distribution (using methods in the previous paper[20]) to investigate lower catalyst loadings.

The discussion below invokes bonds between metal atoms and adsorbates, as well as the oxygen atoms of the silanol and siloxane sites. However, the local environment of each site (before grafting and during catalysis) is described entirely by the positions of atoms in the silica support surrounding the metal center. The coordinates used to describe the local environment are shown in fig. 3.2 They are: (i) the distance between siloxane groups, d_1, (ii) the distance between silanolate groups, d_2, and (iii) the angle between the silanolate and siloxane groups, θ.

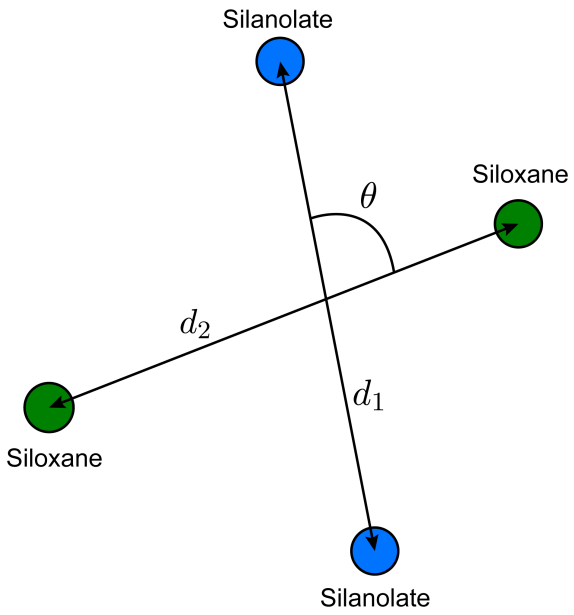


Figure 3.2: Coordinates used to describe the local environment of a grafting site.

The selected coordinates are nearly orthogonal in the sense that their gradients have little or no overlap. Note, however, that the coordinates are incomplete. For a grafting site on our two dimensional surface model, the four nearest neighbors are fully described by five internal coordinates ($8 - 2 \times$ (center of mass) $- 1 \times$ (rotation)). We use only three

coordinates in the kernel regression model, and the results below will show that just two of these coordinates are sufficient to predict site-averaged kinetics. We also emphasize that some calculations below involve other coordinates at intermediate stages, but that the overall kinetics and the kinetics of individual sites ultimately depends only on the coordinates in fig. 3.2.

3.3 Model for catalysis at grafted sites

We consider a simple model of a catalyst site, $\mathbf{M}^*$, comprised of a metal center M and its surrounding support environment, *. We will consider the case in which the catalytic reaction at each site has the same rate-limiting step and the same most abundant surface intermediate (MASI). We further assume that the site does not deactivate. The model reaction has a simple Langmuir-Hinshelwood mechanism:

$$\mathbf{A} + \mathbf{M}^* \rightleftharpoons \mathbf{AM}^* \quad K(\mathbf{x})$$
$$\mathbf{AM}^* \rightarrow \mathbf{B} + \mathbf{M}^* \quad k_2$$

We further assume that

1. the equilibrium constant K for adsorption of reactant A depends on the local environment of site i, $\mathbf{x}_i$,
2. the adsorbed molecule $\mathbf{A}$ ($\mathbf{AM}^*$) is irreversibly converted into the gas phase product B and a bare site $\mathbf{M}^*$,
3. $K(\mathbf{x})c_{\mathbf{A}} \ll 1$ for all sites, so that the bare site is the MASI.

The bond strengths chosen in this work, described below, ensure that these three

assumptions are true for all sites. Fig. 3.3 depicts the Langmuir-Hinshelwood mechanism and the three simplifying assumptions for the simple model system in this work.

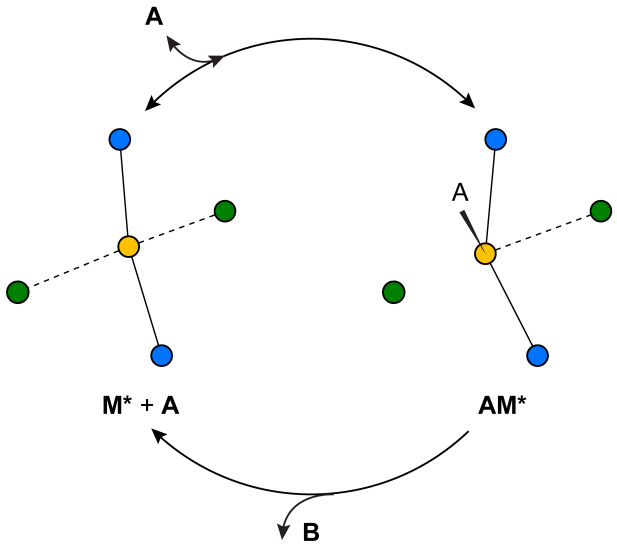


Figure 3.3: The equilibrated adsorption step and irreversible chemical reaction steps for the model reaction $\mathbf{A} \rightarrow \mathbf{B}$, and the $\mathbf{M}^*$ sites described in this work.

The Langmuir-Hinshelwood mechanism leads to a rate law of form

$$\dot{\imath} = \frac{k_2 K(\mathbf{x}) c_{\mathbf{A}}}{1 + K(\mathbf{x}) c_{\mathbf{A}}} \tag{3.1}$$

Because of assumption 3., the rate law simplifies to a power law rate expression of the form

$$\dot{\imath} = k(\mathbf{x}) c_{\mathbf{A}} \tag{3.2}$$

where the pseudo-first-order rate constant is:

$$k(\mathbf{x}) = k_2 K(\mathbf{x}) \tag{3.3}$$

Note that we have also assumed that the rate constant k_2 for the second step in the Langmuir-Hinshelwood mechanism is the same for all sites. In principal, k_2 could also

depend on $\mathbf{x}$, but a model for $k_2(\mathbf{x})$ would require additional parameters to create a model for the saddle region on the potential energy surface. The more elaborate model system with $\mathbf{x}$-dependence in k_2 would still lead to an apparent rate constant k_2K that is one function of $\mathbf{x}$.

The apparent rate constant $k(\mathbf{x})$ depends on the local site geometry through $K(\mathbf{x})$. The adsorption constant is

$$K(\mathbf{x}) = \exp\left[-\frac{\Delta H(\mathbf{x}) - T\Delta S}{k_B T}\right] \tag{3.4}$$

where $\Delta H(\mathbf{x})$ is a site-dependent adsorption enthalpy. $T\Delta S$ is assumed to be constant because its main contributions are the loss of translational and rotational freedom upon adsorption. The rate constant, according to transition state theory, will be of the form

$$k_2 = \frac{k_B T}{h} \exp\left[\frac{\Delta S^{\ddagger}}{k_B}\right] \exp\left[-\frac{\Delta H^{\ddagger}}{k_B T}\right] \tag{3.5}$$

Here, the entropy and enthalpy of activation for the reaction step are assumed to be the same for all sites.

The site-dependent enthalpy of adsorption, $\Delta H(\mathbf{x})$, is modeled by

$$\Delta H(\mathbf{x}) = V_{\mathbf{AM}*}(\mathbf{x}) - (V_{\mathbf{M}*}(\mathbf{x}) + \varepsilon_{\mathbf{A}} + k_B T) \tag{3.6}$$

where $\mathbf{x}$ is the position of $\mathbf{M}$, $V_{\mathbf{AM}^*}(\mathbf{x})$ is the energy with $\mathbf{A}$ chemisorbed to the metal site, $V_{\mathbf{M}^*}(\mathbf{x})$ is the energy of the bare metal site, $\varepsilon_{\mathbf{A}}$ is the gas phase energy of $\mathbf{A}$, and $k_B T$ is the PV contribution to the gas phase enthalpy of $\mathbf{A}$. The same Morse potentials that we used to model grafting[20] are now used to describe the $\mathbf{M} - \mathrm{OSi} \equiv$ bond energies

and $\mathbf{M}(\text{OSi}\equiv)_2$ bond energies. Specifically, the individual interaction energies are

$$\varepsilon_i(r) = D_i\left(1-\exp\left[-a_i(r-r_{i,eq})^2\right]\right) - D_i \tag{3.7}$$

where i is the bond type, D_i is the bond dissociation energy, a_i is inversely related to the vibrational well width, r is the bond length, and $r_{eq,i}$ is its equilibrium bond length. The energy of the bare metal site is

$$V_{\mathbf{M}*}(\mathbf{x}) = \varepsilon_{\mathbf{M}-O}(r_1) + \varepsilon_{\mathbf{M}-O}(r_2) + \varepsilon_{\mathbf{M}\cdots O}(r'_1) + \varepsilon_{\mathbf{M}\cdots O}(r'_2) \tag{3.8}$$

where $\varepsilon_{\mathbf{M}-\text{O}}(r_i)$ is the energy of the $\mathbf{M}-\text{OSi}\equiv$ bonds, r_i is the metal-oxygen bond distance, $\varepsilon_{\mathbf{M}\cdots\text{O}}(r'_i)$ is the energy of the $\mathbf{M}\cdots(\text{OSi}\equiv)_2$ metal-siloxane bond, and r'_i is the metal-siloxane bond distance.

We model adsorption of **A** onto the grafted metal center as an **M-A** bond with energy $\varepsilon_{\mathbf{M}-\mathbf{A}}$. The length of the M-A bond is not explicitly optimized. Instead, we assume that the **M- A** bond displaces the longest and most weakly-coordinated siloxane ($\mathbf{M}\cdots(\text{OSi}\equiv)_2$) from **M**. The displaced siloxane can still exert a repulsive interaction on **M**. We model the close-range repulsion with a Weeks-Chandler-Andersen potential:[36]

$$\varepsilon^{WCA}_{\mathbf{AM}\cdots\mathbf{O}}(r) = D_{\mathbf{AM}\cdots\mathbf{O}}\left(1-\exp\left[-a_{\mathbf{AM}\cdots\mathbf{O}}(r-r_{\mathbf{AM}\cdots\mathbf{O},eq})^2\right]\right) \tag{3.9}$$

for $r \le r_{\mathbf{AM}\cdots\text{O,eq}}$ and $\varepsilon^{WCA}_{\mathbf{AM}\cdots\mathbf{O}}(r) = 0$ otherwise. Thus, the energy of state $\mathbf{AM}^*$ is

$$V_{\mathbf{AM}*}(\mathbf{x}) = \varepsilon_{\mathbf{M}-\mathbf{A}} + \varepsilon_{\mathbf{M}-O}(r_1) + \varepsilon_{\mathbf{M}-O}(r_2) + \varepsilon_{\mathbf{M}\cdots O}(r'_1) + \varepsilon^{WCA}_{\mathbf{M}\cdots O}(r^*) \tag{3.10}$$

where r^* is the longest $\mathbf{M}\cdots\text{O}$ bond prior to the adsorption of **A**.

With these definitions, eq. 3.6 presents a geometry optimization problem much like

that encountered in *ab initio* calculations. The interior atoms must be optimized subject to constraints on peripheral atoms around the metal center. The equilibrium configurations of $\mathbf{M}^*$ and $\mathbf{AM}^*$ are found by changing the M atom position with fixed silanolate and siloxane group positions to minimize 3.8 and 3.10, respectively. This procedure creates a collection of model sites with quenched disorder and limited local flexibility, somewhat like a real amorphous catalyst.

3.3.1 Parameter selection

The quenched disordered lattice was created by starting with a square lattice with spacing 1. Random displacements of the lattice sites were drawn from an isotropic 2D Gaussian distribution with $\sigma^2_{\text{lattice}} = 0.00022$ in the x and y directions. We used the same fractions of silanol, siloxane, and empty sites ($f_{\text{silanol}} = 0.3$, $f_{\text{siloxane}} = 0.3$, and $f_{\text{empty}} = 0.4$) as our previous paper.[20] All rate calculations in this work were performed for a temperature of 300 K and a reactant pressure of 1 atm. The metal-adsorbate bond dissociation energy was modeled as the Cr-C bond dissociation energy for a ($\equiv$ SiO)$_2$Cr(III) alkyl site – the widely accepted active site for Cr/SiO$_2$ olefin polymerization catalysts.[11,37] Based on DFT calculations (Section S1) and reported values for the Cr-C bond,[38] we set $\varepsilon_{\mathbf{M-A}} = 160$ kJ/mol. A list of the parameters and their values are summarized in Table 3.1.

Table 3.1: Parameter values for the quenched disorder lattice, Langmuir-Hinshelwood mechanism, and model chemistry

Parameter	Value
T	300K
$P_{\mathbf{A}}$	1 atm
$\Delta H^{\ddagger}$	65 kJ/mol
$\sigma^2_{\text{lattice}}$	0.00022
f_{Silanol}	0.3
f_{Siloxane}	0.3
f_{Empty}	0.4
$D_{\mathbf{M}\cdots\text{O}}$	120 kJ/mol
$a_{\mathbf{M}\cdots\text{O}}$	1.3
$\hat{r}_{\mathbf{M}\cdots\text{O,eq}}$	1.16
$D_{\mathbf{M-O}}$	500 kJ/mol
$a_{\mathbf{M-O}}$	1.7
$\hat{r}_{\mathbf{M-O,eq}}$	1.0
$\varepsilon_{\mathbf{A}}$	0 kJ/mol
$\varepsilon_{\mathbf{M-A}}$	160 kJ/mol

3.3.2 Site-averaged kinetics

Each metal site has a unique environment, and the different environments lead to a distribution of kinetic properties. For example, the sites will exhibit a distribution of turnover frequencies and activation energies. In contrast, a conventional experiment measures just one site-averaged value for each kinetic property. In this paper, we focus on the site-averaged activation energy. From eqs. 3.2 - 3.5, the activation energy for site i is

$$E_a(\mathbf{x}_i) = -\frac{d\ln \textit{r}_i}{d\beta} = \Delta H(\mathbf{x}_\mathrm{i}) + \Delta H^\ddagger + 2k_BT \tag{3.11}$$

where $\beta = 1/k_BT$. A derivation of eq. 3.11 can be found in section B.2 of appendix B. In this calculation, we assume that E_a, ΔS, $\Delta S^\ddagger$, and $\Delta H^\ddagger$ are not functions of temperature.[39] For ΔH, the temperature dependence from k_BT (eq. 3.6) is considered, but other temperature-dependent terms such as partition functions are ignored. (In practice, all of these properties will probably exhibit some temperature dependence.)

Naively, one might estimate $E_a(\mathbf{x})$ for a large sample of sites and then average them to obtain the site-averaged activation energy. This straightforward average does not give the correct value, even in the limit of large sample sizes. The correct site-averaged activation energy,[40] $\langle E_a\rangle_k$ is obtained from a derivative of the site-averaged rate:

$$\begin{aligned}
\langle E_a\rangle_k &\equiv -\partial \ln \langle \textit{r}\rangle / \partial\beta \\
&= -\frac{\partial}{\partial\beta}\ln\int d\mathbf{x}\rho(\mathbf{x})k(\mathbf{x})\prod\nolimits_i c_i^{\alpha_i} \\
&= -\frac{\int d\mathbf{x}\rho(\mathbf{x})\partial k(\mathbf{x})/\partial\beta}{\int d\mathbf{x}\rho(\mathbf{x})k(\mathbf{x})} \\
&= \frac{\int d\mathbf{x}\rho(\mathbf{x})k(\mathbf{x})\{E_a(\mathbf{x}) + \beta E_a(\mathbf{x})\partial\ln E_a(\mathbf{x})/\partial\beta\}}{\int d\mathbf{x}\rho(\mathbf{x})k(\mathbf{x})} \\
&= \langle E_a\{1+\beta\cdot\partial\ln E_a/\partial\beta\}\rangle_{\rho(\mathbf{x})k(\mathbf{x})}
\end{aligned} \tag{3.12}$$

In this work, we ignore the temperature dependence of E_a. In practice, the rates at individual sites cannot be probed, nor are the temperature intervals in which the rates are measured wide enough to see definitive curvature in the Arrhenius plot. We also expect the correction to be small. Using E_a from eq. 3.11, $E_a\beta\partial \ln E_a/\partial\beta = 2\beta^{-1}$, which will be relatively small compared to a typical $E_a(\mathbf{x})$. Moreover, we anticipate that $E_a\beta\partial \ln E_a/\partial\beta$ term will be similar across different sites, so that conclusions about characteristics of highly active and abundant sites will be unaffected. Using eqs. 3.2 - 3.5, the derivatives and integrations yield:

$$\langle E_a \rangle_k = \left\langle \Delta H(\mathbf{x}) + \Delta H^{\ddagger} + 2k_B T \right\rangle_k. \tag{3.13}$$

The subscript k indicates that the average is computed with probability weights $\rho(\mathbf{x})k(\mathbf{x})$, instead of $\rho(\mathbf{x})$. In practice, this average can be computed in two different ways.

The first strategy is to randomly choose sites from $\rho(\mathbf{x})$ and reweight each of them by $k(\mathbf{x})$ when computing the average:

$$\tilde{E}_a = \sum\nolimits_{i=1}^{n} k(\mathbf{x}_i)E_a(\mathbf{x}_i) \Big/ \sum\nolimits_{i} k(\mathbf{x}_i) \tag{3.14}$$

The numerator and denominator are both exponential averages. As shown in previous work,[35] this strategy usually requires an enormous sample size to converge.

The second strategy is to directly sample sites according to probability weight $\rho(\mathbf{x})k(\mathbf{x})$. This is difficult, because we do not know $k(\mathbf{x})$ precisely prior to performing *ab initio* calculations at $\mathbf{x}$. However, if such a sampling algorithm could be devised (see below), the site-averaged activation energy would become a simple arithmetic average:

$$\bar{E}_a = \frac{1}{n} \sum\nolimits_{i=1}^{n} E_a(\mathbf{x}_i). \tag{3.15}$$

This second strategy enables fast convergence to the site-averaged activation energy according to the central limit theorem. Confidence intervals on the precision of E_a follow from the usual statistical formulae

$$\langle E_a \rangle_k = \bar{E}_a \pm \frac{1}{\sqrt{n}} \hat{S}_{Ea} t_{X,n} \tag{3.16}$$

where $t_{X,n}$ is the student-t statistic for an $X\%$ confidence interval with sample size n, and where the standard error is:

$$\hat{S}^2_{E_a} = \frac{n}{n-1} \left\{ \overline{E^2_{a,i}} - \overline{E_{a,i}}^2 \right\}. \tag{3.17}$$

Of course, these estimates and error formulas do not account for systematic errors in the *ab initio* predictions. Moreover, to sample the distribution $\rho(\mathbf{x})k(\mathbf{x})$, we use a kernel regression model to predict $k(\mathbf{x})$ at sites that have not yet been investigated. The error formulas above also do not account for errors in the kernel regression estimates. In our calculations, the kernel regression errors are much smaller than the intrinsic width of the E_a-distribution, so they can probably be ignored. However, section B.3 of appendix B shows how the typical kernel regression errors could be included in cases where they are large enough to be important.

3.4 Kernel regression

To sample the distribution $\rho(\mathbf{x})k(\mathbf{x})$ starting from a large collection of sites, *i.e.*, from $\rho(\mathbf{x})$, we require preliminary estimates for $k(\mathbf{x})$ at each site. Given accurate calculations as training data at a modest collection of sites, kernel regression can estimate $E_a(\mathbf{x})$ at all the remaining sites.[41,42] The estimated activation energy at site with environment x_i

is a weighted average of the training data, eq. 3.18:

$$\hat{E}_a(\mathbf{x}_i) = \frac{\sum_j E_a(\mathbf{x}_j) w_{ij} \left(d(\mathbf{x}_i, \mathbf{x}_j)\right)}{\sum_j w_{ij} \left(d(\mathbf{x}_i, \mathbf{x}_j)\right)} \tag{3.18}$$

Here $\hat{E}_a(\mathbf{x}_i)$ is the predicted activation energy, and the $E_a(\mathbf{x}_j)$ are computed activation energies. The w_{ij} are Gaussian kernels

$$w_{ij} = \exp[-d^2(\mathbf{x}_i, \mathbf{x}_j)] \tag{3.19}$$

that depend on a Mahalanobis distance,[43] d:

$$d^2(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{S} (\mathbf{x}_i - \mathbf{x}_j). \tag{3.20}$$

Here, $\mathbf{S}$ is a $\dim(\mathbf{x}) \times \dim(\mathbf{x})$ dimensional, positive definite, and symmetric matrix. The kernel regression model is trained by finding the elements of $\mathbf{S}$ which minimizes the leave-one-out loss to best fit the training data. We use Python library tools to implement the kernel regression.[44] Further details about the kernel regression procedures can be found in section 2.3 of chapter 2.

3.5 Importance learning algorithm

The sections above described rate calculations at individual sites, an importance sampling procedure, and a kernel regression (machine learning) procedure. This section integrates all of these components into one "importance learning" algorithm. Importance learning simultaneously accumulates training data, builds the kernel regression model, and focuses computational effort on kinetically important sites with low activation barriers. The algorithm is shown in fig. 3.4.

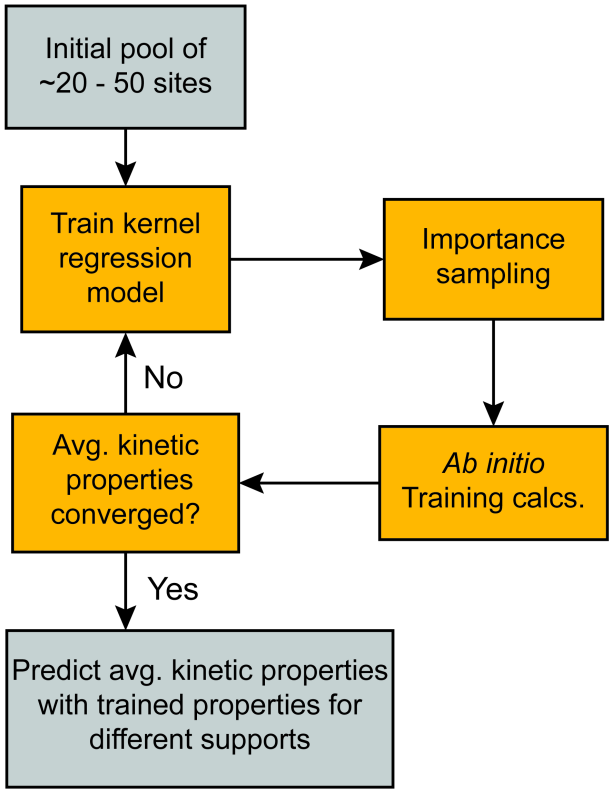


Figure 3.4: The combination of efficient sampling techniques and a machine learning model leads to the "importance learning" algorithm. A set of sites trains a model to learn characteristics of highly active (*i.e.*, important) sites. Efficient sampling techniques select active sites to improve the model and to efficiently predict average kinetic properties. A test for convergence terminates the algorithm when the confidence interval on the site-averaged activation energy shrinks to a prescribed narrow size. In our calculations, the threshold confidence interval was set to 0.75 kJ/mol.

Note that the importance sampling and kernel regression procedures mutually depend on each other. The kernel regression model guides the importance sampling to kinetically important sites. Meanwhile, the accumulated sample of sites and rate calculations teach kernel regression to make accurate preliminary rate predictions.

To compute kinetic properties, precise rate calculations for less active sites are not important, but we need their populations to predict kinetic properties like the overall rate and the fraction of active sites. Therefore, the kernel regression model should also learn to make approximate predictions for inactive sites. For this reason, the importance

learning algorithm begins with rate calculations at a collection of randomly sampled sites. We verified that an initial training set of 20-50 randomly chosen sites is adequate (see section B.4 of appendix B).

3.6 Results

Because the model system is extremely simple, an accurate site-averaged activation energy can be directly calculated without importance learning. Using results for ca. 20,000 sites, we computed the activation energy distribution:

$$\tilde{\rho}(E_a) = \int d\mathbf{x}\rho(\mathbf{x})\delta[E_a(\mathbf{x}) - E_a]. \tag{3.21}$$

and the $k(\mathbf{x})$-weighted activation energy distribution:

$$\tilde{\rho}_k(E_a) = \frac{\int d\mathbf{x}\rho(\mathbf{x})k(\mathbf{x})\delta[E_a(\mathbf{x}) - E_a]}{\int d\mathbf{x}\rho(\mathbf{x})k(\mathbf{x})}. \tag{3.22}$$

Fig. 3.5 shows the essentially exact distributions $\tilde{\rho}(E_a)$ and $\tilde{\rho}_k(E_a)$. The activation energy distribution has support[45] over a range of about 40 kJ/mol. The site-averaged activation energy is 40.4 kJ/mol, about 13 kJ/mol below the (incorrect) average without k-weighting. These results serve as benchmarks for testing the importance learning algorithm. To start the importance learning algorithm, we began with an initial training set of fifty randomly chosen sites. The initial kernel regression model was optimized to minimize the leave-one-out errors. Within this initial training set, the kernel regression model predicts activation energies with a standard error $\sigma \approx 0.8$ kJ/mol. Fig. 3.6 shows how the predicted activation energies compare to the true (precisely computed via eq. 3.11) activation energies for individual sites.

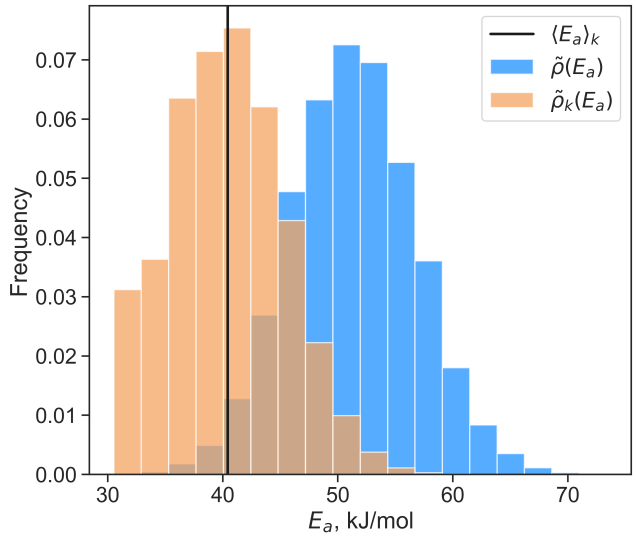


Figure 3.5: Distribution of activation energies (blue) and the rate-weighted activation energy distribution (orange). The solid line shows the site-averaged activation energy.

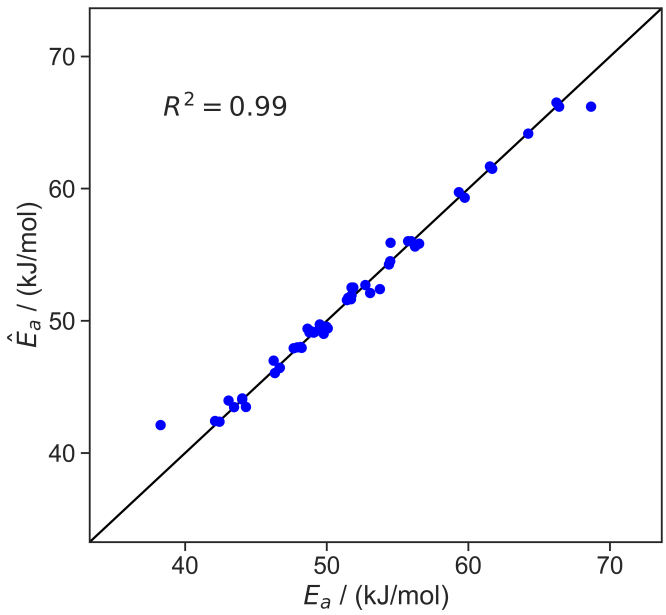


Figure 3.6: Parity plot of predicted activation barriers vs true activation barriers at individual sites. Predictions are from leave-one-out optimization of kernel regression models based on the initial training set of 50 sites. The residuals for all ca. 20,000 sites are approximately Gaussian distributed, with a standard deviation of approximately 0.7 kJ/mol (Figure B.1).

At each iteration of the importance learning algorithm, the activation energy distributions $\tilde{\rho}(E_a)$ and $\tilde{\rho}_k(E_a)$ can be predicted using the kernel regression model. After each iteration, the new calculations are appended to the training set. As more training data is accumulated (primarily at low activation energies), the estimated E_a distributions should become more like the true distribution in the kinetically important range of activation energies. As a corollary, the site-averaged activation energy should also converge to the correct value. Fig. 3.7 shows the predicted distributions at the 0th and 28th iterations of importance learning (the latter being the iteration at which the standard error decreases below 0.75 kJ/mol). A rug shows that the activation energies of the importance sampled sites are indeed centered over the main support of $\tilde{\rho}_k(\hat{E}_a)$.

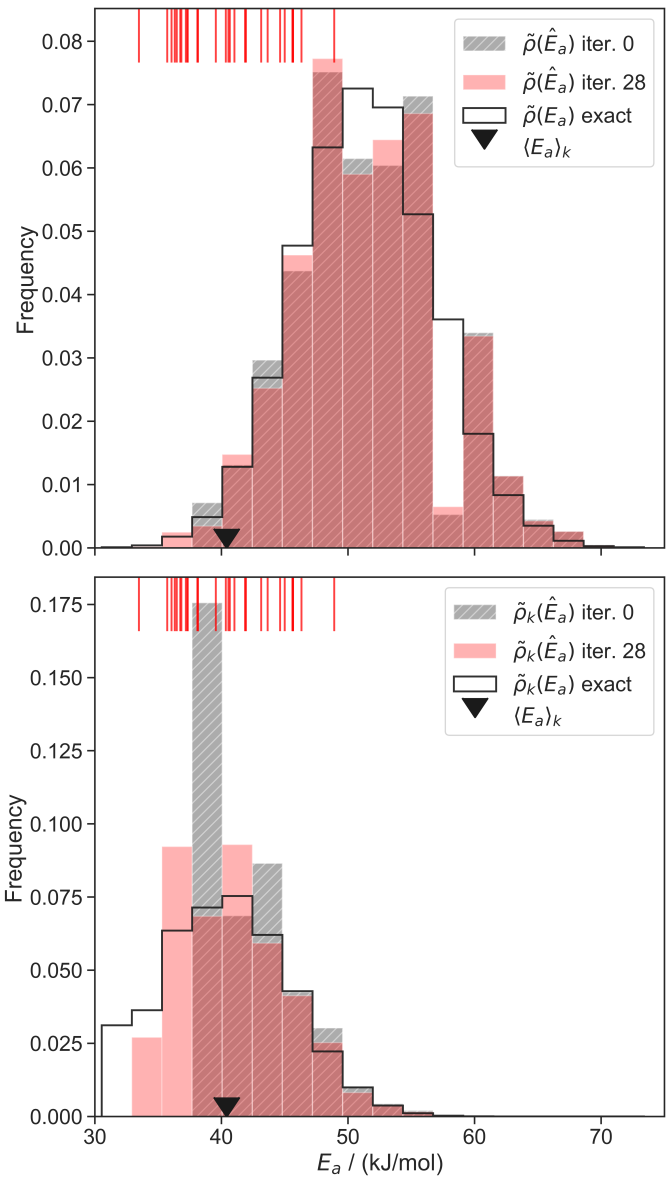


Figure 3.7: Model-predicted activation energy distribution for the unweighted (top) and k-weighted (bottom) distributions at iteration 0 (grey, hatched) and 28 (red) of the importance learning algorithm. Apparent activation energies of importance sampled sites are shown as a rug at the top of each plot. The ▼ symbol shows the correct site averaged E_a.

Prior to importance learning, the initial training set contained only one site with an activation energy under 40 kJ/mol. Importance learning discovers sites with activation energies below 40 kJ/mol, which dominate the overall kinetics. After 28 iterations of im-

portance learning, the low activation energy tail of the predicted $\tilde{\rho}(\hat{E}_a)$ closely resembles that of the exact $\tilde{\rho}(E_a)$. More importantly, the main support of the predicted $\tilde{\rho}_k(\hat{E}_a)$ closely resembles that of the exact $\tilde{\rho}_k(E_a)$. Both distributions are inaccurately predicted at high activation energies, but these sites make vanishingly small contributions to the observed kinetics. They only need to be counted in the normalization of $\tilde{\rho}(E_a)$ to predict the kinetic properties.

Fig. 3.8 shows the convergence of $\langle E_a \rangle_k$ estimates from importance sampling using standard errors. A higher degree of confidence could also be computed using eq. 3.16.

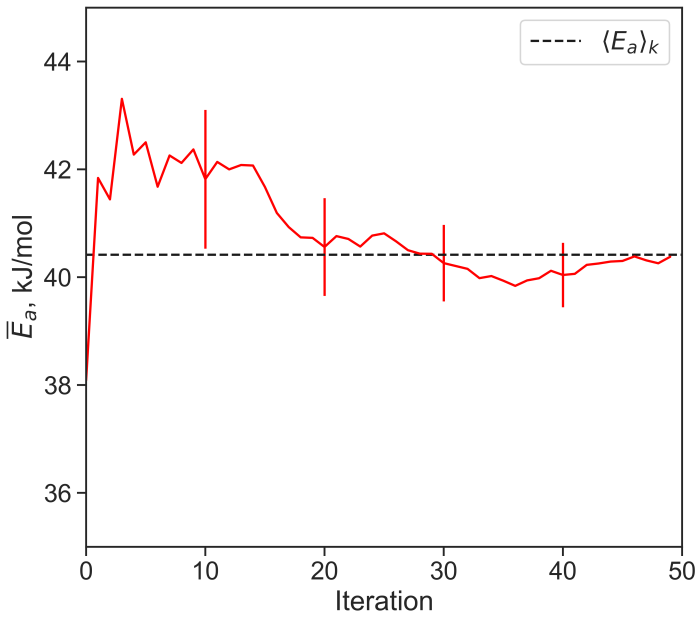


Figure 3.8: The importance learning algorithm converges to within 0.75 kJ/mol of the correct site-averaged $\overline{E}_a$ in 28 iterations. By comparison, a reweighted random sample requires about 200,000 samples to compute $\tilde{E}_a$ with the same level of confidence (Section B.5).

3.7 Identifying characteristics of active sites

In real applications, optimizing the Mahalanobis matrix is inexpensive compared to generating training data from *ab initio* calculations. Therefore, an importance learning calculation can include all potentially important coordinates. However, a central goal of these calculations is to discover those few key characteristics that distinguish active from inactive sites. Intuition would suggest that the most important coordinates can be identified from the largest diagonal elements in the Mahalanobis matrix. The optimized matrix obtained in this work, using d_1, d_2, and θ, is:

$$\mathbf{S} = \begin{array}{c} \begin{matrix} d_1 & d_2 & \theta \end{matrix} \\ \begin{bmatrix} 54370 & -39653 & -11 \\ -39653 & 29308 & 13 \\ -11 & 13 & 0 \end{bmatrix} \end{array} \tag{3.23}$$

Coordinates d_1 and d_2 have the largest diagonal elements, and they indeed have the strongest influence on site activity. The coordinates d_1 and d_2 correspond to silanolate – silanolate distances and siloxane – siloxane distances, respectively. In hindsight, these coordinates should have primary importance because the potential energies are defined in terms of these coordinates.

In general, the diagonal matrix elements are not reliable indicators of the most important structural characteristics. For example, the diagonal matrix elements change magnitude depending on the units used to represent the coordinates. In addition, diagonal matrix elements indicate sensitivity to local structural changes. They do not account for differences in the extent to which sites vary along different structural coordinates within the global ensemble of sites. Off-diagonal matrix elements may also be important. Large off-diagonal matrix elements may indicate that special combinations of the coordinates are important. Alternatively, off-diagonal elements may compensate for

non-orthogonality or redundancy in the set of trial coordinates. The latter complications can be avoided by choosing coordinates that are orthogonal, in the sense:

$$\frac{\partial q_i}{\partial \mathbf{x}} \cdot \frac{\partial q_j}{\partial \mathbf{x}} \approx 0 \tag{3.24}$$

More general guidelines are that

1. Good coordinates should suffice to predict differences in activity over the region with support in distribution $\rho(\mathbf{x})k(\mathbf{x})$.

2. The kernel regression model should predict activation energies with errors that are much smaller than the range of activation energies in $\tilde{\rho}_k(E_a)$.

These two guidelines suggest ranking models according to the fraction of the actual E_a variance that is explained by the model. In linear regression, this is the familiar R^2 statistic. Models that include more input coordinates will generally give larger R^2 values, but small models are preferred, as long as they give accurate site-averaged rate predictions. The fit quality of the kernel regression models trained on different sets of coordinates are shown in Fig 3.9.

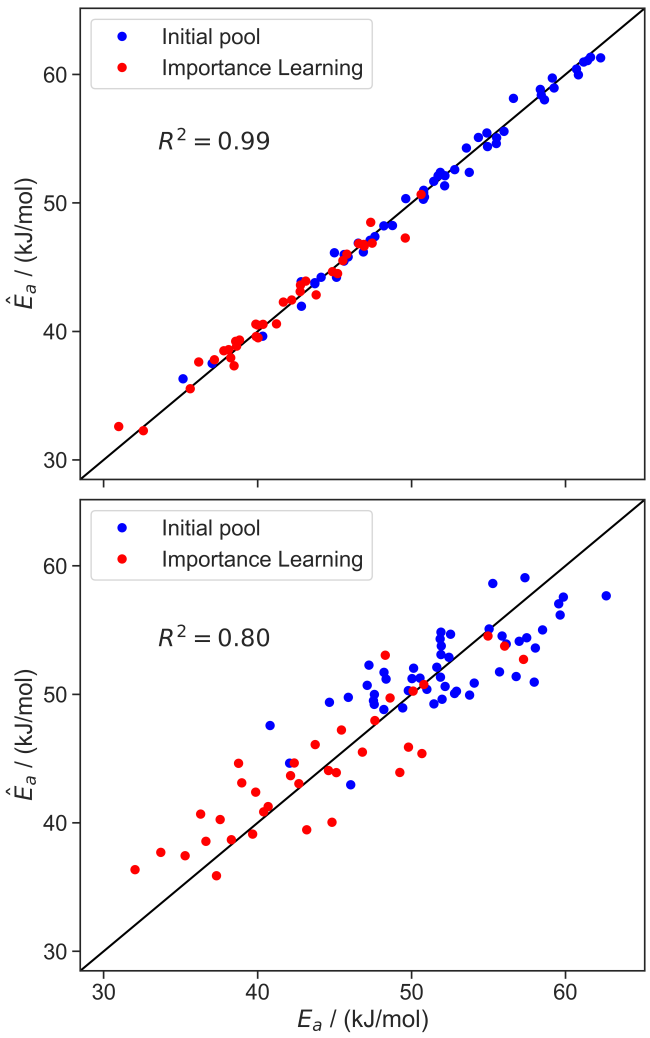


Figure 3.9: Parity plot of model trained with d_1, d_2 (top) and d_1, θ (bottom) at iteration 30 of the importance learning algorithm. As shown in Table 2, d_1 and d_2 are sufficient (without the extra variable θ) to allow kernel regression to predict activation energies across the range of values.

Table 3.2: R^2 values of trained model with different combination of local coordinates at iteration 30 of the importance learning algorithm.

Coordinates	R^2
d_1	0.80
d_2	0.16
θ	-0.03
d_1, d_2	0.99
d_1, θ	0.82
d_2, θ	0.34
d_1, d_2, θ	0.99

The R^2 values identify θ as a kinetically unimportant structural characteristic. The kernel regression model trained only on θ completely fails to make predictions based on the local environment. Models based only on d_1 or d_2 begin to predict coarse trends in the activation energies. The model trained using d_1 and d_2 together makes extremely accurate predictions across the whole range of activation energies. Note that d_1 and d_2 are just two of the five total coordinates that define the local site environment. The model-predicted E_a is plotted as a function of d_1 and d_2 in fig. 3.10. This plot reveals that d_1 and d_2 compensate for each other in active sites. Among sites with the same activation energy, one length increases while the other decreases. Fig. 3.10 also illustrates that the most active sites have shorter d_1 (silanolate-silanolate) distances and longer d_2 (O-O distance of the siloxane ligands) distances relative to the unperturbed distance of 2.00.

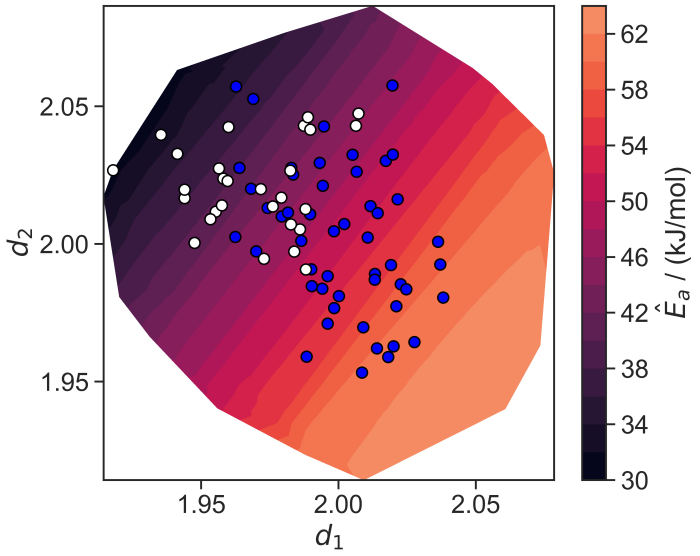


Figure 3.10: Activity of sites as function of the local environment. The upper plot shows the true barriers and the bottom plot shows the model-predicted barriers at iteration 30 of the importance learning loop. Blue points correspond to the initial pool and white points are importance sampled sites.

3.8 Conclusions

Several industrially important or promising catalysts are single metal atoms grafted onto an amorphous support such as silica. These catalysts tend to be poorly understood because the amorphous support gives each site a unique local environment. Moreover, the distribution of disordered environments around each site is quenched, history dependent, and thus largely unpredictable. Each site has a different activation energy, and the variance in activation energies is exponentially magnified in the distribution of activities. Accordingly, active sites tend to be rare, with less than 20% of sites accounting for most of the catalytic activity. The small fraction of active sites hampers both experimental characterization and theoretical modeling efforts.

This paper presented an importance learning algorithm to overcome the theoretical challenges of modeling the activity of such catalysts. It combines machine learning

techniques (kernel regression) and importance sampling techniques (to focus effort on the most active and abundant sites). To illustrate the algorithm, we developed a simple model of a Langmuir-Hinshelwood reaction at sites on a quenched and disordered support. We used the algorithm to compute the site-averaged activation energy.

The algorithm rapidly converged estimates of the site-averaged E_a with uncertainties less than 0.75 kJ/mol, even though the individual sites in the model have activation energies that span a range of nearly 40 kJ/mol. Estimating the site-averaged E_a with the same level of confidence as importance learning requires through standard sampling methods requires 200,000 samples (compared to 75 samples in the importance learning algorithm) for this system. Furthermore, the kernel regression model generated by the algorithm can accurately predict the activation energies using just two structural characteristics of the local environment. The new importance learning algorithm, if combined with *ab initio* calculations and realistic models of amorphous silica, should enable the first rigorous site-averaged computational studies and quantitative predictions for this important family of catalysts.

Bibliography

1. Zhang, X., Shi, H. & Xu, B. Q. Catalysis by gold: Isolated surface Au_3^+ ions are active sites for selective hydrogenation of 1,3-butadiene over Au/ZrO_2 catalysts. *Angew. Chem., Int. Ed.* **44,** 7132–7135. https://www.doi.org/10.1002/anie.200502101 (2005).

2. Hackett, S. F. J. *et al.* High-activity, single-site mesoporous Pd/Al_2O_3 catalysts for selective aerobic oxidation of allylic alcohols. *Angew. Chem., Int. Ed.* **46,** 8593–8596. https://www.doi.org/10.1002/anie.200702534 (2007).

3. Wei, H. *et al.* FeOx -supported platinum single-atom and pseudo-single-atom catalysts for chemoselective hydrogenation of functionalized nitroarenes. *Nat. Commun.* https://www.doi.org/10.1038/ncomms6634 (2014).

4. Thomas, J. M., Raja, R. & Lewis, D. W. Single-site heterogeneous catalysts. *Angew. Chem., Int. Ed.* **44,** 6456–6482. https://www.doi.org/10.1002/anie.200462473 (2005).

5. Wang, A., Li, J. & Zhang, T. Heterogeneous single-atom catalysis. *Nat. Rev. Chem.* **2,** 65–81. https://doi.org/10.1038/s41570-018-0010-1 (2018).

6. Flytzani-Stephanopoulos, M. & Gates, B. C. Atomically Dispersed Supported Metal Catalysts. *Annu. Rev. Chem. Biomol. Eng.* https://doi.org/10.1146/annurev-chembioeng-062011-080939 (2012).

7. McDaniel, M. P. & Welch, M. B. The activation of the phillips polymerization catalyst: I. Influence of the hydroxyl population. *J. Catal.* **82,** 98–109. https://www.doi.org/10.1016/0021-9517(83)90121-5 (1983).

8. Maksasithorn, S., Praserthdam, P., Suriye, K. & Debecker, D. P. Preparation of super-microporous WO_3/SiO_2 olefin metathesis catalysts by the aerosol-assisted sol-gel process. *Microporous Mesoporous Mater.* **213,** 125–133. https://www.doi.org/10.1016/j.micromeso.2015.04.020 (2015).

9. Carrick Wayne, L. *et al.* Ethylene polymerization with supported bis(triphenylsilyl) chromate catalysts. *J. Polym. Sci., Part A-1: Polym. Chem.* **10,** 2609–2620. https://www.doi.org/10.1002/pol.1972.150100909 (1972).

10. Theopold, K. H. Organochromium(III) Chemistry: A Neglected Oxidation State. *Acc. Chem. Res.* **23,** 263–270. https://www.doi.org/10.1021/ar00176a005 (1990).

11. McDaniel, M. P. A Review of the Phillips Supported Chromium Catalyst and Its Commercial Use for Ethylene Polymerization. English. *Adv. Catal.* **53,** 123–606. https://www.doi.org/https://doi.org/10.1016/S0360-0564(10)53003-7 (2010).

12. Ruddy, D. A. & Tilley, T. D. Kinetics and mechanism of olefin epoxidation with aqueous H_2O_2 and a highly selective surface-modified TaSBA15 heterogeneous catalyst. *J. Am. Chem. Soc.* **130,** 11088–11096. https://www.doi.org/10.1021/ja8027313 (2008).

13. Shannon, I. J. *et al.* Metallocene-derived, isolated Mo(VI) active centres on mesoporous silica for the catalytic dehydrogenation of methanol. *J. Chem. Soc., Faraday Trans.* **94,** 1495–1499. https://www.doi.org/10.1039/a800054i (1998).

14. Vining, W. C., Strunk, J. & Bell, A. T. Investigation of the structure and activity of $VO_x/CeO_2/SiO_2$ catalysts for methanol oxidation to formaldehyde. English. *J. Catal.* **281,** 222–230. https://www.doi.org/10.1016/j.jcat.2011.09.024 (2012).

15. Amakawa, K. *et al.* In situ generation of active sites in olefin metathesis. *J. Am. Chem. Soc.* **134,** 11473. https://www.doi.org/10.1021/ja3011989 (2012).

16. McDaniel, M. P. & Martin, S. J. Poisoning Studies on Cr/Silica. 2. Carbon Monoxide. *J. Phys. Chem.* **95,** 3289–3293. https://doi.org/10.1021/j100161a059 (1991).

17. Chauvin, Y. & Commereuc, D. Chemical counting and characterization of the active sites in the rhenium oxide/alumina metathesis catalyst. *J. Chem. Soc., Chem. Commun.* **6,** 462–464. https://doi.org/10.1039/C39920000462 (1992).

18. Howell, J. G., Li, Y. P. & Bell, A. T. Propene Metathesis over Supported Tungsten Oxide Catalysts: A Study of Active Site Formation. *ACS Catal.* **6,** 7728–7738. https://doi.org/10.1021/acscatal.6b01842 (2016).

19. Brunelli, N. A. & Jones, C. W. Tuning acid-base cooperativity to create next generation silica-supported organocatalysts. *J. Catal.* **308,** 60–72. https://doi.org/10.1016/j.jcat.2013.05.022 (2013).

20. Khan, S. A., Vandervelden, C. A., Scott, S. L. & Peters, B. Grafting metal complexes onto amorphous supports: From elementary steps to catalyst site populations: Via kernel regression. *React. Chem. Eng.* **5,** 66–76. https://www.doi.org/10.1039/c9re00357f (2020).

21. Baştuğ, T. & Kuyucak, S. Application of Jarzynski's equality in simple versus complex systems. *Chem. Phys. Lett.* **436,** 383–387. https://www.doi.org/10.1016/j.cplett.2007.01.078 (2007).

22. Bustamante, C., Liphardt, J. & Ritort, F. The Nonequilibrium Thermodynamics of Small Systems. *Phys. Today* **58,** 43–48. https://doi.org/10.1063/1.2012462 (2005).

23. Sear, R. P. Non-self-averaging nucleation rate due to quenched disorder. *J. Phys.: Condens. Matter* **24,** 052205. https://doi.org/10.1088/0953-8984/24/5/052205 (2012).

24. Wang, F. & Landau, D. P. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **86,** 2050. https://doi.org/10.1103/PhysRevLett.86.2050 (2001).

25. Espelid, Ø. & Børve, K. J. Theoretical Models of Ethylene Polymerization over a Mononuclear Chromium(II)/Silica Site. *J. Catal.* **195,** 125–139. https://www.doi.org/10.1006/jcat.2000.2986 (2000).

26. Fong, A., Vandervelden, C., Scott, S. L. & Peters, B. Computational Support for Phillips Catalyst Initiation via Cr-C Bond Homolysis in a Chromacyclopentane Site. *ACS Catal.* **8,** 1728–1733. https://doi.org/10.1021/acscatal.7b03724 (2018).

27. Jystad, A. M., Biancardi, A. & Caricato, M. Simulations of Ammonia Adsorption for the Characterization of Acid Sites in Metal-Doped Amorphous Silicates. *J. Phys. Chem. C* **121,** 22258–22267. https://www.doi.org/10.1021/acs.jpcc.7b08113 (2017).

28. Floryan, L., Borosy, A. P., Nunez-Zarur, F., Comas-Vives, A. & Coperet, C. Strain Effect and Dual Initiation Pathway in Cr(III)/SiO_2 Polymerization Catalysts from Amorphous Periodic Models. English. *J. Catal.* **346,** 50–56. https://www.doi.org/10.1016/j.jcat.2016.11.037 (2017).

29. Handzlik, J. Properties and metathesis activity of monomeric and dimeric Mo centres variously located on γ-alumina - A DFT study. *Surf. Sci.* **601,** 2054–2065. https://www.doi.org/10.1016/J.SUSC.2007.03.002 (2007).

30. Guesmi, H. & Tielens, F. Chromium oxide species supported on silica: A representative periodic DFT model. *J Phys Chem C* **116,** 994–1001. https://doi.org/10.1021/jp209680r (2012).

31. Fong, A., Yuan, Y., Ivry, S. L., Scott, S. L. & Peters, B. Computational Kinetic Discrimination of Ethylene Polymerization Mechanisms for the Phillips (Cr/SiO_2) Catalyst. English. *ACS Catal.* **5,** 3360–3374. https://www.doi.org/10.1021/acscatal.5b00016 (2015).

32. Delley, M. F. *et al.* Proton Transfers Are Key Elementary Steps in Ethylene Polymerization on Isolated Chromium(III) Silicates. *Proc. Natl. Acad. Sci.* **111,** 11624–11629. https://www.doi.org/10.1073/pnas.1405314111 (2014).

33. Gierada, M. & Handzlik, J. Active Sites Formation and Their Transformations During Ethylene Polymerization by the Phillips CrO_x/SiO_2 Catalyst. *J. Catal.* **352,** 314–328. https://www.doi.org/10.1016/j.jcat.2017.05.025 (2017).

34. Ewing, C. S., Bhavsar, S., Veser, G., McCarthy, J. J. & Johnson, J. K. Accurate amorphous silica surface models from first-principles thermodynamics of surface dehydroxylation. *Langmuir* **30,** 5133–5141. https://doi.org/10.1021/la500422p (2014).

35. Goldsmith, B. R., Peters, B., Johnson, J. K., Gates, B. C. & Scott, S. L. Beyond Ordered Materials: Understanding Catalytic Sites on Amorphous Solids. *ACS Catal.* **7,** 7543–7757. https://www.doi.org/10.1021/acscatal.7b01767 (2017).

36. Weeks, J. D., Chandler, D. & Andersen, H. C. Role of Repulsive Forces in Determining Equilibrium Structure of Simple Liquids. English. *J. Chem. Phys.* **54,** 5237–5247. https://doi.org/10.1063/1.1674820 (1971).

37. McGuinness, D. S., Davies, N. W., Horne, J. & Ivanov, I. Unraveling the Mechanism of Polymerization with the Phillips Catalyst. *Organometallics* **29,** 6111–6116. https://www.doi.org/10.1021/om100883n (2010).

38. Luo, Y.-R. *Comprehensive handbook of chemical bond energies* English (CRC Press, Boca Raton, 2007).

39. Peters, B. & Scott, S. L. Single Atom Catalysts on Amorphous Supports: A Quenched Disorder Perspective. *J. Chem. Phys.* **142,** 104708. https://www.doi.org/10.1063/1.4914145 (2015).

40. Wu, C., Schmidt, D., Wolverton, C. & Schneider, W. Accurate coverage-dependence incorporated into first-principles kinetic models: Catalytic NO oxidation on Pt (111). *Journal of Catalysis* **286,** 88–94. ISSN: 0021-9517. https://doi.org/10.1016/j.jcat.2011.10.020 (2012).

41. Weinberger, K. Q. & Tesauro, G. Metric Learning for Kernel Regression. *J. Mach. Learn. Res.,* 8. http://proceedings.mlr.press/v2/weinberger07a.html (2007).

42. Hofmann, T., Schölkopf, B. & Smola, A. J. Kernel methods in machine learning. *Ann. Stat.* **36,** 1171–1220. https://www.doi.org/10.1214/009053607000000677 (2008).

43. Mahalanobis, P. C. On the generalized distance in statistics. *Proc. Natl. Inst. Sci. India* **2,** 49–55. https://doi.org/10.1007/s13171-019-00164-5 (1936).

44. Carey, C. J. & Tang, Y. *metric-learn* https://github.com/metric-learn/metric-learn, 2019. https://github.com/metric-learn/metric-learn.

45. Logan, J. D. *Applied mathematics* 3rd. **14204421,** 529 (Wiley-Interscience, Hoboken, N.J., 2006).

Chapter 4

Grafting $TiCl_4$ onto amorphous silica: modeling effects of silanol heterogeneity

4.1 Introduction

Amorphous silicas are preferred supports for many industrially relevant atomically-dispersed catalysts.[1] For example, Cr/SiO_2 catalysts are used in the polymerization of ethylene,[2] Ti/SiO_2 is active for the epoxidation of alkenes,[3] and Mo/SiO_2 and W/SiO_2 catalysts are active in olefin metathesis.[4] Silica supports offer high mechanical and thermal stability and the absence of strong Lewis or Brønsted acidity which can lead to undesired side-reactions.[1] Furthermore, compared to crystalline silicas, amorphous silicas provide a larger surface area.[1] These properties also make amorphous silica important in fields beyond catalysis including microelectronics,[5] communication,[6] and the pharmaceutical industry.[7] Amorphous silica surfaces are terminated by siloxane bridges (≡SiOSi≡) and terminal silanols (≡SiOH). The silanols can be classified by their proximity to other silanols: isolated or hydrogen-bonded. They can also be classified according to their connectivity to Si atoms: isolated (one OH on a Si atom with no OH groups on adja-

cent Si atoms), geminal pairs (two OH groups on the same Si atom), and vicinal pairs (two OH groups on adjacent Si atoms, directly linked by one siloxane bridge).[1] Vicinal silanol pairs may or may not be H-bonded,[1] and each member of the vicinal pair may also be part of a geminal pair. A distance-based classification has also been proposed. For example, silanols separated by 4-6 Å have been referred to as nearly free silanols and silanols separated by greater than 6 Å have been referred to as fully free silanols.[8] Amorphous silicas exhibit short range order and limited medium range order,[9,10] but the lack of long range order prevents a precise structural characterization.[1] Nevertheless, some structural insight can be obtained from spectroscopic analyses. For example, the ≡(SiO-H) bands in IR spectra can distinguish non-H-bonded from H-bonded silanols,[1] and 29Si chemical shifts in MAS-NMR spectra differ for Si atoms bearing one or two OH groups.[11] Vicinal silanols are particularly difficult to characterize, but their abundance has been inferred from the correlated grafting locations of molecular precursors. For example, the reaction stoichiometry for the grafting of $GaMe_3$ onto amorphous silica coupled with EXAFS analysis of the resulting Ga dimers suggest that most silanols on fumed and precipitated silicas occur in vicinal pairs that persist even when the silica is highly dehydroxylated, due to the very high strain that their removal requires.[12,13] Beyond these clues about the relative abundance of isolated, geminal, and vicinal sites, we have little information about local structural differences. For example, the Si-OH bonds present in a vicinal pair may be oriented parallel to each other or adopt a large dihedral angle. Such differences are likely to influence adsorption and reactivity at vicinal sites. Metal complexes have also been grafted onto amorphous silica to synthesize atomically dispersed catalysts.[4,14–16] The $TiCl_x$ sites generated via the grafting of $TiCl_4$ on fumed silica are examined in a companion publication.[17] In that study, a non-porous, fumed silica (Aerosil-380) with a surface area of 340 m^2/g was fully dehydrated and partially dehydroxylated in vacuo at either 100 or 500 °C. Henceforth, we will refer to these sil-

icas as A380-100 and A380-500. Their silanol densities are 1.4 and 2.0 silanols/nm^2, respectively. When excess $TiCl_4$ vapor in a storage vessel, in equilibrium with its liquid, was transferred to a 300 mL reaction flask containing A380-100 or A380-500, (ca. 15 mg), all silanols were consumed (according to IR). The nature of the grafted species was inferred to be monopodally-grafted [$\equiv SiOTiCl_3$], by IR, elemental analysis, and EXAFS. In principle, grafting on isolated and vicinal silanols may proceed via eqs. 4.1 to 4.3.

$$\equiv SiOH + TiCl_4 \rightleftarrows \equiv SiOTiCl_3 + HCl. \tag{4.1}$$

$$\equiv SiOH + \equiv SiOTiCl_3 \rightleftarrows (\equiv SiO)_2TiCl_2 + HCl. \tag{4.2}$$

$$2 \equiv SiOTiCl_3 \rightleftarrows (\equiv SiO)_2TiCl_2 + TiCl_4. \tag{4.3}$$

However, experimental measurements cannot shed light on the contributions of eqs. 4.2 and 4.3, since bipodally grafted $(\equiv SiO)_2TiCl_2$ was not observed. Computational studies on the mechanism and kinetics of grafting can, in principle, provide information beyond that available from experiments, but these studies have primarily used single-site models[18–21] and model compounds like silsesquioxanes.[22] Of course, single-site models cannot account for the structural diversity of site environments on silica. The few studies that go beyond single-site models only examined a small sample of sites and focused on thermodynamic stability of different grafted species.[23–25] None of the previous studies examine grafting kinetics at an ensemble of disordered sites and at realistic experimental conditions. Recent studies have developed large scale atomistic models for amorphous silica that match key properties like silanol density and IR spectra with experimentally synthesized silicas.[25–28] Others developed ab initio methods for modeling

single-atom catalysts supported on amorphous silica.[29–31] We developed a framework to model the grafting of metal complexes onto amorphous supports using machine learning and population balance modeling.[32–34] In this work, we use ab initio calculations and the population balance modeling framework to understand the mechanism of $TiCl_4$ grafting onto amorphous silica. The intermediates and transition states are computed using DFT for $TiCl_4$ interacting with vicinal silanol pairs possessing a range of dihedral angles. Free energy profiles are used to construct continuously varying rate constants and equilibrium constants as functions of the dihedral angle. Then a population balance model for the intermediates and grafted species is constructed as a function of dihedral angle and time, and solved for an ensemble of silica sites. Finally, the population of sites predicted by the population balance model is compared to the results of $TiCl_4$ grafting experiments.

4.2 Model chemistry

Computations were performed using DFT with the ωB97X-D functional.[35] The def2-TZVP basis set[36] was used for Ti, and the TZVP basis set[37,38] was used for all other atoms, unless otherwise specified. Energy minima and transition states were obtained using the Berny algorithm, as implemented in Gaussian 09.[39] The RMS and maximum forces were required to be less than 3.0×10^{-4} and 4.5×10^{-4} Hartrees/Bohr, respectively, unless otherwise specified. The RMS and maximum displacements were required to be less than 1.2×10^{-3} and 1.8×10^{-3} Bohr, respectively. All minima are vibrationally stable, while transition states have a single imaginary frequency. Free energies of gas phase species include rotations, translations, and vibrations (within the harmonic approximation), while free energies of surface species include only vibrations. Cluster models with fixed peripheral atoms were used to model the grafted Ti sites. Vibrational modes for the cluster models do not include any motions of the fixed peripheral atoms.

Free energies were computed at standard temperature (298.15 K) and 760 Torr partial pressures for both HCl and $TiCl_4$. The silicas used in the companion experimental study (see Introduction) described in the introduction have surface silanol densities of 1.4 and 2.0 silanols/nm_2. If all of these silanols were converted to monopodal [$\equiv SiOTiCl_3$] sites, the HCl pressure for a 15 mg silica sample in a 300 mL reactor would be 0.75 and 1.07 Torr, respectively. At 298.15 K, the vapor pressure of $TiCl_2$ is 10 Torr.[40] To obtain free energies at conditions close to those in the experiments, values were computed at 10 Torr $TiCl_4$ and 1 Torr HCl.

4.3 Generation of silanol clusters

We modeled the grafting of $TiCl_4$ to a cluster model for a vicinal silanol pair. Our preliminary investigations revealed that the grafting free energy depends on the relative orientation of the vicinal silanols. Therefore, we developed a procedure to generate vicinal silanol clusters with a range of O-Si-Si-O dihedral angles. First, each vicinal silanol cluster was optimized with capping fluorine atoms (instead of hydroxyl groups) at the periphery, fig. 4.1a. The fluorines allow us to optimize interior Si, O, and H atom positions without interference from spurious H-bonds among the peripheral atoms. A series of 47 vicinal silanol clusters was generated with fixed dihedral angles ranging from 0 to 92° in increments of 2°. In these energy minimizations, the 6-31G basis[37,38] (smaller than TZVP) was used for the F atoms, to better mimic the electronegativity of a hydroxyl group.[41]

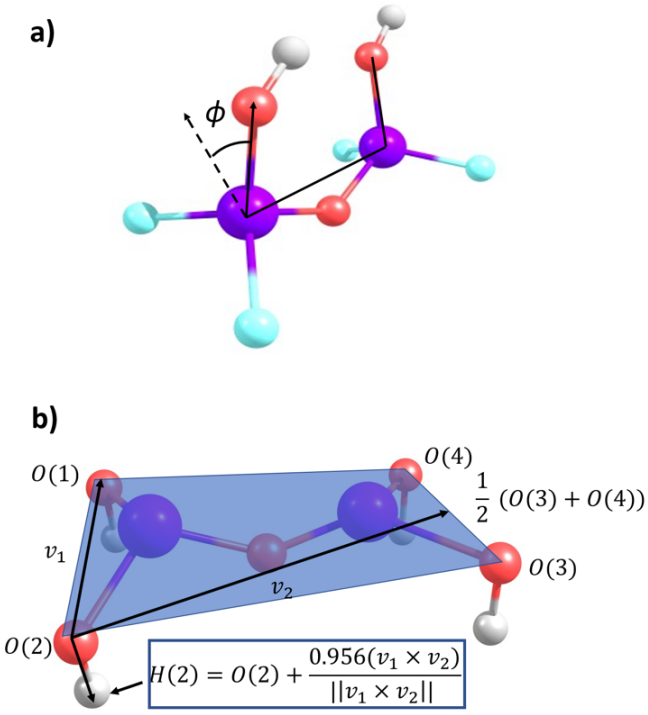


Figure 4.1: Optimized structure of the F-terminated vicinal silanol cluster. Constraining the marked dihedral angle at a series of values from 0 to 92° in increments of 2°, while optimizing all other degrees of freedom, generated 47 distinct clusters. Color scheme: O (red), H (white), Si (purple), F (blue).

Next, the F atoms of each optimized cluster were replaced by peripheral O atoms. The relevant Si-O bond distances were set at 1.648 Å (corresponding to the Si-O bond distance in an optimized $Si(OH)_4$ cluster, described in the SI). An H atom was placed 0.956 Å from each peripheral O atom (corresponding to the O-H bond distance in the optimized $Si(OH)_4$ cluster). In addition, the peripheral O-H bonds were directed along the cross-product of vectors ν_1 and ν_2, as illustrated in fig. 4.1b. This orientation ensures that the peripheral O-H bonds point away from the "active" vicinal hydroxyls, again to avoid spurious H-bonding interactions during analysis of grafting reactions. Finally, each vicinal silanol cluster was reoptimized while keeping the positions of the peripheral hydroxyl groups fixed. The peripheral hydroxyl groups were also fixed in all subsequent

computations.

Fig. 4.2 shows representative vicinal silanol clusters with O-Si-Si-O dihedral angles ϕ of 0° and 66°. In vicinal silanol pairs with O-Si-Si-O dihedral angles less than 60°, one silanol forms a hydrogen bond to the other. This hydrogen bond is potentially important in grafting reactions, because it may weaken/activate the two O-H bonds. For example, the O-H bond for the H-bond donor is slightly elongated (from 0.956 with no H-bond to 0.960 Å with the H-bond). Vicinal silanols with dihedral angles larger than 60° do not engage in mutual hydrogen bonding.

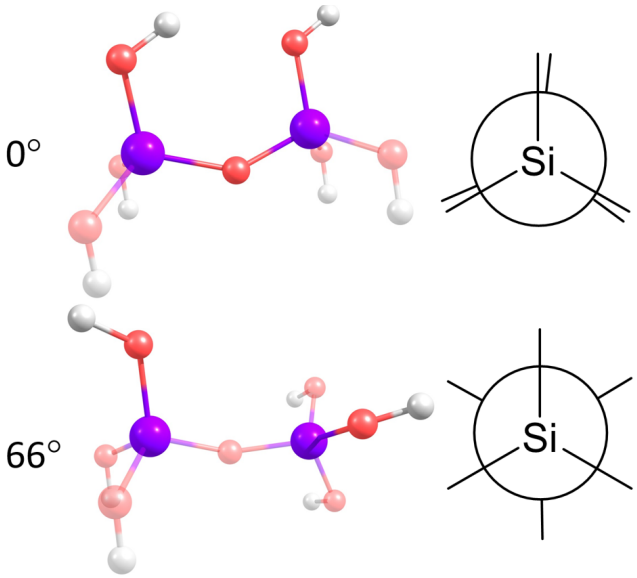


Figure 4.2: Optimized vicinal silanol clusters with different O-Si-Si-O dihedral angles: 0° (top) and 66° (bottom). Newman projections are shown on the right. Color scheme: O (red), H (white), Si (purple). Fixed peripheral atoms are indicated as transparent.

4.4 Generation of amorphous silica models

Slab models for amorphous silica were generated using classical molecular dynamics simulations within the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) software.[42] The silica interactions were described by the van Beest, Kramer,

van Santen (BKS) empirical force field,[43] which consists of pairwise Buckingham and electrostatic potentials. Each slab was generated by a melt-cleave-quench-functionalize procedure that is described in detail elsewhere.[44] Briefly, a β-cristobalite crystal consisting of 192 Si and 384 O atoms was heated to 8000 K at 1 atm in the NPT ensemble. The resulting molten silica was cleaved and the simulation box size was increased along one dimension to form a liquid slab sitting within a vacuum and equilibrated in the NPzAT ensemble. The molten slab was quenched to form the amorphous solid by cooling at a rate too rapid to allow for crystallization; the models in the present work were formed using a cooling rate of 1 K/ps. The two surfaces (top and bottom) of the slab were functionalized by aprocedure designed to produce only vicinal silanol pairs as described below. A 1 fs timestep was used for all MD simulations. Constant temperature was maintained with a Berendsen thermostat using a 1 ps damping constant.46 The NPT and NPzAT simulations used a Berendsen barostat with a 1 ps time constant and a modulus of 360,000 atm.[45] The final simulation cell has dimensions of 21.0155×21.0155 along its cross-section and 90.83 Å in the direction nominally normal to the slab surfaces; the precise dimension varies between slabs due to the NPzAT quenching step. A three-dimensional Ewald summation47 with a tolerance of 1.0×10^{-4} was used to treat the long-range electrostatics. Silanols were formed by the addition of a water molecule across an Si-O-Si bond, creating two Si-OH groups. A vicinal pair results when the two Si atoms involved are part of a two-membered (2M) Si_2O_2 ring (here "two" indicates the number of Si atoms in the ring). Thus, an ensemble of vicinal silanols was created by functionalizing every 2M ring at the surface in 100 amorphous silica slabs, each generated with the procedure described above. For each slab, the atoms at the surface and those involved in 2M rings were identified. For each 2M ring with both Si atoms at the surface, the O atom closer to the vacuum phase was chosen for functionalization. It was replaced by two OH groups, one attached to each Si atom; the net reaction is thus the

addition of a water molecule, ensuring charge neutrality is maintained. The location of the O atom in each OH group was determined using a Monte Carlo procedure to find an orientation that does not overlap with other atoms in the slab. Then, the corresponding H atom in the OH moiety was placed along the nominal surface normal. The resulting OH-functionalized surface was equilibrated by a 1 ns NVT trajectory; the silanol interactions are based on a modified BKS force field that maintains charge neutrality and adds Si-O and O-H harmonic stretching potentials as well as harmonic Si-O-H bending potentials.[44] An example of this functionalization is shown in fig. 4.3.

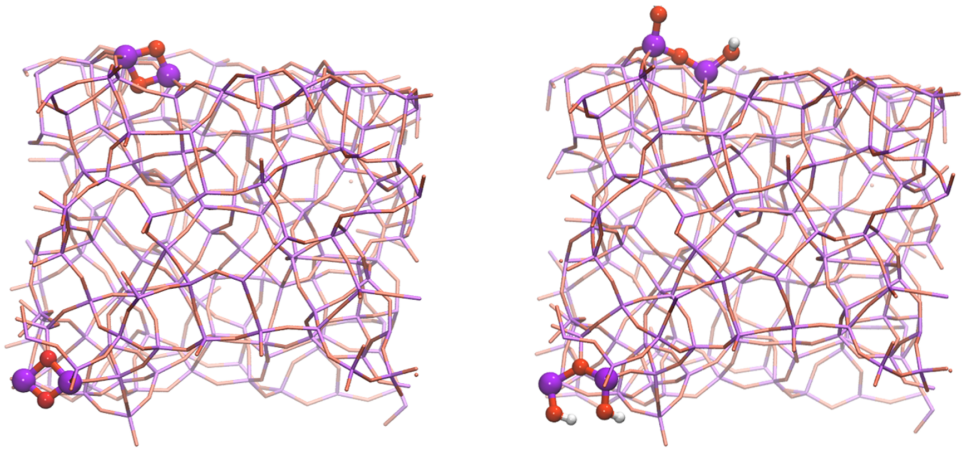


Figure 4.3: Amorphous silica surface model (wireframe) with unfunctionalized 2M rings at the surfaces (left; ball-and-stick) and the vicinal silanol pairs (right; ball-and–stick) resulting from their hydrolysis. Color scheme: O (red), H (white), Si (purple).

For the 100 different amorphous silica slabs created in this work, 388 vicinal silanol pairs were generated in total and used to describe the dihedral angle distribution. The number of vicinal pairs per slab (including both the top and bottom surfaces) ranges from one to seven, with an average of ∼4 pairs. The average surface silanol density over all the slabs generated is 0.88 OH/nm^2. Note that our approach generates only vicinal silanols in one specific way, and that the silanol densities are lower than those obtained using other methods to generate different silanol types.[44] Further studies are needed to

determine which procedures and models most accurately replicate the properties and structure of real silicas.

4.5 Results

4.5.1 Computational assessment of $TiCl_4$ grafting onto vicinal silanols

Given the previously established prevalence of paired silanols on partially dehydroxylated amorphous silica surfaces,[12,13] we investigated the grafting of $TiCl_4$ onto vicinal disilanol sites. First, we investigated $TiCl_4$ grafting at a vicinal site with aligned silanols, i.e., an O-Si-Si-O dihedral angle ϕ of 0° and a hydrogen-bond between them (fig. 4.2). This site is one of a training set of vicinal disilanol clusters, with varying dihedral angles. We use it to obtain initial, detailed information about $TiCl_4$ grafting kinetics and thermodynamics at vicinal sites. Intermediates and transition states are shown in fig. 4.4, with free energies computed for both standard and low-pressure conditions. Intermediates involving physisorbed $TiCl_4$ are not shown, but all have higher free energies than the corresponding precursor states with no bound $TiCl_4$ molecule.

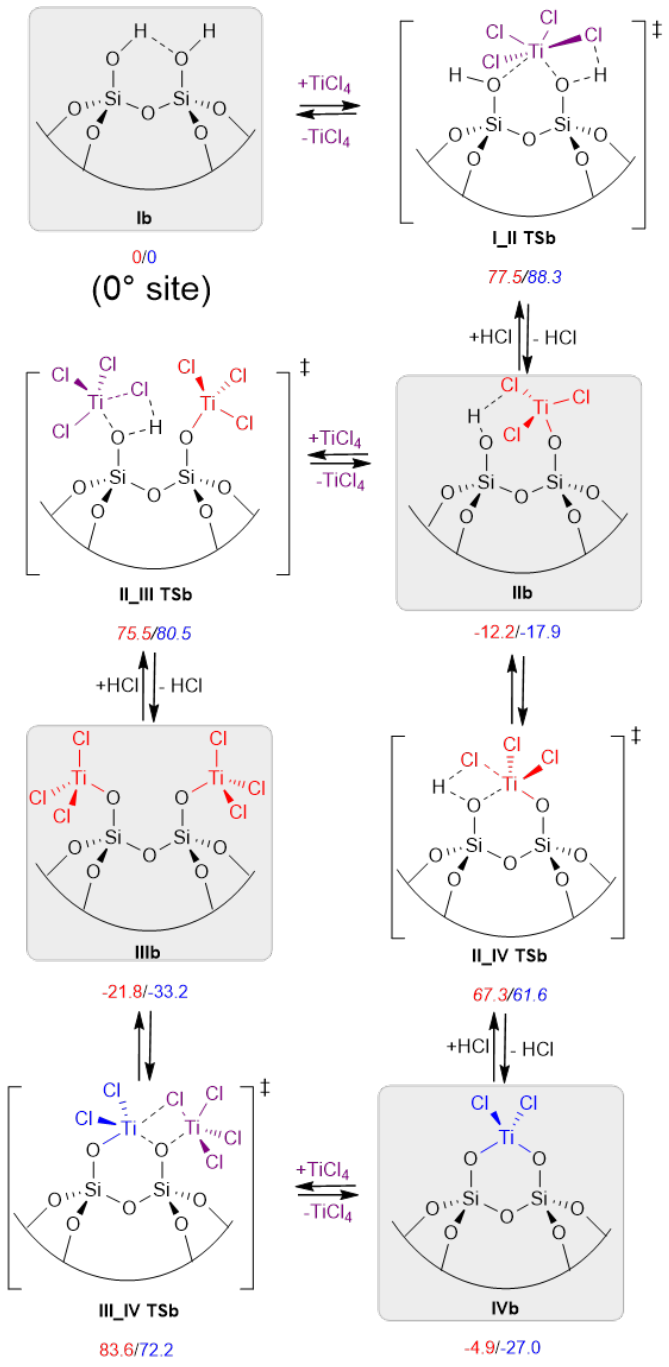


Figure 4.4: Grafting of $TiCl_4$ onto a vicinal silanol pair with parallel silanols (dihedral angle 0°). Free energies (kJ/mol) are shown below each structure for 298.15 K and standard pressure (760 Torr each $TiCl_4$ and HCl, left, in red), and for low pressure (10 Torr $TiCl_4$ and 1 Torr HCl, right, in blue). The gray boxes indicate stable or metastable surface intermediates. Other species are transition states.

In the first step, $TiCl_4$ and one of the two silanols react to eliminate HCl and form a monopodal Ti grafted site [≡SiOTiCl_3] (**IIb**) adjacent to a silanol site. When the reacting silanol is the H-bond donor in the vicinal pair, the barrier is 92.8 kJ/mol (not shown). The barrier for reaction with the hydrogen-bond acceptor silanol is slightly lower, 88.3 kJ/mol. In both transition states, the hydrogen bond between the silanols is disrupted. In the final structure **IIb**, a new hydrogen-bond is formed between the remaining silanol and a chloride ligand. The ratio of **Ib** and **IIb** is obtained from the equilibrium expression expressed in eq. 4.4:

$$K_2 = exp[-\beta \Delta G_2^0] = [\theta_{\mathbf{IIb}}(P_{HCl}/P^0)]/[\theta_{\mathbf{Ib}}(P_{sat}/P^0)] \tag{4.4}$$

where ΔG_2 is the free energy for the reaction **Ib** → **IIb**, calculated at the reference pressures of HCl and $TiCl_4$ (P_0 = 760 Torr), P_{sat} = 10 Torr and P_{HCl} = 1 Torr (low pressure conditions as defined in section 4.2). At equilibrium, $\theta_{\mathbf{IIb}}/\theta_{\mathbf{Ib}} \approx 1.4 \times 10^3$, i.e., **IIb** is strongly favored. Since the barrier for the reverse reaction (**IIb** → **Ib**) is 106.2 kJ/mol, the first step is essentially irreversible at room temperature under the low pressure conditions.

The remaining silanol can react with a second $TiCl_4$ molecule, via elimination of HCl, to form a vicinal pair of monopodally-grafted Ti species, 2[≡SiOTiCl_3] (**IIIb**). We searched for a potential **IIIb** structure with a Ti-Ti dimer. Multiple optimizations were performed with different starting points, where the Ti atoms were placed at a close distance. However, all initial positions optimized to the structure with the non-interacting Ti atoms, as shown in fig. 4.4. Alternatively, **IIb** can undergo intramolecular elimination of HCl, converting to the "bipodal" site [$(\equiv SiO)_2TiCl_2$] (**IVb**). Bipodal **IVb** can also revert to **IIIb** via reaction with free $TiCl_4$. Even at low $TiCl_4$ pressure (relevant to the experimental grafting conditions), the order of stability is **IIIb** > **IVb**. However, the

barrier for intramolecular conversion of **IIb** to **IVb** is much smaller (79.5 kJ/mol) than for the intermolecular reaction of **IIb** with $TiCl_4$ to give **IIIb** (98.4 kJ/mol). Therefore, most **IIb** is initially converted to the kinetic product, IVb. Eventually, most **IVb** reacts with $TiCl_4$ to give the thermodynamic product, **IIIb**, with a barrier of 99.2 kJ/mol. Since **Ib** and **IIb** are much less stable than **IIIb** and **IVb**, their equilibrium surface coverages are expected to be negligible under reaction conditions. Accordingly, we can model the outcome of grafting as an equilibrated mixture of **IIIb** and **IVb** using eqs. 4.5 and 4.6:

$$\theta_{\mathbf{eqIIIb}} = \frac{P_{sat}/P^0}{P_{sat}/P^0 + K_3} \tag{4.5}$$

$$\theta_{\mathbf{eqIVb}} = \frac{K_3}{P_{sat}/P^0 + K_3} \tag{4.6}$$

Here, $K_3 = exp[-\beta \Delta G_3]$ is the equilibrium constant for the reaction **IIIb** $\rightarrow$ **IVb** + $TiCl_4$, and ΔG_3 is the free energy of reaction calculated at reference pressure of $TiCl_4$ (P_0=760 Torr). The predicted surface coverages are 0.92 and 0.08 for **IIIb** and **IVb**, respectively. Based on a recent study benchmarking DFT methods using the ωB97X-D functional against CCSD(T) for several types of reactions (oxidation, hydration, metathesis, and epoxidation) catalyzed by Mo and W,[46] the estimated uncertainty in free energy differences is ca. $\pm$8 kJ/mol. Note that additional errors may result from our use of rigid peripheral atom constraints to represent connections to a silica matrix that is nevertheless slightly flexible.[47] Since the uncertainty is larger than the free energy difference (6 kJ/mol), we cannot definitively predict which of the two sites will be more abundant for parallel vicinal silanol pairs.

4.5.2 Effect of dihedral angle on the population of grafted species

The previous section showed that a vicinal site with a dihedral angle of 0° should react with $TiCl_4$ to give **IIIb** and **IVb** as the predominant species upon completion of grafting. However, amorphous silica has a distribution of vicinal site geometries. Silanol pairs with non-zero dihedral angles may react differently. Therefore, we analyzed the effect of dihedral angle in vicinal silanol sites on the stability of the two grafted sites. Based on the findings in section 4.5.1, we assume that all silanols react even under low pressure conditions, and that only **IIIb** and **IVb** are eventually present. The free energy difference between **IIIb** and **IVb** was computed for sites with dihedral angles between 0 and 92°, fig. 6. To obtain a smooth curve, we were obliged to impose much tighter convergence criteria for these calculations. The RMS and maximum forces were required to be less than 1.00×10^{-5} and 1.50×10^{-5} Hartrees/Bohr, respectively, while the RMS and maximum displacements were required to be less than 4.00×10^{-5} and 6.0×10^{-5} Bohr, respectively. Fig. 6 is approximately, but not perfectly, monotonic.

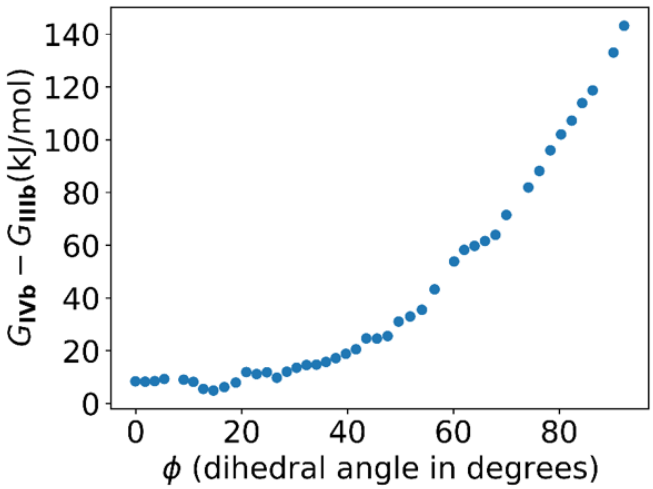


Figure 4.5: Free energy differences between the bipodal site (**IVb**) and a vicinal pair of monopodal sites (**IIIb**), as a function of the initial silanol dihedral angle in state **Ib**. Free energies are computed at 298.15 K and 10 Torr $TiCl_4$.

In general, the free energy difference increases with the dihedral angle ϕ, consistent with an increase in ring strain in **IVb** which increasingly de-stabilizes this site relative to **IIIb**. However, the free energy difference is small (5-10 kJ/mol) and nearly constant for dihedral angles below ca. 20°. Thereafter, it increases rapidly. This is because, on increasing the dihedral angle, the titanasiloxane ring slowly distorts up to 40° and starts to deform rapidly after that. When the dihedral angle reaches 60°, the hydrogen bond in the vicinal silanol **Ib** disappears. At this point, **IIIb** is favored by 42 kJ/mol. The difference continues to increase up to 143 kJ/mol for the maximum dihedral angle studied here (92°). The free energy difference was used in conjunction with eqs. 4.5 and 4.6 to calculate the surface coverage as a function of dihedral angle, fig. 4.6.

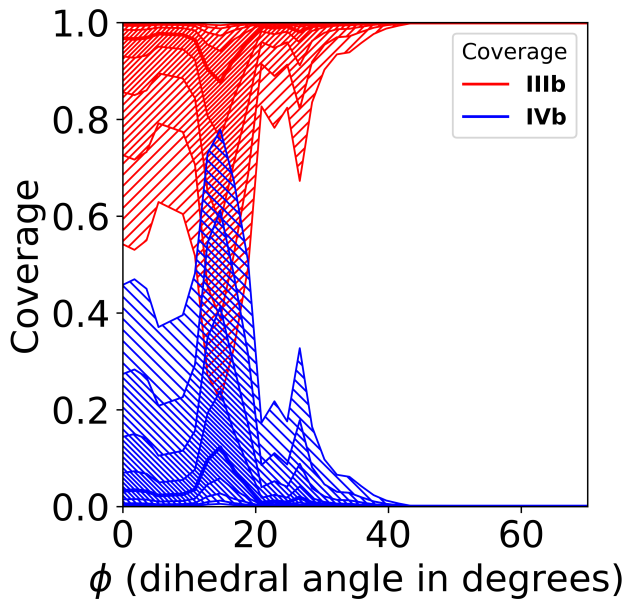


Figure 4.6: Surface coverages of the vicinal pair of monopodal Ti sites (**IIIb**) and the bipodal Ti site (**IVb**) as a function of the dihedral angle in the vicinal silanol site (low pressure conditions). The shaded regions show calculated uncertainties in the predicted coverages, based on errors of ±2 kJ/mol, ±4 kJ/mol, ±6 kJ/mol, and ±8 kJ/mol in the DFT-computed free energies. Ref. 48 reported maximum differences of 8 kJ/mol between calculated barriers from DFT (using the ωB97X-D functional) and CCSD(T). Their analysis was based on a suite of reactions (oxidation, hydration, metathesis, and epoxidation) catalyzed by Mo and W.

For all dihedral angles, the vicinal pair of monopodal Ti sites (**IIIb**) is the dominant surface species. The surface coverage of bipodal **IVb** is small (<0.15) for dihedral angles up to 20°, and negligible for angles greater than 20°. Fig. 4.6 also shows the estimated uncertainty in the calculated coverages, based on an estimated uncertainty of ±8 kJ/mol in the DFT-computed free energy differences.48 For dihedral angles 30°, the large uncertainty allows for both species to be present in significant amounts at equilibrium. For dihedral angles larger than 30°, we can conclude with confidence that **IIIb** is more abundant. The computed dependence of the grafting free energy on the dihedral angle ϕ can be used to make predictions about experimental outcomes, provided we know the distribution of dihedral angles for the silica. An approximate distribution of vicinal dihedrals, $\rho(\phi)$, was obtained from the atomistic silica slab models described in section 4.4. Fig. 4.7 shows the distribution of vicinal dihedral angles (ϕ).

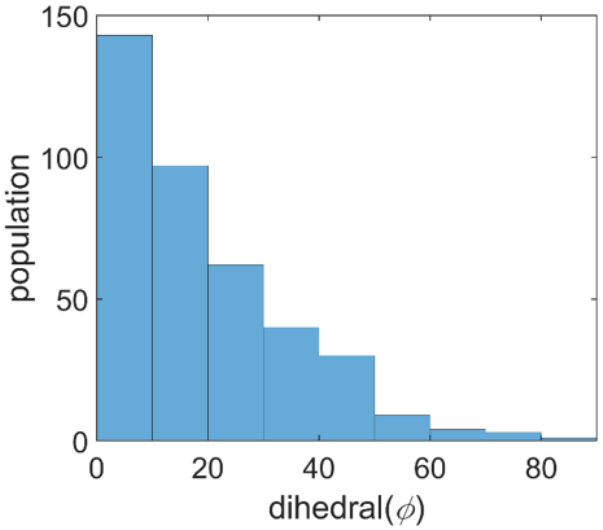


Figure 4.7: Distribution of vicinal dihedral angles (ϕ) extracted from the atomistic silica slab models. The histogram bin-width is 10°.

The average coverages of **IIIb** and **IVb** (averaged over the distribution of the dihedral angles) are given in eq. 4.7:

$$\overline{\theta}_{\mathbf{IIIb}} = \int \rho(\phi)\theta_{\mathbf{IIIb}}(\phi)d\phi$$
$$\overline{\theta}_{\mathbf{IVb}} = \int \rho(\phi)\theta_{\mathbf{IVb}}(\phi)d\phi \tag{4.7}$$

where $\theta_{\mathbf{IIIb}}(\phi)$ and $\theta_{\mathbf{IVb}}(\phi)$ are the ϕ-dependent surface coverages of **IIIb** and **IVb**, respectively (see fig. 4.6), and $\overline{\theta}_{\mathbf{IIIb}}$ and $\overline{\theta}_{\mathbf{IVb}}$ are the average coverages of **IIIb** and **IVb**, respectively. The predicted average coverages ($\overline{\theta}_{\mathbf{IIIb}}$ and $\overline{\theta}_{\mathbf{IVb}}$) can be compared with experimental observations about the relative abundances of **IIIb** and **IVb**. The resulting fractional coverages of **IIIb** and **IVb** are 0.97 and 0.03, respectively. With an uncertainty of ± 8 kJ/mol in the calculated free energy differences, the range of fractional coverages are $0.65 < \overline{\theta}_{\mathbf{IIIb}} < 1.00$ and $0.00 < \overline{\theta}_{\mathbf{IIIb}} < 0.35$. The ranges are large because the uncertainty in the predicted coverages is large for sites with small dihedral angles, and the silica model has a large fraction of such sites. In agreement with these predictions, the experimental coverages are $\overline{\theta}_{\mathbf{IIIb}} \approx 1$ and $\overline{\theta}_{\mathbf{IVb}} \approx 0$ for both silica materials (A380-100 and A380-500). Although our model predicts a predominance of **IIIb** states at equilibrium, in agreement with our experiments,[17] other studies have reported the formation of a mixture of **IIIb** and **IVb** sites.[48–50] To understand how a mixture might result in kinetic **IVb** states, we develop a population balance model of the complete grafting process at realistic conditions in the next section.

4.5.3 *Ab initio* population balance model of grafting

In the grafting experiment, precursor molecules ($TiCl_4$) react with a distribution of silanol sites. As shown in section 4.5.2, the vicinal dihedral angle has a significant effect on the relative free energies of **IIIb** and **IVb**. Hence, we anticipate that the vicinal dihedral will also influence the free energies of other intermediates and transition states

in the grafting pathway. By extension, the grafting kinetics will be a function of the dihedral angle. Nearly all the vicinal pairs are too far from other silanols to influence each other. However, HCl liberated upon grafting can influence the driving force and kinetics of grafting at other sites. In this section we develop a population balance model coupled to a species balance for HCl to simulate $TiCl_4$ grafting kinetics in a batch reactor. The following population balance equations describe the evolution of the different species (**Ib**, **IIb**, **IIIb**, and **IVb**) as functions of time and vicinal dihedral angle (ϕ). Note that P_{HCl} is also a function of time.

$$\begin{aligned}\frac{d\theta_{\mathbf{IIb}}(\phi,t)}{dt} &= k_1(\phi)(1-\theta_{\mathbf{IIb}}(\phi,t)-\theta_{\mathbf{IIIb}}(\phi,t)-\theta_{\mathbf{IVb}})(\phi,t))\frac{P_{sat}}{k_BT} \\ &- k_{-1}(\phi)\theta_{\mathbf{IIb}}(\phi,t)\frac{P_{HCl}}{k_BT} - k_2(\phi)\theta_{\mathbf{IIb}}(\phi,t)\frac{P_{sat}}{k_BT} + k_{-2}(\phi)\theta_{\mathbf{IIIb}}(\phi,t)\frac{P_{HCl}}{k_BT} \\ &- k_4(\phi)\theta_{\mathbf{IIb}}(\phi,t) + k_{-4}(\phi)\theta_{\mathbf{IVb}}(\phi,t)\frac{P_{HCl}}{k_BT}\end{aligned} \tag{4.8}$$

$$\begin{aligned}\frac{d\theta_{\mathbf{IIIb}}(\phi,t)}{dt} &= k_2(\phi)\theta_{\mathbf{IIb}}(\phi,t)\frac{P_{sat}}{k_BT} - k_{-2}(\phi)\theta_{\mathbf{IIb}}(\phi,t)\frac{P_{HCl}}{k_BT} - k_3(\phi)\theta_{\mathbf{IIIb}}(\phi,t) \\ &+ k_{-3}(\phi)\theta_{\mathbf{IVb}}(\phi,t)\frac{P_{sat}}{k_BT}\end{aligned} \tag{4.9}$$

$$\begin{aligned}\frac{d\theta_{\mathbf{IVb}}(\phi,t)}{dt} &= k_4(\phi)\theta_{\mathbf{IIb}}(\phi,t) - k_{-4}(\phi)\theta_{\mathbf{IVb}}(\phi,t)\frac{P_{HCl}}{k_BT} + k_3(\phi)\theta_{\mathbf{IIIb}}(\phi,t) \\ &- k_{-3}(\phi)\theta_{\mathbf{IVb}}(\phi,t)\frac{P_{sat}}{k_BT}\end{aligned} \tag{4.10}$$

Here, σ is the number density of surface silanols, a is the area per unit mass of silica, m is the mass of silica (σ is multiplied by a factor of 1/2 because each vicinal site contributes two silanols), V is the headspace volume of the reactor, ϕ is the vicinal site dihedral angle, t is time, and $\theta_{\mathbf{IIb}}$, $\theta_{\mathbf{IIIb}}$, and $\theta_{\mathbf{IVb}}$ are the fractions of **IIb**, **IIIb**,

and **IVb**, respectively. The population of **Ib**, $\theta_{\mathbf{Ib}}$, can be determined using species balance ($\theta_{\mathbf{Ib}}1 - \theta_{\mathbf{IIb}} - \theta_{\mathbf{IIIb}} - \theta_{\mathbf{IVb}}$). k_i are rate constants of different steps in the grafting pathway (see section C.1 of the appendix). The rate constants are calculated using harmonic transition state theory with quantized vibrations. Some of the rate constants are a function of the vicinal dihedral (ϕ), as discussed later. P_{sat} is constant and equal to the saturation pressure of $TiCl_4$ at 298.15 K. We calculate $P_{HCl}(t)$ by summing over reacted silanols as

$$P_{HCl}(t) = \frac{\sigma}{2}\frac{amk_BT}{V}\int \rho(\phi)\{\theta_{\mathbf{IIb}}(\phi,t) + 2(\theta_{\mathbf{IIb}}(\phi,t) + \theta_{\mathbf{IVb}}(\phi,t))\}d\phi \tag{4.11}$$

In the numerical simulation, eq. 11 should be carefully discretized because the ϕ-dependences in $\rho(\phi)$ and $\theta_i(\phi,t)$ are discretized differently. See section C.4 of the appendix for details. Now, to solve the population balance model we need a model for $k_i(\phi)$ and a distribution of vicinal silanol sites $\rho(\phi)$. To obtain the dependence of k_i on ϕ, we compute the full free energy pathway for 5 vicinal sites (ϕ = 0°, 20°, 40°, 56°, and 60°). We find that the free energies of **I_II TSb**, **II_IV TSb**, **III_IV TSb**, and **IVb** relative to the bare silica site (**Ib**), are a strong function of ϕ. Whereas, the free energies of **IIb**, **IIIb**, and **II_IV TSb** relative to **Ib** are almost constant and do not strongly depend on ϕ. We parametrize polynomial functions to fit free energies of species with a strong dependence on ϕ. And we set the free energies of species with a weak dependence on ϕ constant (equal to their average free energy). We express k_i as a function of free energy barriers using harmonic TST to obtain parametrized models for the rate constants ($k_i(\phi)$). The free energies and the polynomial fits are shown in section C.3 of the appendix. We use the atomistic silica model described in section 4.4 to estimate $\rho(\phi)$. Finally, we use the following physical parameters from the grafting

experiment setup: a 300 mL headspace volume, a 15 mg sample of silica, and 340 m^2/g surface area of silica. All the sites are initialized in the unreacted state (**Ib**), and we use the ode15s solver in MATLAB to solve the ODEs.[51] Fig. 4.8 shows the evolution of $\theta_{\mathbf{Ib}}$, $\theta_{\mathbf{IIb}}$, $\theta_{\mathbf{IIIb}}$, and $\theta_{\mathbf{IVb}}$ as functions of ϕ and t for a silanol density of 1.4 silanols/nm^2 (A380-500). Fig. 4.9 shows the evolution of P_{HCl}, $\overline{\theta}_{\mathbf{IIb}}$, $\overline{\theta}_{\mathbf{IIIb}}$, and $\overline{\theta}_{\mathbf{IVb}}$.

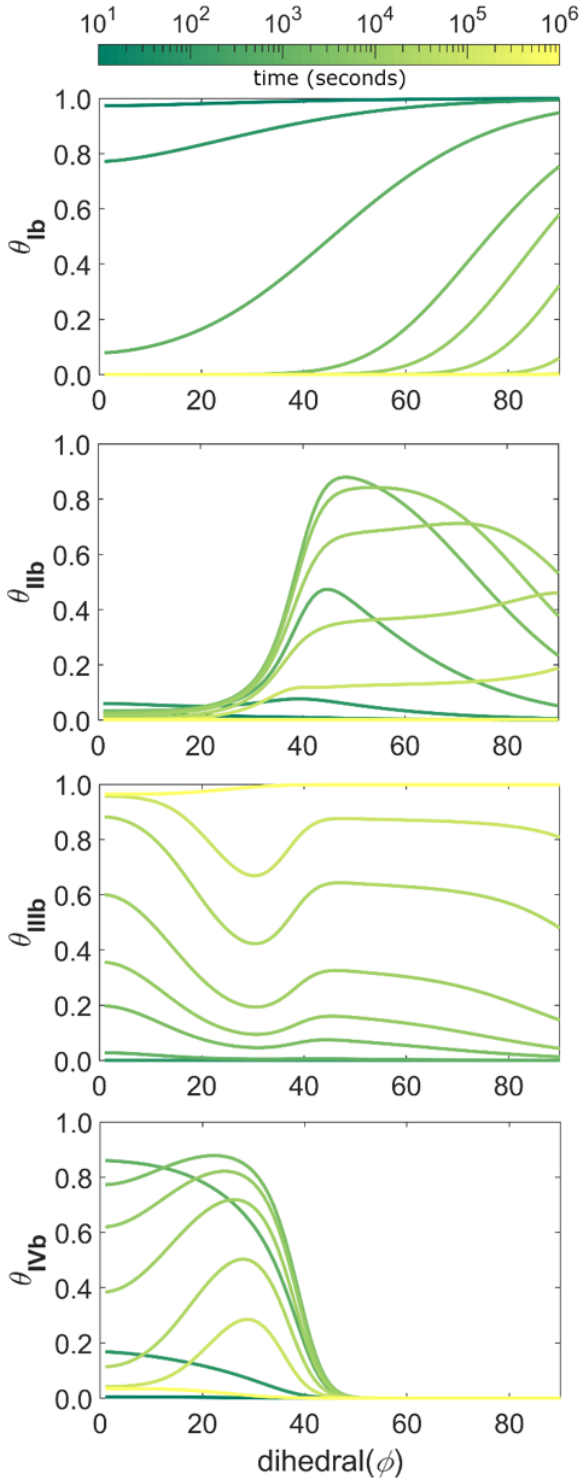


Figure 4.8: Evolving population of **Ib**, **IIb**, **IIIb**, and **IVb** as predicted by the population balance model. The population balance model has been solved at 298.15 K for 300 mL headspace volume, a 15 mg silica sample, 340 m^2/g surface area of silica, and 1.4 silanols/nm^2 silanol number density.

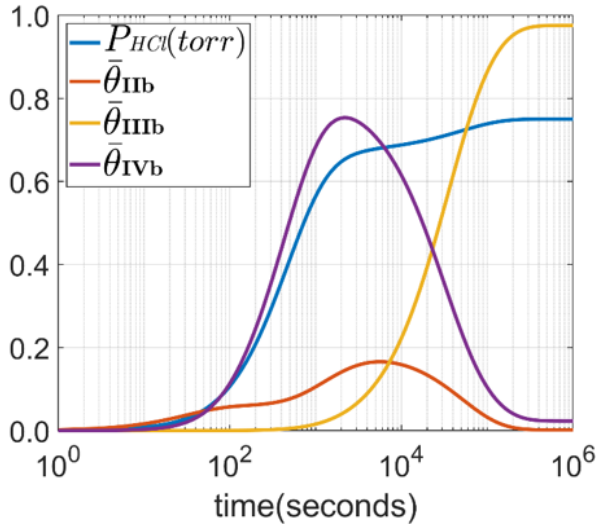


Figure 4.9: Predicted evolution of HCl pressure, $\bar{\theta}_{\mathbf{IIb}}$, $\bar{\theta}_{\mathbf{IIIb}}$, and $\bar{\theta}_{\mathbf{IVb}}$ as a function of time. The population balance model has been solved at 298.15 K for 300 mL headspace volume, a 15 mg silica sample, 340 m^2/g surface area of silica, and 1.4 silanols/nm^2 silanol number density.

Small dihedral sites (ϕ<40°) react first, over times of approximately 10^3 seconds giving intermediate **IIb**. This is followed by rapid conversion to **IVb**. This is further followed by slow conversion to **IIIb** over times of approximately 10^4 – 10^5 seconds. The **IVb**→**IIIb** rate constant ($k_{-3}(\phi)$) monotonically decreases on increasing ϕ from 0° to 40°. Hence, the time spent by sites in state **IVb** before converting to **IIIb** monotonically increases from 0° to 40°. For $0° < \phi < 20°$, sites convert from **IVb** to **IIIb** quickly, while for $20° < \phi < 40°$, sites remain in state **IVb** for a long time. As shown in section 4.5.2, the calculations predict that nearly all **IVb** sites, even those with $20° < \phi < 40°$ will eventually convert to structure **IIIb** at equilibrium. Large dihedral sites ($\phi > 40°$) take longer to convert to the intermediate **IIb** sites (approximately 10^4 seconds), but these **IIb** sites directly convert to **IIIb**. The **IIb**→**IVb** pathway is unfavorable for large dihedral sites because of increasing ring strain. On reaching equilibrium all such sites ($\phi > 40°$) convert to **IIIb**. Sites near the $\phi = 40°$ boundary can react via both

mechanisms. Some of them react as **Ib**→**IIb**→**IVb**⇌**IIIb**, while some of them react as **Ib**→**IIb**→**IIIb**. Our results show that the dihedral angle ϕ at a vicinal site affects the kinetics of grafting and the grafting mechanism. Small dihedral sites ($\phi < 20°$) react as **Ib**→**IIb**→**IVb**⇌**IIIb**, with rapid conversion from **IVb** to **IIIb**. Sites with medium dihedral angles ($20° < \phi < 40°$) react as **Ib**→**IIb**→**IVb**⇌**IIIb**, with slow conversion from **IVb** to **IIIb**. Large dihedral sites ($\phi > 40°$) react via the mechanism **Ib**→**IIb**→**IIIb**. A complete conversion of **IIIb** (at long times) agrees with experimental measurements on both silica samples (A380-100 and A380-500). For the atomistic silica model used in this study, about 62% sites react via the first mechanism (**Ib**→**IIb**→**IVb**⇌**IIIb**), 26% sites react via the second mechanism (**Ib**→**IIb**→**IVb**⇌**IIIb**), and the rest (12%) react via the third mechanism (**Ib**→**IIb**→**IIIb**). Grafting completes in about 10^5 seconds and the final HCl pressure is ca. 0.75 Torr. The total fraction of reacted silanols can be calculated as $(P_{HCl}V)/(\sigma a m k_B T)$. We get $(P_{HCl}V)/(\sigma a m k_B T) = 0.99$ on completion, hence almost all silanols react. Although we cannot ensure accurate predictions because of the inexact silica models and DFT calculations, our calculations show how catalysts and insights derived from grafting experiments can depend on the time provided for the grafting kinetics. Fig. 4.9 shows that the Ti:silanol ratio is a function of grafting time. Ca. 87% of the sites react by 2×10^3 seconds. If grafting is stopped at 2×10^3 seconds most of the sites will end up in state **IVb** with a Ti:silanol ratio of ca. 1:2. Most of the sites convert to **IIIb** if the reaction is allowed to go to completion and the Ti:silanol ratio will be 1:1. Stopping the reaction between 2×10^3 and 10^5 seconds will lead to intermediate Ti:silanol ratios. These results can be used to explain the variability in the Ti:silanol ratios observed in different Ti grafting experiments in the literature.[17,48–50] The simple thermodynamic description in section 4.5.1 makes a number of predictions that haven't yet been tested. For example, if $TiCl_4$ is completely evacuated, **IVb** should become the dominant species. The model presented here can predict the distribution of

times required for the conversion back to state **IVb**.

This demonstrates the importance of considering a distribution of sites in computational studies of atomically dispersed catalysts on amorphous supports. Interpretations drawn from calculations on single sites can lead to incorrect conclusions about grafting kinetics, grafting mechanisms, and subsequent catalytic activities.

4.6 Conclusions

Characterization of metals grafted to amorphous silica using spectroscopic techniques has been challenging because of the absence of long-range structural order. Computational studies have also been rare, with most of these relying on single site models even though these disordered materials clearly have a distribution of sites.[52] In this work we used DFT calculations, a computational model of silica, and a population balance model to develop an ab initio model of the $TiCl_4$ grafting kinetics on the ensemble of vicinal silanols. We provide an efficient and theoretically rigorous computational framework that predicts kinetics for the entire ensemble of sites from a modest set of DFT calculations. Our results predict the silanol population density as a function of a key dihedral angle, how the kinetics and thermodynamics of key grafting intermediates depend on the dihedral angle, and how the populations of different grafting intermediates evolves in time. The model predicts, in agreement with experimental results, that almost all sites ultimately yield a vicinal pair of monopodally-grafted Ti sites. However, in contrast to commonly assumed single-exponential models of the grafting kinetics,[53–56] the grafting reaction proceeds via different mechanisms at different rates, resulting in non-exponential kinetics for the ensemble of sites. Specifically, the model predicts that ca. 87% of the sites initially react to form a bipodal $Cl_2Ti(OSi{\equiv})_2$ site giving a Ti:Silanol ratio of 1:2. However, if the grafting reaction continues for 10-100 times longer under $TiCl_4$ vapor, nearly

all sites convert to vicinal pairs of monopodal $Cl_3TiOSi\equiv$ sites giving a Ti:Silanol ratio of 1:1. These predictions may explain why some grafting experiments report Ti:silanol ratios near 1:1 while others report ratios near 1:2.[17,48–50] The model also makes predictions that haven't yet been tested, e.g. it should be possible to drive equilibrium toward the bipodal $Cl_2Ti(OSi\equiv)_2$ sites by evacuation of $TiCl_4$ after grafting. In future work, these computational techniques should help to predict the structures and populations catalyst sites that result from grafting of precursors on amorphous silica, ultimately providing a route to predictions about their catalyst activity.

Bibliography

1. Vansant, E. E., Van Der Voort, P. & Vrancken, K. C. *Characterization and Chemical Modification of the Silica Surface* (1995).

2. McDaniel, M. P. A Review of the Phillips Supported Chromium Catalyst and Its Commercial Use for Ethylene Polymerization. English. *Adv. Catal.* **53,** 123–606. https://www.doi.org/https://doi.org/10.1016/S0360-0564(10)53003-7 (2010).

3. Cativiela, C., Fraile, J., García, J. & Mayoral, J. A new titanium-silica catalyst for the epoxidation of alkenes. *Journal of Molecular Catalysis A: Chemical* **112,** 259–267. ISSN: 1381-1169. https://doi.org/10.1016/1381-1169(96)00131-8 (1996).

4. Rhers, B. *et al.* A Well-Defined, Silica-Supported Tungsten Imido Alkylidene Olefin Metathesis Catalyst. *Organometallics* **25,** 3554–3557. ISSN: 0276-7333. https://doi.org/10.1021/om060279d (2006).

5. Gupta, T. K. & Jean, J.-H. Principles of the development of a silica dielectric for microelectronics packaging. *Journal of Materials Research* **11,** 243–263. ISSN: 2044-5326. https://doi.org/10.1557/JMR.1996.0030 (1996).

6. Tomozawa, M. & Hepburn, R. W. Surface structural relaxation of silica glass: a possible mechanism of mechanical fatigue. *Journal of Non-Crystalline Solids* **345-**

346, 449–460. ISSN: 0022-3093. https://doi.org/10.1016/j.jnoncrysol.2004.08.065 (2004).

7. Lai, C.-Y. *et al.* A Mesoporous Silica Nanosphere-Based Carrier System with Chemically Removable CdS Nanoparticle Caps for Stimuli-Responsive Controlled Release of Neurotransmitters and Drug Molecules. *Journal of the American Chemical Society* **125,** 4451–4459. ISSN: 0002-7863. https://doi.org/10.1021/ja0286501 (2003).

8. Pavan, C. *et al.* Nearly free surface silanols are the critical molecular moieties that initiate the toxicity of silica particles. *Proceedings of the National Academy of Sciences* **117,** 27836. https://doi.org/10.1073/pnas.2008006117 (2020).

9. Elliott, S. R. Medium-range structural order in covalent amorphous solids. *Nature* **354,** 445–452. ISSN: 1476-4687. https://doi.org/10.1038/354445a0 (1991).

10. Sørensen, S. S., Biscio, C. A. N., Bauchy, M., Fajstrup, L. & Smedskjaer, M. M. Revealing hidden medium-range order in amorphous materials using topological data analysis. *Science Advances* **6,** eabc2320. https://doi.org/10.1126/sciadv.abc2320 (2020).

11. Liu, C. C. & Maciel, G. E. The Fumed Silica Surface: A Study by NMR. *Journal of the American Chemical Society* **118,** 5103–5119. ISSN: 0002-7863. https://doi.org/10.1021/ja954120w (1996).

12. Taha, Z. A., Deguns, E. W., Chattopadhyay, S. & Scott, S. L. Formation of Digallium Sites in the Reaction of Trimethylgallium with Silica. *Organometallics* **25,** 1891–1899. ISSN: 0276-7333. https://doi.org/10.1021/om051034o (2006).

13. Fleischman, S. D. & Scott, S. L. Evidence for the Pairwise Disposition of Grafting Sites on Highly Dehydroxylated Silicas via Their Reactions with $Ga(CH_3)_3$. *J. Am. Chem. Soc.* **133,** 4847–4855. https://www.doi.org/10.1021/ja108905p (2011).

14. Zhong, L. *et al.* Spectroscopic and structural characterization of $Cr(II)/SiO_2$ active site precursors in model Phillips polymerization catalysts. *J. Catal.* **293,** 1–12. https://www.doi.org/10.1016/j.jcat.2012.05.014 (2012).

15. Coperet, C. *et al.* Surface Organometallic and Coordination Chemistry toward Single-Site Heterogeneous Catalysts: Strategies, Methods, Structures, and Activities. *Chem. Rev.* **116,** 323–421. https://www.doi.org/10.1021/acs.chemrev.5b00373 (2016).

16. Mania, P., Verel, R., Jenny, F., Hammond, C. & Hermans, I. Thermal Restructuring of Silica-Grafted TiClx Species and Consequences for Epoxidation Catalysis. *Chemistry - A European Journal* **19,** 9849–9858. ISSN: 0947-6539. https://doi.org/10.1002/chem.201300842 (2013).

17. Khan, S. W., Deguns, E. W., Peters, B. & Scott, S. L. Quantifing abundances of grafted [(SiO)4-xTiClx] sites on amorphous silica. *in preparation.*

18. Del Rosal, I., Gerber, I. C., Poteau, R. & Maron, L. Grafting of Lanthanide Complexes on Silica Surfaces: A Theoretical Investigation. *The Journal of Physical Chemistry A* **114,** 6322–6330. ISSN: 1089-5639. https://doi.org/10.1021/jp101495n (2010).

19. Cheng, R., Liu, X., Fang, Y., Terano, M. & Liu, B. High-resolution 29Si CP/MAS solid state NMR spectroscopy and DFT investigation on the role of geminal and single silanols in grafting chromium species over Phillips Cr/silica catalyst. *Applied Catalysis A: General* **543,** 26–33. ISSN: 0926-860X. https://doi.org/10.1016/j.apcata.2017.05.011 (2017).

20. Hu, Z. & Turner, C. H. Initial Surface Reactions of TiO_2 Atomic Layer Deposition onto SiO_2 Surfaces: Density Functional Theory Calculations. *The Journal of Physical Chemistry B* **110,** 8337–8347. ISSN: 1520-6106. https://doi.org/10.1021/jp060367b (2006).

21. Hu, Z. & Turner, C. H. Atomic Layer Deposition of TiO_2 from TiI_4 and H_2O onto SiO_2 Surfaces: Ab Initio Calculations of the Initial Reaction Mechanisms. *Journal of the American Chemical Society* **129,** 3863–3878. ISSN: 0002-7863. https://doi.org/10.1021/ja066529z (2007).

22. Skowronska-Ptasinska, M. D., Duchateau, R., van Santen, R. A. & Yap, G. P. A. Methyl Aluminosilsesquioxanes, Models for Lewis Acidic Silica-Grafted Methyl Aluminum Species. *Organometallics* **20,** 3519–3530. ISSN: 0276-7333. https://doi.org/10.1021/om0102596 (2001).

23. Guesmi, H., Gryboś, R., Handzlik, J. & Tielens, F. Characterization of molybdenum monomeric oxide species supported on hydroxylated silica: a DFT study. *Physical Chemistry Chemical Physics* **16,** 18253–18260. ISSN: 1463-9076. http://dx.doi.org/10.1039/C4CP02296C (2014).

24. Gueddida, S., Lebégue, S. & Badawi, M. Interaction between transition metals (Co, Ni, and Cu) systems and amorphous silica surfaces: A DFT investigation. *Applied Surface Science* **533,** 147422. ISSN: 0169-4332. https://doi.org/10.1016/j.apsusc.2020.147422 (2020).

25. Guesmi, H. & Tielens, F. Chromium oxide species supported on silica: A representative periodic DFT model. *J Phys Chem C* **116,** 994–1001. https://doi.org/10.1021/jp209680r (2012).

26. Chizallet, C. Toward the Atomic Scale Simulation of Intricate Acidic Aluminosilicate Catalysts. *ACS Catalysis* **10,** 5579–5601. https://doi.org/10.1021/acscatal.0c01136 (2020).

27. Halbert, S., Ispas, S., Raynaud, C. & Eisenstein, O. Modelling the surface of amorphous dehydroxylated silica: the influence of the potential on the nature and density of defects. *New Journal of Chemistry* **42,** 1356–1367. ISSN: 1144-0546. http://dx.doi.org/10.1039/C7NJ03922K (2018).

28. Leydier, F. *et al.* Brønsted acidity of amorphous silica–alumina: The molecular rules of proton transfer. *Journal of Catalysis* **284,** 215–229. ISSN: 0021-9517. https://doi.org/10.1016/j.jcat.2011.08.015 (2011).

29. Berro, Y. *et al.* Imprinting isolated single iron atoms onto mesoporous silica by templating with metallosurfactants. *Journal of Colloid and Interface Science* **573,** 193–203. ISSN: 0021-9797. https://doi.org/10.1016/j.jcis.2020.03.095 (2020).

30. Handzlik, J. & Ogonowski, J. Structure of Isolated Molybdenum(VI) and Molybdenum(IV) Oxide Species on Silica: Periodic and Cluster DFT Studies. *The Journal of Physical Chemistry C* **116,** 5571–5584. https://www.doi.org/10.1021/jp207385h (2012).

31. Jystad, A. M., Biancardi, A. & Caricato, M. Simulations of Ammonia Adsorption for the Characterization of Acid Sites in Metal-Doped Amorphous Silicates. *J. Phys. Chem. C* **121,** 22258–22267. https://www.doi.org/10.1021/acs.jpcc.7b08113 (2017).

32. Khan, S. A., Vandervelden, C. A., Scott, S. L. & Peters, B. Grafting metal complexes onto amorphous supports: From elementary steps to catalyst site popula-

tions: Via kernel regression. *React. Chem. Eng.* **5,** 66–76. https://www.doi.org/10.1039/c9re00357f (2020).

33. Vandervelden, C. A., Khan, S. A., Scott, S. L. & Peters, B. Site-averaged kinetics for catalysts on amorphous supports: an importance learning algorithm. *React. Chem. Eng.* **5,** 77–86. https://www.doi.org/10.1039/C9RE00356H (2020).

34. Vandervelden, C. A., Khan, S. A. & Peters, B. Importance learning estimator for the site-averaged turnover frequency of a disordered solid catalyst. *The Journal of Chemical Physics* **153,** 244120. https://doi.org/10.1063/5.0037450 (2020).

35. Chai, J.-D. & Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **10,** 6615–6620. https://www.doi.org/10.1039/B810189B (2008).

36. Feller, D. The role of databases in support of computational chemistry calculations. *J. Comput. Chem.* **17,** 1571–1586. https://www.doi.org/10.1002/(SICI)1096-987X(199610)17:13%3C1571::AID-JCC9%3E3.0.CO;2-P (1996).

37. Schuchardt, K. L. *et al.* Basis Set Exchange: A Community Database for Computational Sciences. *J. Chem. Inf. Model.* **47,** 1045–1052. https://www.doi.org/10.1021/ci600510j (2007).

38. Godbout, N., Salahub, D. R., Andzelm, J. & Wimmer, E. Optimization of Gaussian-type basis sets for local spin density functional calculations. Part I. Boron through neon, optimization technique and validation. *CJC University of Calgary 50th Anniversary Virtual Issue* **01,** 560–571. https://doi.org/10.1139/v92-079 (1992).

39. Frisch, M. J. *et al.* Gaussian16. Gaussian Inc. Wallingford CT 2016.

40. Luchinskii, G. & Khim, Z. F. *Physico-chemical study of titanium halides 1. Parameters of phase transitions of titanium tetrachloride* 593–598 (1966).

41. Goldsmith, B. R., Sanderson, E. D., Bean, D. & Peters, B. Isolated Catalyst Sites on Amorphous Supports: A Systematic Algorithm for Understanding Heterogeneities in Structure and Reactivity. *J. Chem. Phys.* **138,** 204105. https://www.doi.org/10.1063/1.4807384 (2013).

42. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics* **117,** 1–19. ISSN: 0021-9991. https://doi.org/10.1006/jcph.1995.1039 (1995).

43. Van Beest, B. W. H., Kramer, G. J. & van Santen, R. A. Force fields for silicas and aluminophosphates based on ab initio calculations. *Physical Review Letters* **64,** 1955–1958. https://doi.org/10.1103/PhysRevLett.64.1955 (1990).

44. Wimalasiri, P., Nguyen, J., Senanayake, H. S., Laird, B. & Thompson, W. H. Model amorphous silica surfaces for use in simulations of dynamics and catalysis. *in preparation.*

45. Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. V., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81,** 3684–3690. ISSN: 0021-9606. https://doi.org/10.1063/1.448118 (1984).

46. Hu, L. & Chen, H. Assessment of DFT Methods for Computing Activation Energies of Mo/W-Mediated Reactions. *Journal of Chemical Theory and Computation* **11,** 4601–4614. ISSN: 1549-9618. https://doi.org/10.1021/acs.jctc.5b00373 (2015).

47. Ewing, C. S. *et al.* Impact of Support Interactions for Single-Atom Molybdenum Catalysts on Amorphous Silica. *Ind Eng Chem Res* **55,** 12350–12357. https://www.doi.org/10.1021/acs.iecr.6b03558 (2016).

48. Haukka, S., Lakomaa, E. L., Jylha, O., Vilhunen, J. & Hornytzkyj, S. Dispersion and distribution of titanium species bound to silica from titanium tetrachloride. *Langmuir* **9,** 3497–3506. https://doi.org/10.1021/la00036a026 (1993).

49. Haukka, S., Lakomaa, E. L. & Root, A. An IR and NMR study of the chemisorption of titanium tetrachloride on silica. *The Journal of Physical Chemistry* **97,** 5085–5094. https://doi.org/10.1021/j100121a040 (1993).

50. Kytökivi, A. & Haukka, S. Reactions of HMDS, $TiCl_4$, $ZrCl_4$, and $AlCl_3$ with Silica As Interpreted from Low-Frequency Diffuse Reflectance Infrared Spectra. *The Journal of Physical Chemistry B* **101,** 10365–10372. https://www.doi.org/10.1021/jp971822u (1997).

51. Neidinger, R. D. Introduction to Automatic Differentiation and MATLAB Object-Oriented Programming. *SIAM Review* **52,** 545–563. ISSN: 0036-1445. https://doi.org/10.1137/080743627 (2010).

52. He, S.-Y. *et al.* Screening silica-confined single-atom catalysts for nonoxidative conversion of methane. *The Journal of Chemical Physics* **154,** 174706. ISSN: 0021-9606. https://doi.org/10.1063/5.0048962 (2021).

53. Cullen, R. J. *et al.* Spontaneous Grafting of Nitrophenyl Groups on Amorphous Carbon Thin Films: A Structure–Reactivity Investigation. *Chemistry of Materials* **24,** 1031–1040. ISSN: 0897-4756. https://doi.org/10.1021/cm2030262 (2012).

54. Chaudhari, C. V. *et al.* Ethylene vinyl acetate based radiation grafted hydrophilic matrices: Process parameter standardization, grafting kinetics and characterization. *Radiation Physics and Chemistry* **125,** 213–219. ISSN: 0969-806X. https://doi.org/10.1016/j.radphyschem.2016.04.017 (2016).

55. Sha, X., Xu, X., Sohlberg, K., Loll, P. J. & Penn, L. S. Evidence that three-regime kinetics is inherent to formation of a polymer brush by a grafting-to approach. *RSC Advances* **4,** 42122–42128. http://dx.doi.org/10.1039/C4RA05663A (2014).

56. Siefering, K. L. & Griffin, G. L. Kinetics of Low Pressure Chemical Vapor Deposition of TiO_2 from Titanium Tetraisopropoxide. *Journal of The Electrochemical Society* **137,** 814–818. http://dx.doi.org/10.1149/1.2086561 (1990).

Chapter 5

On the various computational models of silica: a statistical analysis of structural differences

5.1 Introduction

Developments in computational catalysis have advanced our understanding of homogeneous and ordered heterogenous catalysts tremendously.[1–14] In contrast, amorphous catalysts pose many challenges and remain elusive on several fronts. These challenges primarily arise because of a quenched distribution of active site structures which is a function of the non-equilibrium preparation history of these materials.[15,16] Atomically dispersed catalysts on amorphous supports are one such class of catalysts. Where the active sites, often transition metals, are dispersed on amorphous supports like silica (SiO_2).[17–21] Silica is one of the most commonly used support because of its inert nature, mechanical strength, and thermal stability.[22–24] Amorphous silica surfaces are terminated by siloxanes (≡SiOSi≡) and silanols (≡SiOH). Silanols are generally classified as isolated,

geminal, or vicinal, depending on their connectivity to other silanols.[25] These classifications are described in more detail in section 5.2.2. Active metal sites are grafted onto amorphous silica via the reaction of surface silanols with fluid phase organometallic precursors.[20,21,26–28] Industrially important examples of these catalysts include, WO_3/SiO_2 for olefin metathesis,[29] Cr/SiO_2 for the polymerization of ethene,[17] etc. A complete in silico model of atomically dispersed amorphous catalysts will require the following: 1) an accurate atomistic model of the amorphous support, 2) methods to model the grafting of metal complexes onto amorphous supports, and 3) methods to efficiently calculate site-averaged kinetic properties (activation barriers and turnover frequencies) of the grafted catalyst. Recently, methods have been developed to efficiently model the grafting of metal ions to amorphous supports and calculate site-averaged kinetic properties of the grafted catalyst.[30–32] These methods start from an atomistic model of amorphous silica and model the subsequent grafting and catalysis steps. Given a model of the amorphous support, the evolving population of catalytic sites during grafting is predicted using machine learning (ML) parametrized population balance models.[30] The site-averaged kinetic properties of the grafted catalyst are calculated using importance learning (importance sampling + ML).[31,32]

A few atomistic models of amorphous silica have been developed.[33–37] These models are typically prepared by melting bulk silica at a high temperature ($\sim 3000 - 7000$ K) followed by rapidly quenching the melt. Following this, surfaces are cleaved and functionalized by hydroxyl groups. Models with different silanol densities are generated by incrementally condensing pairs of silanols. Different studies have used different protocols and parameter choices and make many ad hoc assumptions. And many questions remain unanswered. For example, what should be the ideal melting temperature and quench rate? What is the best procedure to functionalize the surface of the quenched melt and condense silanols? Which force-field or ab initio method should be used? Moreover, real

silicas are not synthesized via this process.

Commercial silica supports used for atomically dispersed catalysts can be non-porous or mesoporous. Aerosil is an example of a non-porous silica support.[38] It is synthesized via flame pyrolysis of $SiCl_4$.[39] SBA-15[40] and MCM-41[41,42] are examples of mesoporous silica supports. These supports are synthesized by creating an ordered template with parallel cylindrical micelles using surfactants. The template is infused with a silica precursor, like tetraethoxysilane (TEOS), which covers the micelles. Finally, the material is calcined and oxidized to remove the micelles to obtain mesoporous silica. Silicas synthesized via all of these methods can be calcined at different temperatures to condense neighboring silanol groups to achieve the desired surface silanol density.[22]

The alchemical computational protocols attempt to replicate the structure of these experimentally synthesized silicas. Some studies have validated their computational models by comparing computed IR spectra to experimentally obtained spectra.[33,35–37] Some studies have compared the silanol density vs. calcination temperature curves to experiments.[33,34] However, validating the structure of computationally generated atomistic models of amorphous materials is challenging. The absence of long-range order results in broad peaks compared to sharp features found in the spectra of crystalline materials.[43] Hence, unlike crystalline materials, the exact structure of amorphous materials cannot be obtained. Obtaining information about the surface structure of amorphous materials is even more challenging.[44]

Computational studies with different protocols have claimed that their silica models are representative of real silicas. Different protocols can lead to different structural features. Differences in the structural features of the silica support, particularly at the surface, can lead to different grafting kinetics/thermodynamics and different catalytic kinetics and mechanisms.[30,31,45–48] Some studies have also claimed that larger models are more accurate.[49] Development of accurate atomistic models of amorphous silicas is crit-

ical in modeling atomically dispersed amorphous catalysts. While it is not possible to conclusively determine the structure of amorphous silicas, we can compare models generated using different protocols and comment on their similarities and differences. Here, we investigate if different protocols lead to atomistic silica models with different structural features. Finally, we outline a list of metrics based on experimental measurements which future models should be validated against.

5.2 Methods

5.2.1 Silica models

We use silica models from the following studies: Ugliengo et al.,[35] Comas-Vives,[33] and Tielens et al.[37]. The models of Comas-Vives and Ugliengo are non-porous, while Tielen's models are mesoporous. Each study has developed silica models with a range of silanol densities. Schematics summarizing the different preparation protocols are shown in sections D.1,D.2, and D.3 of the appendix.

We reference the silica models using an abbreviation of the author's name followed by their silanol density, separated by an '_'. For example, the 5.8 silanols/nm^2 model from Tielens is called TI_5.8.

5.2.2 Metrics on which models are compared

Amorphous silicas can be characterized by their population of different silanol types (isolated, geminal, vicinal, etc. described in more detail below) and the distribution of local structural features of silanol sites (bond lengths, bond angles etc.). The categorization of silanols into different types corresponds to a discrete distribution. Distributions corresponding to structural features are continuous (continuous distributions of bond

lengths, bond angles, etc.). Differences in the population of different silanol types can influence the nature of the dispersed metal sites and their subsequent catalytic kinetics and mechanisms.[15] Differences in the local structure of silanol sites can also influence grafting and catalysis.[45,46,50]

A reproducible protocol to generate an atomistic model of amorphous silica will correspond to unique distributions of silanol types and local structural features of silanol sites. Models generated using protocols corresponding to different distributions will lead to different grafting kinetics/thermodynamics and different catalytic kinetics and mechanisms. Here, we compare silica models generated using different protocols to investigate if different protocols lead to different distributions.

We compare silica models based on the distribution of silanol types and three structural features of vicinal silanol sites, which have been shown to influence grafting kinetics/thermodynamics and catalysis kinetics and mechanisms.[46,50] Fig. 5.1a shows the classification of silanol types and fig. 5.1b shows the structural features of the vicinal silanol sites considered here. The three categories of silanols considered in this study are described as follows: i) isolated silanol groups, where the Si atom of the silanol (≡SiOH) is bonded to three other O atoms, which in turn are not bonded to any other silanol group, ii) geminal silanol groups, where two OH groups are bonded to a single Si atom, and iii) vicinal silanol groups, where two silanols (≡SiOH) are connected by a Si-O-Si bridge. We count geminal groups by counting the total number of silanols that are part of a geminal group and divide the total by two. Similarly, we count vicinal groups by counting the total number of silanols that are part of a vicinal group and divide the total by two. A silanol can be part of a geminal and vicinal group or two vicinal groups at the same time. All silanols which are not part of either group are categorized as isolated. The three structural features of vicinal sites are described as follow: i) ϕ is the dihedral angle formed by the O-Si-Si-O atoms, ii) d is the distance between the two silanol oxygens, and

iii) θ is the angle formed by the triangle shown in fig. 5.1b. The triangle is formed by rotating the two silanol groups to the same plane while keeping other degrees of freedom fixed.

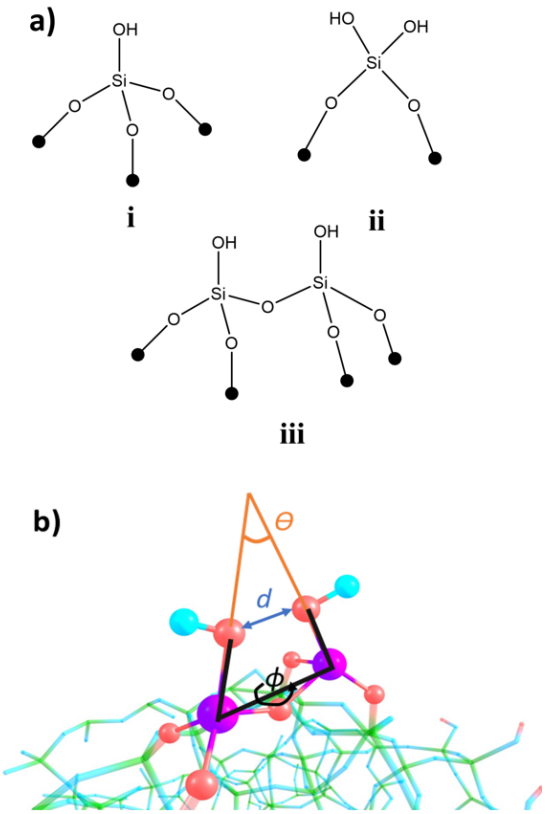


Figure 5.1: **a)** The three categories of silanols considered in this study are shown: i) isolated silanols, ii) geminal silanol groups, and iii) vicinal silanol groups. Black circles represent the extended silica matrix. **b)** The three structural features of vicinal silanol sites (ϕ, d, and θ) which are used to compare silica models are shown. Color scheme: H(blue), O(red), Si(purple), and the extended silica matrix is shown in the stick format colored in green and blue.

The population of isolated, geminal, and vicinal sites is a discrete distribution with 3 categories. And the distributions of the three structural features of the vicinal site (ϕ, d, and θ) are continuous distributions. These distributions can be considered converged for large models of amorphous silica with thousands of silanol sites. However, model sizes published in the literature are small, with tens of sites. Hence, these distributions

cannot be considered converged, thus precluding a direct comparison between models. Sites extracted from small models of amorphous silica are samples drawn from the distribution characteristic of the model preparation protocol. Each silica model corresponds to samples of the four distributions described above. To compare two models, we investigate if the samples are drawn from the same distributions. We test the null hypothesis ($\mathbf{H}_0$) defined as follows: the two samples being compared have been drawn from the same distribution. $\mathbf{H}_0$ is rejected or not based on a statistical hypothesis test. Rejection of $\mathbf{H}_0$ implies that the two samples are not from the same distribution. This further implies that the models from which the samples are extracted are different. Not being able to reject $\mathbf{H}_0$ implies that we cannot conclude that the two models being compared are different. We use two statistical hypothesis testing methods in this study described in the following sections.

5.2.3 Kolmogorov Smirnov test

We use the two-sample Kolmogorov-Smirnov (KS) test to compare samples drawn from continuous probability distributions (ϕ, d, and θ).[51] The two-sample KS test uses the maximum difference between cumulative distribution functions (CDF) as the test statistic given by sup|$\mathrm{CDF}_1(x)$ - $\mathrm{CDF}_2(x)$|. Where, CDF_1 and CDF_2 are the CDFs of the samples being compared and x is the random variable (ϕ, d, and θ). The magnitude of sup|$\mathrm{CDF}_1(x)$ - $\mathrm{CDF}_2(x)$| determines the probability that the two samples are from the same distribution. The probability is expressed as a p-value. It is defined as the probability of observing events at least as extreme as sup|$\mathrm{CDF}_1(x)$ - $\mathrm{CDF}_2(x)$| if H_0 is true. H_0 is rejected if $p < \alpha$ (a pre-determined significance level, usually 5%-10%). We use the two-sample KS test as implemented in MATLAB.[52]

5.2.4 Chi-sq test

The distribution of the types of silanols (isolated, geminal, and vicinal) is discrete. The chi-square test can be used to test $\mathbf{H}_0$ for discrete probability distributions using the chi-square statistic calculated as[53]

$$\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}. \tag{5.1}$$

Where, m is the total number of samples being compared (we compare 2 samples at a time), n is the total number of categories (3 categories in our case: isolated, geminal, and vicinal), O_{ij} is the observed value for the jth category in the ith sample, and E_{ij} is the expected value for the jth category in the ith sample. The expected values, E_{ij}, are calculated by pooling all the samples together as follows:

$$E_{ij} = \frac{\sum_{k=1}^{m} O_{kj}}{\sum_{k=1}^{m} \sum_{l=1}^{n} O_{kl}} \sum_{l=1}^{n} O_{il}. \tag{5.2}$$

For our case (comparing 2 samples with 3 categories), χ^2 is distributed according to the chi-square distribution with 2 degrees of freedom ($\chi^2(2)$) given $\mathbf{H}_0$ is true.[53] Similar to the description in section 5.2.3, a p-value can be calculated. It is the probability of observing a value at least as extreme as χ^2 if $\mathbf{H}_0$ is true. $\mathbf{H}_0$ is rejected if $p < \alpha$ (a pre-determined confidence, usually 5%-10%).

5.3 Results

In this section, we compare the different silica models using the KS and the chi-square tests. It should be noted that, in all studies, models with the largest silanol densities have the largest number of silanols. Smaller models are derived from the larger models by

incrementally condensing silanol pairs to achieve the desired silanol density. We compare models with similar silanol densities. Results of applying the KS test to test $\mathbf{H}_0$ for samples of continuous distributions (ϕ, d, and θ) are shown in Table 5.1.

Table 5.1: p-values on testing $\mathbf{H}_0$: $\rho_1(x) = \rho_2(x)$ using the 2-sample KS test for continuous distributions (ϕ, d, and θ) are shown. Where, x are the continuous random variables (ϕ, d, and θ). The quantities in brackets following silanol densities are sample sizes.

Tielens silanols/nm²	**Comas-Vives silanols/nm²**	p-value		
		$\mathbf{H}_0$: $\rho_1(\phi) = \rho_2(\phi)$	$\mathbf{H}_0$: $\rho_1(d) = \rho_2(d)$	$\mathbf{H}_0$: $\rho_1(\theta) = \rho_2(\theta)$
5.8(47)	5.9(17)	0.10	7.66×10^{-5}	5.94×10^{-4}
4.8(35)	4.6(13)	0.07	3.55×10^{-5}	8.57×10^{-4}
3.6(18)	3.3(5)	0.02	0.05	0.03
2.3(8)	2.4(3)	0.03	0.01	0.03
1.7(2)	1.5(1)	0.20	0.20	0.20
Tielens silanols/nm²	**Ugliengo silanols/nm²** $_0$:	$\mathbf{H}_0$: $\rho_1(\phi) = \rho_2(\phi)$	$\mathbf{H}_0$: $\rho_1(d) = \rho_2(d)$	$\mathbf{H}_0$: $\rho_1(\theta) = \rho_2(\theta)$
5.8(47)	5.4(9)	0.30	0.47	0.67
4.8(35)	4.5(6)	0.39	0.44	0.20
2.3(8)	2.4(2)	0.65	0.07	0.03
Comas-Vives silanols/nm²	**Ugliengo silanols/nm²**	$\mathbf{H}_0$: $\rho_1(\phi) = \rho_2(\phi)$	$\mathbf{H}_0$: $\rho_1(d) = \rho_2(d)$	$\mathbf{H}_0$: $\rho_1(\theta) = \rho_2(\theta)$
7.2(23)	7.2(17)	0.15	6.23×10^{-6}	1.68×10^{-5}
5.9(17)	5.4(9)	0.31	0.01	0.03
4.6(13)	4.5(6))	0.32	0.18	0.18
2.4(3)	2.4(2))	0.06	0.78	0.78

We obtain five pairs with similar silanol densities from Tielens and Comas-Vives models. p-values are in the $10^{-5} - 10^{-3}$ range for d and θ sample comparisons for TI_5.8 vs. CO_5.9 and TI_4.8 vs. CO_4.6. Hence, it is extremely unlikely that these samples were drawn from the same distribution. p-values are in the $10^{-2} - 5 \times 10^{-2}$ range for d and θ sample comparisons for TI_3.6 vs. CO_3.3 and TI_2.3 vs. CO_2.4. For these cases, $\mathbf{H}_0$ can be rejected at a 5% significance level. p-values are equal to 0.10 and 0.07 for ϕ sample comparisons for TI_5.8 vs. CO_5.9 and TI_4.8 vs. CO_4.6, respectively. For these cases, $\mathbf{H}_0$ can be rejected at a 10% significance level. p-values are equal to 0.02

and 0.03 for ϕ sample comparisons for TI_3.6 vs. CO_3.3 and TI_2.3 vs. CO_2.4, respectively. For these cases, $\mathbf{H}_0$ can be rejected at a 5% significance level. p-values are equal to 0.20 for all sample comparisons (ϕ, d, and θ) between TI_1.5 and CO_1.7. For these cases, we cannot reject $\mathbf{H}_0$ at any reasonable significance level.

We obtain three pairs with similar silanol densities from Tielens and Ugliengo models. The p-values are greater than 0.10 for all cases, except for d and θ comparisons for TI_2.3 vs. UG_2.4. Where, the p-values are equal to 0.07 and 0.03 for d and θ comparisons, respectively. $\mathbf{H}_0$ can be rejected at 10% and 5% significance levels, respectively. We obtain four pairs with similar silanol densities from Comas-Vives and Ugliengo models. p-values are equal to 6.23×10^{-6} and 1.68×10^{-5} for CO_7.2 vs. UG_7.2 for d and θ comparisons, respectively. Hence, it is extremely unlikely that these samples were drawn from the same distribution. The p-values are equal to 0.01 and 0.03 for CO_5.9 vs. UG_5.4 for d and θ comparisons, respectively. $\mathbf{H}_0$ can be rejected at a 5% significance level for these cases. p-values are greater than 0.10 for d and θ comparisons for CO_4.6 vs. UG_4.5 and CO_2.4 vs. UG_2.4. For these cases, we cannot reject $\mathbf{H}_0$ at any reasonable significance level. p-values are greater than 0.10 for all ϕ comparisons, except for CO_2.4 vs. UG_2.4. Where, the p-value is equal to 0.06 and $\mathbf{H}_0$ can be rejected at a 10% significance level.

It should be noted that hypothesis testing methods are not reliable for small sample sizes. The KS test is not reliable for cases where either model has less than 5 samples. In summary, Comas-Vives models are different from Tielens models on all three metrics (ϕ, d, and θ). Comas-Vives models are different from Ugliengo models based on d, and θ distributions. However, we are not able to reject $\mathbf{H}_0$ for most sample comparisons between Ugliengo and Tielens models. To further compare the models, we test $\mathbf{H}_0$ for the discrete distribution of silanol types (isolated, geminal, and vicinal) using the chi-square test. Table 5.2 shows the comparison between models of similar silanol densities

from different studies. The χ^2statistics and p-values are tabulated.

Table 5.2: Testing $\mathbf{H_0}$ between samples of silanol types (isolated, geminal, and vicinal) from different models of similar silanol densities using the chi-square test. The χ^2 statistic and p-value are shown.

Tielens silanols/nm²	**Comas-Vives silanols/nm²**	$\boldsymbol{\chi^2}$	***p*-value $\mathbf{H_0}$:$\boldsymbol{\rho_1}$(silanol types) $\boldsymbol{\rho_2}$(silanol types)**
5.8	5.9	5.97	0.05
4.8	4.6	4.89	0.09
3.6	3.3	3.50	0.17
2.3	2.4	-*	-*
1.7	1.5	-*	-*
Tielens silanols/nm²	**Ugliengo silanols/nm²**	$\boldsymbol{\chi^2}$	***p*-value $\mathbf{H_0}$:$\boldsymbol{\rho_1}$(silanol types) $\boldsymbol{\rho_2}$(silanol types)**
5.8	5.4	6.85	0.03
4.8	4.5	4.97	0.08
2.3	2.4	2.91	0.23
Comas-Vives silanols/nm²	**Ugliengo silanols/nm²**	$\boldsymbol{\chi^2}$	***p*-value $\mathbf{H_0}$:$\boldsymbol{\rho_1}$(silanol types) $\boldsymbol{\rho_2}$(silanol types)**
7.2	7.2	2.63	0.27
5.9	5.4	3.40	0.18
4.6	4.5	4.90	0.09
2.4	2.4	1.24	0.54

* χ^2 diverges if both samples have zero observations in the same category. TI_2.3, TI_1.7, CO_2.4, and CO_1.5 have zero geminal silanol groups.

p-values are equal to 0.05 and 0.09 for TI_5.8 vs. CO_5.9 and TI_4.8 vs. CO_4.6, respectively. For these cases, $\mathbf{H_0}$ can be rejected at a 10% significance level. The p-value is equal to 0.17 for TI_3.6 vs. CO_3.3 and we cannot reject $\mathbf{H_0}$ at any reasonable significance level. We cannot perform the chi-square test for TI_2.3 vs. CO_2.4 and TI_1.7 vs. CO_1.5. This is because χ^2 diverges if both samples have zero observations in the same category. TI_2.3, TI_1.7, CO_2.4, and CO_1.5 have zero geminal silanol

groups. The p-values are 0.03 and 0.08 for TI_5.8 vs. UG_5.4 and TI_4.8 vs. UG_4.5, respectively. For these cases, $\mathbf{H}_0$ can be rejected at 5% and 10% significance levels, respectively. The p-value is equal to 0.23 for TI_2.3 vs. UG_2.4 and $\mathbf{H}_0$ cannot be rejected at any reasonable significance level. The p-values are greater than 0.1 for all comparisons between Comas-Vives and Ugliengo models, except for CO_4.6 vs. UG_4.5. For this case, the p-value is equal to 0.09 and $\mathbf{H}_0$ can be rejected at a 10% significance level.

Similar to the KS test case, it should be noted that the chi-square test is not reliable for small samples. In summary, based on the distribution of silanol types, Tielens models are different from Comas-Vives and Ugliengo models. However, we are not able to reject $\mathbf{H}_0$ for most sample comparisons between Comas-Vives and Ugliengo models. Fig. 5.2 shows the comparison between the three models based on all 4 distributions (3 continuous: ϕ, d, and θ and 1 discrete: types of silanols).

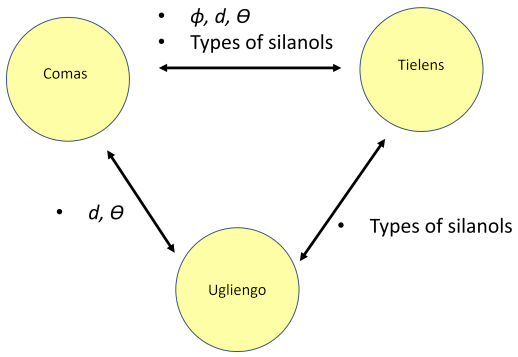


Figure 5.2: Differences between models are shown. Tielens and Comas-Vives models differ on all four metrics. Tielens and Ugliengo models differ on the distribution of silanol types. Comas-Vives and Ugliengo models have different d and θ distributions.

All the models differ on at least one metrics. Hence, different protocols to generate atomistic models of amorphous silica lead to different structural features. Grafting kinetics/thermodynamics and catalytic kinetics/mechanisms computed using different models will be different.

Another point to note are the similarities and differences between mesoporous and non-porous models. Tielen's models, representative of MCM-15, are different from non-porous Comas-Vives models on all four metrics. While they only differ from the Ugliengo models on the distribution of silanol types. The differences between mesoporous and non-porous models can be attributed to meso-scale structural differences between them or the inherent differences in their generation protocols. At present, we do not know the extent to which these reasons affect the final structure. Further systematic investigations are required to elucidate differences between mesoporous and non-porous models.

We have shown that models of silica published in the literature are different. But we cannot conclusively determine if any of them represent real silicas because of reasons outlined in the introduction. However, some information about the distribution of silanols can be obtained by grafting organometallic precursors onto amorphous silicas.[27,54] There is some evidence about the occurrence of majority of the silanols in vicinal pairs on amorphous silicas (both nonporous and mesoporous). This evidence was obtained from the EXAFS spectra of Ga modified non-porous and mesoporous amorphous silicas.[27,54] Ga was grafted onto amorphous silicas calcined at different temperatures (silicas with different silanol densities) via their reaction with vapor phase GaR_3. On all silicas, each silanol ($\equiv$SiOH) reacted with a single GaR_3 molecule via the elimination of a CH_4 molecule. A prominent Ga-Ga path was observed in the EXAFS spectra of all Ga modified silicas. The paired nature of all the grafted Ga species implies the pairwise occurrence of silanols on all silicas. If all silanols are paired, then on calcination, vicinal pairs should be the last to condense. This is because of the formation of highly strained four membered Si-O-Si-O rings on condensation of vicinal silanols. Therefore, most silanols on silicas calcined at high temperatures are part of vicinal pairs. Furthermore, the EXAFS spectra were uniform for silicas with different silanol densities. This implies that the nature of silanols is similar on all silicas. Hence, most silanols occur as vicinal pairs on all silicas.

For all simulation protocols considered here, larger models have a large fraction of vicinal sites while smaller models have a small fraction of vicinal sites. In this respect, larger models are more representative of real silicas, while smaller models are not consistent with the occurrence of majority of the silanols in vicinal pairs.

In order to validate models, several computable properties should be compared with experimental measurements including, the IR spectrum, the silanol density vs. calcination temperature curves, and the nature of silanol groups (isolated, geminal, vicinal etc.). Further validation tools are provided by the recently developed methods to model grafting and catalysis of atomically dispersed catalysts. As described in the introduction, the evolution of the active sites during grafting can be modeled using ML parametrized population balance models.[30] The predictions of the grafting model can be compared to some experimental observables. For example, the nature of the final grafted sites can be inferred from a combination of spectroscopic measurements and elemental balance experiments.[21,54] Grafting kinetics can be inferred by measuring the rate of evolution of gas phase species eliminated in grafting reactions. For example, HCl is eliminated on grafting of $TiCl_4$ and CH_4 is eliminated on grafting $GaMe_3$.[27,55] The site-averaged kinetic properties (site-averaged activation barrier and turnover frequency) of the grafted catalyst can be calculated using the importance learning algorithm and compared with experimental measurements of these quantities.[31,32] These comparisons can be made for different grafting and catalytic reactions. Validity of a model can only be established after several such successful comparisons.

5.4 Conclusions

The development of surface models of amorphous silica is critical for modeling atomically dispersed catalysts on amorphous silica supports. A few studies have developed

atomistic models of amorphous silica using different simulation protocols. And all studies have claimed that their models are representative of real silica materials. A few important questions need to be addressed. For example, which of these models, if any, are representative of real silicas? Is comparing their IR spectra and silanol density vs. calcination temperature curves with experiments enough to validate the models? Are bigger models necessarily more accurate?

In this work we have shown that models of amorphous silica generated using different simulation protocols lead to different structures. The reactivity of surface sites is a function of the local silica environment. And hence, different models will lead to different distributions of grafting/catalytic kinetics and thermodynamics. Additionally, the model sizes are small, with tens of silanols. And it is yet not clear if such models can be used to represent silica samples used in experiments which typically contain ca. 10^{18} silanols.

Future studies should systematically investigate the effect of different simulation parameters on the structure of these atomistic models. Models should be validated against several experimental metrics, some of which have been outlined here. Models which agree with the most metrics should be identified. This can potentially elucidate the structural features of real amorphous silica materials and consequently lead to the development of accurate in silico models of atomically dispersed catalysts.

Bibliography

1. Morokuma, K. & Musaev, D. G. *Computational Modeling for Homogeneous and Enzymatic Catalysis: A Knowledge-Base for Designing Efficient Catalysis* (John Wiley & Sons, 2008).

2. Sperger, T., Sanhueza, I. A., Kalvet, I. & Schoenebeck, F. Computational Studies of Synthetically Relevant Homogeneous Organometallic Catalysis Involving Ni, Pd, Ir, and Rh: An Overview of Commonly Employed DFT Methods and Mechanistic Insights. *Chem. Rev.* **115,** 9532–9586. https://www.doi.org/10.1021/acs.chemrev.5b00163 (2015).

3. Goldsmith, B. R., Hwang, T., Seritan, S., Peters, B. & Scott, S. L. Rate-Enhancing Roles of Water Molecules in Methyltrioxorhenium-Catalyzed Olefin Epoxidation by Hydrogen Peroxide. *J. Am. Chem. Soc.* **137,** 9604–9616. https://www.doi.org/10.1021/jacs.5b03750 (2015).

4. Hammer, B. & Nørskov, J. K. in, 71–129 (Academic Press, 2000). https://www.doi.org/10.1016/S0360-0564(02)45013-4.

5. Getman, R. B., Schneider, W. F., Smeltz, A. D., Delgass, W. N. & Ribeiro, F. H. Oxygen-Coverage Effects on Molecular Dissociations at a Pt Metal Surface. *Phys. Rev. Lett.* **102,** 076101. https://www.doi.org/10.1103/PhysRevLett.102.076101 (2009).

6. Greeley, J. Theoretical Heterogeneous Catalysis: Scaling Relationships and Computational Catalyst Design. *Annu Rev Chem Biomol* **7,** 605–635. https://www.doi.org/10.1146/annurev-chembioeng-080615-034413 (2016).

7. Saleheen, M. & Heyden, A. Liquid-Phase Modeling in Heterogeneous Catalysis. *ACS Catal.* **8,** 2188–2194. https://www.doi.org/10.1021/acscatal.7b04367 (2018).

8. Salciccioli, M., Stamatakis, M., Caratzoulas, S. & Vlachos, D. G. A review of multiscale modeling of metal-catalyzed reactions: Mechanism development for complexity and emergent behavior. *Chem. Eng. Sci.* **66,** 4319–4355. https://www.doi.org/10.1016/j.ces.2011.05.050 (2011).

9. Blaszkowski, S. R. & van Santen, R. A. Theoretical study of the mechanism of surface methoxy and dimethyl ether formation from methanol catalyzed by zeolitic protons. *The Journal of Physical Chemistry B* **101,** 2292–2305. https://doi.org/10.1021/jp962006 (1997).

10. Li, Y.-P., Head-Gordon, M. & Bell, A. T. Analysis of the Reaction Mechanism and Catalytic Activity of Metal-Substituted Beta Zeolite for the Isomerization of Glucose to Fructose. *ACS Catalysis* **4,** 1537–1545. https://www.doi.org/10.1021/cs401054f (2014).

11. Newsome, D. & Coppens, M.-O. Molecular dynamics as a tool to study heterogeneity in zeolites - Effect of Na+ cations on diffusion of CO_2 and N_2 in Na-ZSM-5. *Chemical Engineering Science* **121,** 300–312. https://www.doi.org/10.1016/j.ces.2014.09.024 (2015).

12. McFarland, E. W. & Metiu, H. Catalysis by Doped Oxides. *Chemical Reviews* **113,** 4391–4427. https://www.doi.org/10.1021/cr300418s (2013).

13. Le Bahers, T., Rérat, M. & Sautet, P. Semiconductors Used in Photovoltaic and Photocatalytic Devices: Assessing Fundamental Properties from DFT. *The Journal of Physical Chemistry C* **118,** 5997–6008. https://www.doi.org/10.1021/jp409724c (2014).

14. Li, F., Luo, S., Sun, Z., Bao, X. & Fan, L.-S. Role of metal oxide support in redox reactions of iron oxide for chemical looping applications: experiments and density functional theory calculations. *Energy & Environmental Science* **4,** 3661–3667. https://www.doi.org/10.1039/C1EE01325D (2011).

15. Goldsmith, B. R., Peters, B., Johnson, J. K., Gates, B. C. & Scott, S. L. Beyond Ordered Materials: Understanding Catalytic Sites on Amorphous Solids. *ACS Catal.* **7,** 7543–7757. https://www.doi.org/10.1021/acscatal.7b01767 (2017).

16. Yoon, C. & Cocke, D. L. Potential of amorphous materials as catalysts. *Journal of Non-Crystalline Solids* **79,** 217–245. ISSN: 0022-3093. https://doi.org/10.1016/0022-3093(86)90224-3 (1986).

17. McDaniel, M. P. A Review of the Phillips Supported Chromium Catalyst and Its Commercial Use for Ethylene Polymerization. English. *Adv. Catal.* **53,** 123–606. https://www.doi.org/https://doi.org/10.1016/S0360-0564(10)53003-7 (2010).

18. Begley, J. & Wilson, R. The kinetics of propylene disproportionation. *Journal of Catalysis* **9,** 375–395. ISSN: 0021-9517. https://doi.org/10.1016/0021-9517(67)90265-5 (1967).

19. Regalbuto, J. *Catalyst preparation: science and engineering* (CRC press, 2016).

20. Zhong, L. *et al.* Spectroscopic and structural characterization of Cr(II)/SiO_2 active site precursors in model Phillips polymerization catalysts. *J. Catal.* **293,** 1–12. https://www.doi.org/10.1016/j.jcat.2012.05.014 (2012).

21. Coperet, C. *et al.* Surface Organometallic and Coordination Chemistry toward Single-Site Heterogeneous Catalysts: Strategies, Methods, Structures, and Activities. *Chem. Rev.* **116,** 323–421. https://www.doi.org/10.1021/acs.chemrev.5b00373 (2016).

22. Vansant, E. E., Van Der Voort, P. & Vrancken, K. C. *Characterization and Chemical Modification of the Silica Surface* (1995).

23. Flörke, O. W. *et al.* in *Ullmann's Encyclopedia of Industrial Chemistry* (John Wiley & Sons, 2000).

24. Bergna, H. E. & Roberts, W. O. *Colloidal silica: fundamentals and applications* (CRC Press, 2005).

25. Zhuravlev, L. T. The surface chemistry of amorphous silica. Zhuravlev model. *Colloids Surf., A* **173,** 1–38. https://www.doi.org/10.1016/S0927-7757(00)00556-2 (2000).

26. Deguns, E. W., Taha, Z., Meitzner, G. D. & Scott, S. L. An X-ray Absorption Study of Two VOCl3-Modified Silicas: Evidence for Chloride-Silica Interactions. *The Journal of Physical Chemistry B* **109,** 5005–5011. https://www.doi.org/10.1021/jp045886y (2005).

27. Fleischman, S. D. & Scott, S. L. Evidence for the Pairwise Disposition of Grafting Sites on Highly Dehydroxylated Silicas via Their Reactions with $Ga(CH_3)_3$. *J. Am. Chem. Soc.* **133,** 4847–4855. https://www.doi.org/10.1021/ja108905p (2011).

28. Samantaray, M. K. *et al.* Surface organometallic chemistry in heterogeneous catalysis. *Chem. Soc. Rev.* **47,** 8403–8437. https://www.doi.org/10.1039/C8CS00356D (2018).

29. Mol, J. Industrial applications of olefin metathesis. *Journal of Molecular Catalysis A: Chemical* **213.** The 15th. International Symposium on Olefin Metathesis and Related Chemistry, 39–45. ISSN: 1381-1169. https://doi.org/10.1016/j.molcata.2003.10.049 (2004).

30. Khan, S. A., Vandervelden, C. A., Scott, S. L. & Peters, B. Grafting metal complexes onto amorphous supports: From elementary steps to catalyst site populations: Via kernel regression. *React. Chem. Eng.* **5,** 66–76. https://www.doi.org/10.1039/c9re00357f (2020).

31. Vandervelden, C. A., Khan, S. A., Scott, S. L. & Peters, B. Site-averaged kinetics for catalysts on amorphous supports: an importance learning algorithm. *React. Chem. Eng.* **5,** 77–86. https://www.doi.org/10.1039/C9RE00356H (2020).

32. Vandervelden, C. A., Khan, S. A. & Peters, B. Importance learning estimator for the site-averaged turnover frequency of a disordered solid catalyst. *The Journal of Chemical Physics* **153,** 244120. https://doi.org/10.1063/5.0037450 (2020).

33. Comas-Vives, A. Amorphous SiO_2 surface models: Energetics of the dehydroxylation process, strain, ab initio atomistic thermodynamics and IR spectroscopic signatures. *Phys. Chem. Chem. Phys.* **18,** 7475–7482. https://www.doi.org/10.1039/c6cp00602g (2016).

34. Ewing, C. S., Bhavsar, S., Veser, G., McCarthy, J. J. & Johnson, J. K. Accurate amorphous silica surface models from first-principles thermodynamics of surface dehydroxylation. *Langmuir* **30,** 5133–5141. https://doi.org/10.1021/la500422p (2014).

35. Ugliengo, P. *et al.* Realistic models of hydroxylated amorphous silica surfaces and MCM- 41 mesoporous material simulated by large-scale periodic B3LYP calculations. *Adv. Mater.* **20,** 4579–4583. https://www.doi.org/10.1002/adma.200801489 (2008).

36. Tielens, F., Gervais, C., Lambert, J. F., Mauri, F. & Costa, D. Ab initio study of the hydroxylated surface of amorphous silica: A representative model. *Chem. Mater.* **20,** 3336–3344. https://www.doi.org/10.1021/cm8001173 (2008).

37. Gierada, M., Michorczyk, P., Tielens, F. & Handzlik, J. Reduction of chromia-silica catalysts: A molecular picture. *J. Catal.* **340,** 122–135. https://www.doi.org/10.1016/j.jcat.2016.04.022 (2016).

38. Mathias, J. & Wannemacher, G. Basic characteristics and applications of aerosil: 30. The chemistry and physics of the aerosil Surface. *Journal of Colloid and Interface Science* **125,** 61–68. ISSN: 0021-9797. https://doi.org/10.1016/0021-9797(88)90054-9 (1988).

39. Garrett, P. R. *Defoaming: Theory and industrial applications* 329. https://doi.org/10.1201/9781315140827 (CRC Press, Boca Raton, 1992).

40. Zhao, D. *et al.* Triblock Copolymer Syntheses of Mesoporous Silica with Periodic 50 to 300 Angstrom Pores. *Science* **279,** 548–552. ISSN: 0036-8075. https://doi.org/10.1126/science.279.5350.548 (1998).

41. Beck, J. S. *et al.* A new family of mesoporous molecular sieves prepared with liquid crystal templates. *Journal of the American Chemical Society* **114,** 10834–10843. https://doi.org/10.1021/ja00053a020 (1992).

42. Kresge, a. C., Leonowicz, M., Roth, W. J., Vartuli, J. & Beck, J. Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism. *nature* **359,** 710–712. https://doi.org/10.1038/359710a0 (1992).

43. Stachurski, Z. H. On Structure and Properties of Amorphous Materials. *Materials* **4,** 1564–1598. ISSN: 1996-1944. https://doi.org/10.3390/ma4091564 (2011).

44. Bürgler, D. E. *et al.* Atomic-scale scanning tunneling microscopy of amorphous surfaces. *Phys. Rev. B* **59,** 10895–10902. https://doi.org/10.1103/PhysRevB.59.10895 (16 1999).

45. Peters, B. & Scott, S. L. Single Atom Catalysts on Amorphous Supports: A Quenched Disorder Perspective. *J. Chem. Phys.* **142,** 104708. https://www.doi.org/10.1063/1.4914145 (2015).

46. Goldsmith, B. R., Sanderson, E. D., Bean, D. & Peters, B. Isolated Catalyst Sites on Amorphous Supports: A Systematic Algorithm for Understanding Heterogeneities in Structure and Reactivity. *J. Chem. Phys.* **138,** 204105. https://www.doi.org/10.1063/1.4807384 (2013).

47. Floryan, L., Borosy, A. P., Nunez-Zarur, F., Comas-Vives, A. & Coperet, C. Strain Effect and Dual Initiation Pathway in Cr(III)/SiO_2 Polymerization Catalysts from Amorphous Periodic Models. English. *J. Catal.* **346,** 50–56. https://www.doi.org/10.1016/j.jcat.2016.11.037 (2017).

48. Guesmi, H. & Tielens, F. Chromium oxide species supported on silica: A representative periodic DFT model. *J Phys Chem C* **116,** 994–1001. https://doi.org/10.1021/jp209680r (2012).

49. Handzlik, J. Application of the ONIOM (QM/QM) method in the study of molybdena-silica system active in olefin metathesis. *Int. J. Quantum Chem.* **107,** 2111–2119. https://www.doi.org/10.1002/qua.21397 (2007).

50. Goldsmith, B. R., Fong, A. & Peters, B. in *Reaction Rate Constant Computations: Theories and Applications* 213–232 (The Royal Society of Chemistry, 2013). ISBN: 978-1-84973-650-3. http://dx.doi.org/10.1039/9781849737753-00213.

51. Jr., F. J. M. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association* **46,** 68–78. 10.1080/01621459.1951.10500769 (1951).

52. Neidinger, R. D. Introduction to Automatic Differentiation and MATLAB Object-Oriented Programming. *SIAM Review* **52,** 545–563. ISSN: 0036-1445. https://doi.org/10.1137/080743627 (2010).

53. Greenwood, P. E. & Nikulin, M. S. *A guide to chi-squared testing* (John Wiley & Sons, 1996).

54. Taha, Z. A., Deguns, E. W., Chattopadhyay, S. & Scott, S. L. Formation of Di-gallium Sites in the Reaction of Trimethylgallium with Silica. *Organometallics* **25,** 1891–1899. ISSN: 0276-7333. https://doi.org/10.1021/om051034o (2006).

55. Haukka, S., Lakomaa, E. L., Jylha, O., Vilhunen, J. & Hornytzkyj, S. Dispersion and distribution of titanium species bound to silica from titanium tetrachloride. *Langmuir* **9,** 3497–3506. https://doi.org/10.1021/la00036a026 (1993).

Chapter 6

Quantifying abundances of grafted $[(\equiv SiO)_{4-x}TiCl_x]$ sites on amorphous silica

The experiments described in this chapter were performed by students of the Scott group at UCSB: Erica Deguns, Rosemary White, and Ziyad Taha.

6.1 Introduction

Silica-supported titanium catalysts are used for a wide variety of reactions, including olefin epoxidation,[1,2] esterification[3] and transesterification,[4] hydroxylation of phenol,[5] oxyfunctionalization of alkanes,[6] and Baeyer-Villiger oxidation.[7] Titanium complexes can achieve a variety of coordination numbers and bonding with a variety of ligands including alkoxide, siloxide, or halides.[8] One of many important examples is the Ziegler-Natta catalyst (Ti/SiO_2) for the polymerization of olefins.[9] While framework titanium sites may be desired for epoxidation catalysts,[10,11] polymerization activity requires titanium

sites with more accessible coordination spheres.[9] $TiCl_4$ has been widely employed as a precursor for such catalysts due to its availability and its ability to be readily dispersed onto the surface by, for example, atomic layer epitaxy,[12] or vapor phase grafting.[13–15] Preparation of materials by these methods leads to more active materials by preventing the formation of inactive TiO_2 domains.[15,16] Surface silanols react readily with $TiCl_4$ to form chemisorbed Ti species, liberating HCl. Two reactions of $TiCl_4$ with silanols on silica have been proposed, eq. 6.1-6.2,.[9,14,17,18] A third proposed reaction invokes the reaction between $TiCl_4$ and strained siloxane bonds, eq. 6.3.[18]

$$\equiv \mathrm{SiOH} + \mathrm{TiCl_4} \rightleftarrows \equiv \mathrm{SiOTiCl_3} + \mathrm{HCl}. \tag{6.1}$$

$$\equiv \mathrm{SiOH} + \equiv \mathrm{SiOTiCl_3} \rightleftarrows (\equiv \mathrm{SiO})_2\mathrm{TiCl_2} + \mathrm{HCl}. \tag{6.2}$$

$$\equiv \mathrm{SiOSi} \equiv + \mathrm{TiCl_4} \rightleftarrows \equiv \mathrm{SiOTiCl_3} + \equiv \mathrm{SiCl}. \tag{6.3}$$

Many literature reports suggest that eq. 6.1-6.3 occur simultaneously, resulting in unavoidable mixtures of titanium sites.[13,19,20] Although these grafting reactions have been extensively studied by IR,[18,19] Raman, NMR,[19] XAS,[21] XPS,[15] DFT,[22] SEM/TEM[15] and mass balance techniques, the literature is inconclusive as to the structure of the titanium species generated. Furthermore, most grafting studies only attempt to qualitatively determine the nature of grafted species. To the best of our knowledge, no study has attempted to quantify the abundances of different grafted species. In this work, we characterize the structure and stability of the grafted sites created by the vapor phase reaction of TiCtextsubscript4 with a silica that is partially dehydroxylated at two different temperatures: 100 and 500 °C. In addition to reactions described by eq. 6.1-6.3, we

also consider the following route to the bipodal [(≡SiO)$_2$TiCl$_2$] site:

$$2 \equiv \mathrm{SiOTiCl_3} \rightleftarrows (\equiv \mathrm{SiO})_2\mathrm{TiCl_2} + \mathrm{TiCl_4}. \tag{6.4}$$

We study these reactions and the resulting ensemble of grafted sites using IR and mass balance experiments. We also introduce a site balance algebra to quantify the amount of different grafted species. Our results demonstrate how such a model can help to interpret experimental data, check for internal consistency, and help to place confidence intervals on populations of different site types.

6.2 Experimental methods

6.2.1 Sample preparation

The silica used in this study is Aerosil-380 (hereafter referred to as A380), an amorphous, fumed silica from Degussa. It has a surface area of 340 m^2/g, a primary particle size of 7 nm and no significant microporosity. The thermal pretreatment of the silica in each experiment is indicated by an appended number. For example, A380-100 denotes a sample of Aerosil-380 treated at 100°C. To ensure reproducibility, each sample was heated under dynamic vacuum ($<10^{-4}$ Torr) at the designated temperature for a minimum of four hours. $TiCl_4$ (99.99+%, Aldrich) was stored under vacuum in a glass bulb equipped with a Teflon stopcock. Excess $TiCl_4$ was transferred onto powdered silica as the vapor, via an all-glass high vacuum line equipped with ground-glass stopcocks. $TiCl_4$ reacts slowly with hydrocarbon-based vacuum greases. Although this does not appear to affect the preparation of titanium-modified silicas, the grease was replaced frequently to maintain the integrity of the vacuum system. After desorption of unreacted $TiCl_4$ to a liquid N_2 trap, the titanium-modified silica was recovered as a white powder. Titanium

analysis was performed in air at the end of each experiment. Each sample was weighed and stirred in 1.0 M H_2SO_4 to which 3.5% aqueous H_2O_2 (0.03 mL/mL sample solution) was added to extract the titanium as its peroxo complex. The solution was filtered before recording its UV-vis spectrum in a 1 cm quartz cuvette, referenced to a H_2SO_4/H_2O_2 solution containing approximately the same concentration of H_2O_2. Spectra were recorded on a Shimadzu UV2401PC spectrophotometer. The yellow peroxotitanium complex has a distinct peak in the visible at $\lambda_{max} = 408$ nm. A calibration curve was prepared using a standard solution of titanium atomic absorption standard (1004 ppm, Aldrich).

6.2.2 Infrared spectroscopy

IR experiments were performed in a Pyrex gas cell equipped with KCl windows affixed with TorrSeal® (Varian). Its high-vacuum ground-glass stopcock and joints were lubricated with Apiezon® H grease (Varian). A self-supporting pellet of silica was prepared in air by pressing ca. 15 mg of the solid at 40 kg/cm^2 in a 16 mm stainless steel die. It was then mounted in a Pyrex pellet holder. $TiCl_2$ vapor was introduced via a vacuum manifold directly onto the silica pellet, and the excess was desorbed under dynamic vacuum to a liquid N_2 trap. IR spectra of the self-supporting pellet were recorded in transmission mode on a Shimadzu PrestigeIR spectrophotometer equipped with a DTGS detector, and purged with CO_2-free dry air from a Balston 75-52 Purge Gas Generator. Background and sample IR spectra were recorded by co-adding 64 scans at a resolution of 4 cm^{-1}.

6.3 Results

6.3.1 Experimental Observation of $TiCl_4$ Grafting onto Silica

Modification of the surface of silica with $TiCl_4$ vapor results in a slight change in the appearance of the solid, from dull to brilliant white. The extent of reaction between $TiCl_4$ and the surface silanols is observed readily using IR spectroscopy. Fig. 6.1a-b show the IR spectrum of a self-supporting disk of a nonporous silica (Aerosil 380) pretreated at 500 °C (designated A380-500), before and after its exposure to excess $TiCl_4$ vapor at room temperature for several minutes. The sharp band at 3747 cm^{-1}, attributed to the ≡(SiO-H) mode of the non-hydrogen-bonded surface silanols, disappears rapidly and almost completely upon contact with excess $TiCl_4$. Simultaneously, the characteristic rovibrational spectrum of HCl(g), centered at 2884^{-1}, was detected (not shown). Virtually all of the hydroxyl groups of the silica surface are therefore consumed in the reaction with $TiCl_4$(g) at room temperature. Only a weak, broad absorption band remains at 3665 cm^{-1} in the ≡(SiO-H) region. This band, assigned to internal silanols perturbed by the silica matrix,[23] represents hydroxyl groups that are unreactive towards $TiCl_4$[17] as well as even the most aggressive inorganic hydrogen-sequestering agents, including $AlMe_3$[17] and $VOCl_3$.[24] They are therefore considered to be inaccessible silanols.

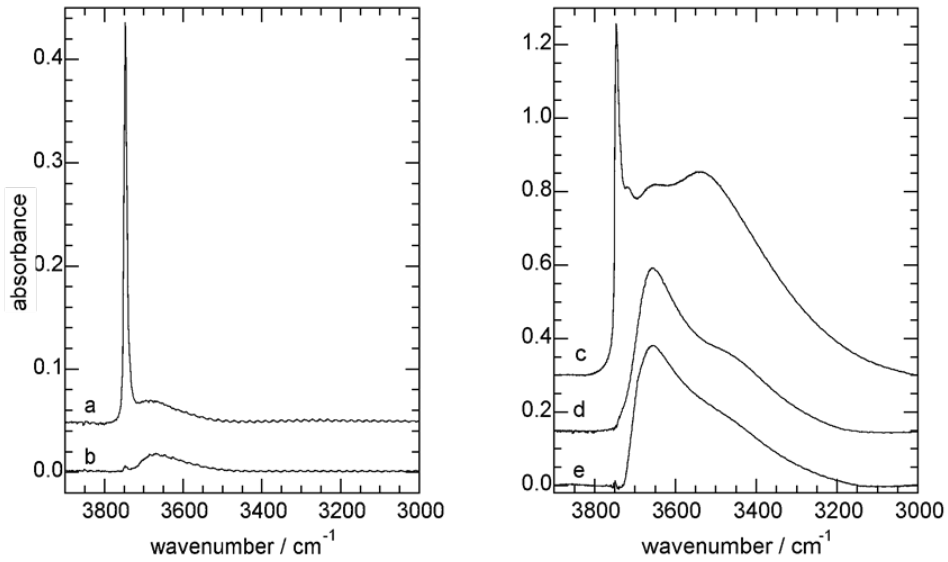


Figure 6.1: IR spectra of self-supporting pellets of A380 silica: (a) pretreated at 500 °C; (b) previous sample, after reaction with excess $TiCl_4$ followed by evacuation of volatiles; (c) pretreated at 100 °C; (d) previous sample, after reaction with excess $TiCl_4$ followed by evacuation of volatiles; and (e) previous sample, after heating at 120 °C for 4 hours under dynamic vacuum.

The nature of the grafted $TiCl_x$ sites can be inferred, in part, from the stoichiometry of the surface reaction: the amount of Ti irreversibly adsorbed, the amount of HCl liberated, and the Ti/Cl ratio present in the modified silica.[25] The titanium content of $TiCl_4$-modified A380-500, measured after desorption of physisorbed $TiCl_4$ by evacuation of the reactor to a liquid N_2 trap, was (0.78 ± 0.05) mmol Ti/g silica (average of 5 independent experiments). Since the number of accessible hydroxyls on this silica was previously measured to be (0.81 ± 0.03) mmol/g silica,[26] the Ti analysis suggests that each accessible silanol reacts with one $TiCl_4$ molecule. The stoichiometry of the grafting reaction on A380-500 is therefore described completely by eq. 6.1. Furthermore, the titanium loading of $TiCl_4$-modified A380-500 remained stable at 0.78 mmol Ti/g silica even after 3 days of continuous evacuation at room temperature, or after heating the material to 120 °C under dynamic vacuum for 4 h. The ratio of grafted Ti to accessible

silanols is almost 1:1. One possible interpretation is that all silanol sites are isolated. IR and mass balance do not determine the absolute populations of isolated and paired silanol sites. However, experiments grafting $GaMe_3$ to amorphous SiO_2 samples of different silanol densities suggest that majority of the silanols are present as vicinal pairs (silanols on adjacent Si atoms connected by a siloxane bridge).[27,28] mass balance suggests that $GaMe_3$ reacts with surface silanols (on all silicas) in a 1:1 ratio via the following reaction:

$$2 \equiv SiOHGaMe_3 \rightleftarrows \equiv SiOGaMe_2 + MeH. \tag{6.5}$$

A prominent Ga-Ga path was observed in the EXAFS spectra of Ga/SiO_2, implying the existence of dimeric Ga sites. This further implies that silanols occur in pairs. Furthermore, the Ga-Ga ($\sim$ 2.99 Å) and Ga-O ($\sim$ 1.94 Å) distances were invariant with the silanol density and similar to the Ga-Ga and Ga-O distances in molecular Ga_2O_2, respectively. Since the structure of grafted Ga dimers is similar across silanol densities, the silanol types on all silicas should also be similar. And, since vicinal sites are the last to condense, most silanols on low density silicas are vicinal. Therefore, majority of the silanols on all silicas should occur in vicinal pairs.

The predicted reaction of vicinal silanols with $TiCl_4$ to produce predominantly [≡SiO$TiCl_3$] is also consistent with the experimental observations provided the vicinal Ti sites do not associate to give a dinuclear structure (unlike grafted $GaMe_x$).

The IR spectrum of silica pretreated at a lower temperature (A380-100) is more complex than that of A380-500 in the region of the O-H stretching vibrations, fig. 6.1c. More of the silanols are perturbed by hydrogen-bonding, although there is no molecular water present (as judged by the absence of the HOH bending mode, at 1640 cm^{-1}). The fraction of inaccessible silanols is also larger on A380-100 compared to A380-500.28 This is evident upon comparing their IR spectra after reaction with excess $TiCl_2$, Figs.

6.1b and 6.1d. However, the signal at 3747 cm^{-1} due to the non-interacting silanols of A380-100 disappears completely upon exposure to excess $TiCl_4$. Initially, the reaction of A380-100 with excess $TiCl_2$ resulted in a titanium loading of (1.25 ± 0.02) mmol/g silica (average of 2 experiments). Since A380-100 contains (1.15 ± 0.01) mmol accessible ≡SiOH/g, these values correspond to an almost 1:1 reaction between the silanol groups and $TiCl_2$. However, after four hours of thermal treatment under dynamic vacuum, the Ti loading slightly decreased to (1.11 ± 0.03) mmol /g silica (average of 3 experiments). This $\sim$ 10 % reduction in the loading could be attributed to the conversion of $\sim$ 15 % adjacent monopodal [≡SiOTiCl_3] sites to bipodal [$(\equiv SiO)_2TiCl_2$] sites via eq. 6.4.

6.3.2 Site balance algebra

In this section we develop a site balance algebra to describe the calcination, grafting, and evacuation processes on silica. We do not obtain information about exact structural features of sites using this method. However, site balance algebra can become an additional tool to interpret experimental data. And in combination with mass balance and spectroscopic methods it can be used to characterize the nature of sites more precisely on atomically dispersed catalysts. The model is used to classify sites into discrete categories like monopodal [≡SiOTiCl_2] and bipodal [$(\equiv SiO)_2TiCl_2$]. We demonstrate how measurements of total Ti loading after grafting and post evacuation can be used to quantify the amounts of monpodal [≡SiOTiCl_3] and bipodal [$(\equiv SiO)_2TiCl_2$] sites. We model the surface of A380-100 as a distribution of isolated and paired silanol sites. Upon heating at 100 °C under vacuum, the surface of silica loses physisorbed water. Further heating at 500 °C causes some adjacent silanols to condense, with elimination of water according to the following reaction:

$$2 \equiv \mathrm{SiOH} \rightleftarrows \equiv \mathrm{SiOSi} \equiv + \mathrm{H_2O}. \tag{6.6}$$

To model the effect of thermal treatment, the silanols are further classified into silanol pairs that condense upon calcination at 500 °C and those which do not. According to IR and mass balance, almost all silanols react with $TiCl_4$ to form monopodal [≡SiOTiCl$_3$] species regardless of the thermal pretreatment of the silica (Section 4.1). However, the pre-evacuation Ti:silanol ratio is slightly different than 1:1 on both the samples. A slightly larger amount of Ti (1.24 mmol/g) vs. accessible silanols (1.15 mmol/g) was grafted on A380-100. And a slightly smaller amount of Ti (0.78 mmol/g) vs. accessible silanols (0.81 mmol/g) was grafted on A380-500. We assume that the amount of grafted Ti is the average of the two values on both silicas. And that the Ti:silanol ratio is 1:1 after grafting on both silicas. Accordingly, the amount of Ti is assumed to be $(1.24 + 1.15)/2 = 1.20$ mmol/g on A380-100 and $(0.78 + 0.81)/2 = 0.80$ mmol/g on A380-500.

On A380-100, prolonged evacuation causes a few of the vicinal [≡SiOTiCl$_3$] sites to be converted to bipodal [(≡SiO)$_2$TiCl$_2$] by elimination of $TiCl_4$. Hence, we further subdivide the vicinal sites into those which eliminate $TiCl_4$ upon evacuation, and those which do not. The various grafting and condensation reactions are summarized in fig. 6.2.

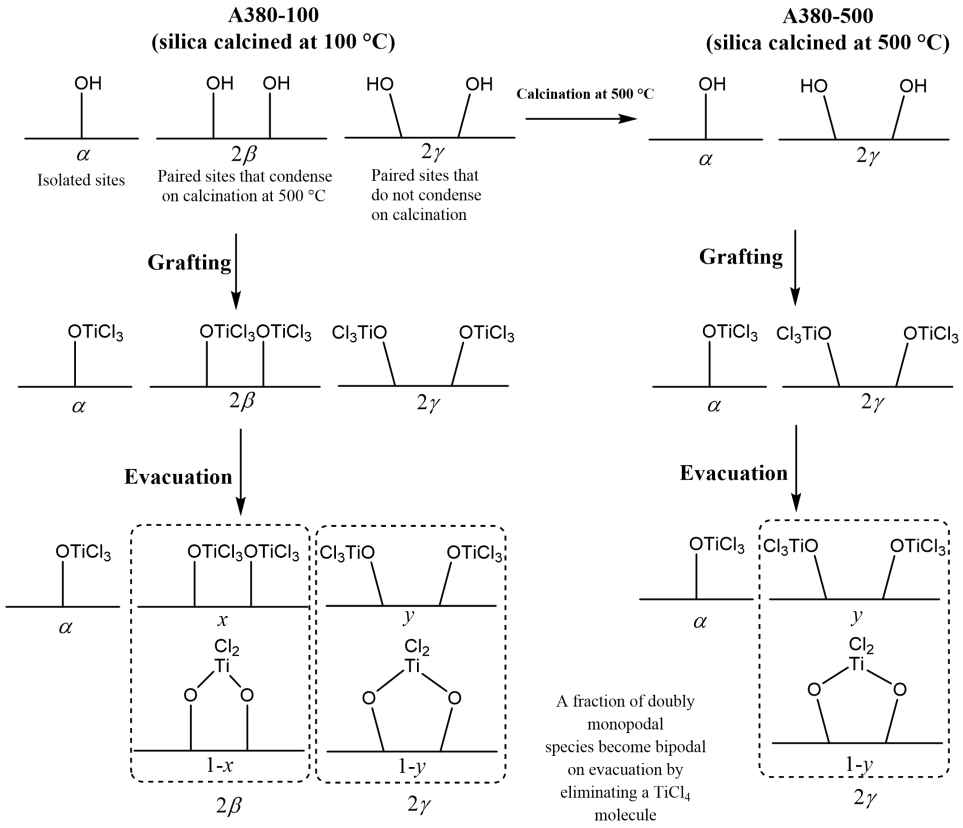


Figure 6.2: Effect of thermal treatment on the populations of various silanol sites and their reaction with $TiCl_4$, followed by prolonged evacuation. The surface silanol populations (α, β, γ) are in mmol/g; x and y are fractions.

Here, α is the population (mmol /g silica) of isolated silanols, 2β is the population (mmol /g silica) of paired silanols which do not condense on thermal treatment at 500 °C, and 2γ is the population (mmol /g silica) of paired silanols that condense on thermal treatment. We define x as the ratio of β sites and y as the ratio of γ sites, respectively, which do not convert to bipodal $[(\equiv SiO)_2TiCl_2]$ on evacuation. The pre- and post-evacuation Ti loadings can be expressed in terms of α, β, γ, x, and y. And we can setup a system of equations by equating these expressions to their experimentally measured values from section 4.1.

The equations are as follows:

A380-100 pre-evacuation

$$\alpha + 2\beta + 2\gamma = n_1 \pm \sigma_1 = 1.20 \pm 0.02 \tag{6.7}$$

A380-100 post-evacuation

$$\alpha + 2x\beta + (1-x)\beta + 2y\gamma + (1-y)\gamma = n_3 \pm \sigma_3 = 1.11 \pm 0.03 \tag{6.8}$$

A380-500 pre-evacuation

$$\alpha + 2\gamma = n_2 \pm \sigma_2 = 0.80 \pm 0.05 \tag{6.9}$$

A380-500 post-evacuation

$$\alpha + 2y\gamma + (1-y)\gamma = n_4 \pm \sigma_1 = 0.80 \pm 0.05 \tag{6.10}$$

Here, n_1 and n_2 are the pre-evacuation Ti loadings on A380-100 and A380-500, respectively. n_3 and n_4 are the post-evacuation Ti loadings on A380-100 and A380-500, respectively. And σ_i are uncertainties in the experimentally measured values of n_i. The system of equations is underdetermined with 4 eqns. and 5 variables (α, β, γ, x, and y). As described in section 3.1, by grafting $GaMe_3$ to amorphous silicas it was established that amorphous silicas have a negligible population of isolated silanols.[27,28] Given this evidence we assume $\alpha = 0$. Setting $\alpha = 0$ gives us a solvable system of equations. And we get

$$\hat{\beta} = \frac{n_1 - n_2}{2}, \tag{6.11}$$

$$\hat{\gamma} = \frac{n_2}{2}, \tag{6.12}$$

$$\hat{x} = 2\frac{n_3 - n_4}{n_2 - n_1} - 1, \tag{6.13}$$

$$\hat{x} = \frac{2n_4}{n_2} - 1. \tag{6.14}$$

Here, $\hat{\beta}$, $\hat{\gamma}$, $\hat{x}$, and $\hat{y}$ are the mean values of β, γ, x, and y respectively. We can also calculate errors in the values by expanding the solutions to first order using Taylor's expression. For example, for β we have

$$d\beta = \Sigma \frac{\partial \beta}{\partial n_i} dn_i. \tag{6.15}$$

Here, $dn_i = \sigma_i$ are the standard deviations (uncertainty) in n_i. We can express β in terms of $\hat{\beta}$ and the uncertainty as

$$\beta = \hat{\beta} \pm \sqrt{\Sigma (\frac{\partial \beta}{\partial n_i})^2 {\sigma_i}^2}. \tag{6.16}$$

Plugging the expression of β from eqn. (11) into eqn. (16) we get

$$\beta = \hat{\beta} \pm \frac{\sqrt{{\sigma_1}^2 + {\sigma_2}^2}}{2}. \tag{6.17}$$

We can derive similar expressions for other variables:

$$\gamma = \hat{\gamma} \pm \frac{\sigma_2}{2} 1 \tag{6.18}$$

$$x = \hat{x} \pm 2\frac{\sqrt{({\sigma_1}^2 + {\sigma_2}^2)(n_3 - n_4)^2 + ({\sigma_3}^2 + {\sigma_4}^2)(n_1 - n_2)^2}}{(n_2 - n_1)^2} \tag{6.19}$$

$$y = \hat{y} \pm \frac{2}{{n_2}^2}\sqrt{{\sigma_2}^2{n_4}^2 + {\sigma_4}^2{n_2}^2}. \tag{6.20}$$

Table 6.1 shows the calculated values of the variables with uncertainties.

Table 6.1: Calculated values and uncertainties of β, γ, x, and y

Variable	**Value**
β	0.20 ± 0.03 mmol/g
γ	0.40 ± 0.02 mmol/g
x	$43 \pm 36\%$
y	$0 \pm 18\%$

We can quantify the abundances of different species in fig. 6.2 using the site balance algebra. 0.40 ± 0.06 mmol/g silanols condense on calcination at 500 °C. Approximately 43±36% (7-79%) of the sites which condense on calcination at 500 °C form bipodal species on evacuation. And $0 \pm 18\%$ (0-18%) of the sites which condense on calcination at 500 °C form bipodal species on evacuation. This implies that sites which are harder to condense on calcination are also less likely to convert to bipodal [(≡SiO)$_2$TiCl$_2$] on evacuation. β (±0.03) and γ (±0.02) have small uncertainties. But the uncertainties in y (±18%), and x (±36%) are large. Uncertainties can be reduced using more precise experimental measurements. We can also check the consistency of experiments using the site balance algebra. The initial set of grafting experiments incorrectly measured the post-evacuation Ti loading on A380-100 to be ca. 0.4 mmol/g silica ($n_2 = 0.4$). This would have implied that almost all monopodal [≡SiOTiCl$_3$] convert to bipodal [(≡SiO)$_2$TiCl$_2$] on evacuation on A380-100. Solving the site balance algebra with $n_2 = 0.4$ gives $x = 2.8$. However, $x = 2.8$ is inconsistent with $0 < x < 1$ (x is a fraction). This led us to repeat the grafting experiments multiple times to achieve a reproduceable post-evacuation Ti loading on

A380-100 ($n_4 = 1.11$ mmol/g).

6.4 Conclusions

Metals dispersed on amorphous supports are used to catalyze several industrially important reactions. Active sites reside in different local environments because of the amorphous nature of the support. And the lack of long-range order in these amorphous materials makes it difficult to characterize the structure and distribution of sites on these catalysts. Furthermore, the structure of grafted sites dictates the activity of these catalysts. And hence, determining the structural distribution of sites is critical in understanding the mechanism and activity of such catalysts. In this work, we identify the structure and distribution of sites on Ti modified silicas using a combination of IR and mass balance. We also introduce a site balance algebra to quantify the populations of grafted Ti species. $TiCl_4$ reacts with amorphous silica to yield monopodally grafted [$\equiv SiOTiCl_3$] species. The nature of grafted species immediately after grafting is invariant with silica calcination temperature (100 °C and 500 °C). Furthermore, 15% monopodal [$\equiv SiOTiCl_3$] convert to bipodal [$(\equiv SiO)_2TiCl_2$] via elimination of $TiCl_4$ under vacuum on A380-100 (silica calcined at 100 °C). However, A380-500 (silica calcined at 500 °C) is stable under vacuum. We introduce a site balance algebra to determine the abundances of monopodal and bipodal sites. The model was also used to identify an incorrectly measured post-evacuation Ti loading on A380-100. And this guided us to reperform the grafting to obtain a reproducible result consistent with the solution of the site balance algebra. We believe that such a site balance algebra can be useful in interpreting the results of and guiding grafting experiments. A complete characterization of atomically dispersed catalysts will require a combination of spectroscopic methods, mass balance, site balance algebra, and recently developed computational population balance modeling

tools.[29] In future work, we will demonstrate a synergistic application of these methods to determine the structure and distribution of sites on atomically dispersed amorphous catalysts like Cr/SiO_2, V/SiO_2, Ga/SiO_2 etc.

Bibliography

1. Joergensen, K. A. Transition-metal-catalyzed epoxidations. *Chemical Reviews* **89,** 431–458. https://doi.org/10.1021/cr00093a001 (1989).

2. Cativiela, C., Fraile, J., García, J. & Mayoral, J. A new titanium-silica catalyst for the epoxidation of alkenes. *Journal of Molecular Catalysis A: Chemical* **112,** 259–267. ISSN: 1381-1169. https://doi.org/10.1016/1381-1169(96)00131-8 (1996).

3. Schwegler, M. A., Van Bekkum, H. & De Munck, N. Heteropolyacids as catalysts for the production of phthalate diesters. *Applied catalysis* **74,** 191–204. http://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=5527417 (1991).

4. Blandy, C., Pellegatta, J.-L. & Cassoux, P. Homogeneous and supported titanates as catalysts for transesterification of acrylic esters. *Catalysis letters* **43,** 139–142. https://doi.org/10.1023/A:1018978203665 (1997).

5. Notari, B. *Synthesis and Catalytic Properties of Titanium Containing Zeolites* (eds Grobet, P., Mortier, W., Vansant, E. & Schulz-Ekloff, G.) 413–425. https://doi.org/10.1016/S0167-2991(09)60618-2 (Elsevier, 1988).

6. Tatsumi, T., Nakamura, M., Negishi, S. & Tominaga, H.-o. Shape-selective oxidation of alkanes with H_2O_2 catalysed by titanosilicate. *J. Chem. Soc., Chem. Commun.,* 476–477. http://dx.doi.org/10.1039/C39900000476 (6 1990).

7. Bhaumik, A., Kumar, P. & Kumar, R. Baeyer-Villiger rearrangement catalysed by titanium silicate molecular sieve (TS-1)/H_2O_2 system. *Catalysis letters* **40,** 47–50. https://doi.org/10.1007/BF00807456 (1996).

8. Gau, H.-M. *et al.* Chemistry of $Ti(OiPr)Cl_3$ with Chloride and Oxygen-Containing Ligands: The Roles of Alkoxide and Solvents in the Six-Coordinate Titanium Complexes. *Journal of the American Chemical Society* **118,** 2936–2941. https://doi.org/10.1021/ja952730q (1996).

9. Murray, J., Sharp, M. & Hockey, J. The polymerization of propylene by the $SiO_2/TiCl_4/AlMe_3$ system. *Journal of Catalysis* **18,** 52–56. ISSN: 0021-9517. https://doi.org/10.1016/0021-9517(70)90311-8 (1970).

10. Thomas, J. M. *et al.* The Identity in Atomic Structure and Performance of Active Sites in Heterogeneous and Homogeneous, Titanium-Silica Epoxidation Catalysts. *The Journal of Physical Chemistry B* **103,** 8809–8813. https://doi.org/10.1021/jp991991+ (1999).

11. Murugavel, R. & Roesky, H. W. Titanosilicates: Recent Developments in Synthesis and Use as Oxidation Catalysts. *Angewandte Chemie International Edition in English* **36,** 477–479. https://doi.org/10.1002/anie.199704771 (1997).

12. Puurunen, R. L. Formation of Metal Oxide Particles in Atomic Layer Deposition During the Chemisorption of Metal Chlorides: A Review. *Chemical Vapor Deposition* **11,** 79–90. https://doi.org/10.1002/cvde.200400021 (2005).

13. Haukka, S., Lakomaa, E. L., Jylha, O., Vilhunen, J. & Hornytzkyj, S. Dispersion and distribution of titanium species bound to silica from titanium tetrachloride. *Langmuir* **9,** 3497–3506. https://doi.org/10.1021/la00036a026 (1993).

14. Chen, L., Chuah, G. & Jaenicke, S. Propylene epoxidation with hydrogen peroxide catalyzed by molecular sieves containing framework titanium. *Journal of Molecular Catalysis A: Chemical* **132,** 281–292. ISSN: 1381-1169. https://doi.org/10.1016/S1381-1169(97)00276-8 (1998).

15. Castillo, R., Koch, B., Ruiz, P. & Delmon, B. Influence of preparation methods on the texture and structure of titania supported on silica. *J. Mater. Chem.* **4,** 903–906. http://dx.doi.org/10.1039/JM9940400903 (6 1994).

16. Gao, X., Bare, S. R., Fierro, J. L. G., Banares, M. A. & Wachs, I. E. Preparation and in-Situ Spectroscopic Characterization of Molecularly Dispersed Titanium Oxide on Silica. *The Journal of Physical Chemistry B* **102,** 5653–5666. https://www.doi.org/10.1021/jp981423e (1998).

17. Morrow, B. A. & Hardin, A. H. Raman spectra of some hydrogen sequestering agents chemisorbed on silica. *The Journal of Physical Chemistry* **83,** 3135–3141. https://doi.org/10.1021/j100487a015 (1979).

18. Morrow, B. & McFarlan, A. Chemical reactions at silica surfaces. *Journal of Non-Crystalline Solids* **120,** 61–71. ISSN: 0022-3093. https://doi.org/10.1016/0022-3093(90)90191-N (1990).

19. Haukka, S., Lakomaa, E. L. & Root, A. An IR and NMR study of the chemisorption of titanium tetrachloride on silica. *The Journal of Physical Chemistry* **97,** 5085–5094. https://doi.org/10.1021/j100121a040 (1993).

20. Kunawicz (née Murray), J., Jones, P. & Hockey, J. A. Reactions of silica surfaces with hydrogen sequestering agents. *Trans. Faraday Soc.* **67,** 848–853. http://dx.doi.org/10.1039/TF9716700848 (0 1971).

21. Sierra, M. S., Ruiz, J., Mayoral, J., Proietti, M. & Fraile, J. Structural study of titanium-modified silica catalysts by EXAFS and XANES. *Journal De Physique Iv* **7.** https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=2050287 (1997).

22. Tanaka, T., Nakajima, T. & Yamashita, K. Density functional study on the adsorption and surface reactions on SiO_2 in TiN-CVD using $TiCl_4$ and NH_3. *Thin Solid Films* **409.** Proceedings of the 2nd Asian Conference on Chemical Vapour Deposition, 51–57. ISSN: 0040-6090. https://doi.org/10.1016/S0040-6090(02)00103-7 (2002).

23. Morrow, B. & McFarlan, A. Infrared and gravimetric study of an aerosil and a precipitated silica using chemical and hydrogen/deuterium exchange probes. *Langmuir* **7,** 1695–1701. https://doi.org/10.1021/la00056a022 (1991).

24. Rice, G. L. & Scott, S. L. Characterization of Silica-Supported Vanadium(V) Complexes Derived from Molecular Precursors and Their Ligand Exchange Reactions. *Langmuir* **13,** 1545–1551. https://doi.org/10.1021/la960679d (1997).

25. For practical reasons, measurement of the Ti uptake by the silica is simpler, more accurate and more reproducible than measurement of either the HCl yield or the Ti/Cl ratio on the surface (via hydrolysis and measurement of the HCl liberated). It is difficult to completely remove the HCl impurity from volatile $TiCl_4$, adventitious HCl is generated by the reaction of $TiCl_4$ with vacuum grease, and HCl itself reacts with grease.

26. Taha, Z. A. *Silica-supported vanadium complexes: Structure, characterization and reactivity, especially towards olefins.* PhD thesis (University of Ottawa (Canada), 2004).

27. Taha, Z. A., Deguns, E. W., Chattopadhyay, S. & Scott, S. L. Formation of Digallium Sites in the Reaction of Trimethylgallium with Silica. *Organometallics* **25,** 1891–1899. ISSN: 0276-7333. https://doi.org/10.1021/om051034o (2006).

28. Fleischman, S. D. & Scott, S. L. Evidence for the Pairwise Disposition of Grafting Sites on Highly Dehydroxylated Silicas via Their Reactions with $Ga(CH_3)_3$. *J. Am. Chem. Soc.* **133,** 4847–4855. https://www.doi.org/10.1021/ja108905p (2011).

29. Khan, S. A., Vandervelden, C. A., Scott, S. L. & Peters, B. Grafting metal complexes onto amorphous supports: From elementary steps to catalyst site populations: Via kernel regression. *React. Chem. Eng.* **5,** 66–76. https://www.doi.org/10.1039/c9re00357f (2020).

Chapter 7

Conclusions

7.1 Summary

In summary, this thesis describes the development of new computational methods to model the synthesis and reactivity of atomically dispersed catalysts on amorphous supports. Shortcomings of current modeling methodologies are discussed and efficient machine learning aided computational methods to overcome these shortcomings are presented. The efficacy of the developed methods is demonstrated on toy systems. Furthermore, the developed methods are applied to real examples and their predictions are validated against experimental outcomes. Our work establishes a systematic procedure to computationally probe amorphous catalysts and paves the way to understand the structure and mechanism of these elusive catalytic materials.

In chapter 2, we develop a ML parametrized population balance modeling framework to model grafting of metal complexes onto amorphous supports. A ML model is trained to learn grafting barriers as a function of local site environment on amorphous silica. The trained ML model parameterizes a population balance model to predict the evolving distribution of grafting sites. We also develop a 2D model of amorphous silica for which

benchmark calculations can be performed for thousands of sites. For this 2D model, the parameterized population balance model is demonstrated to predict the evolving population of ca. 20,000 sites with a ML model trained on 100 sites. Chapter 3 presents an importance learning (IL) algorithm to efficiently calculate the site-averaged activation barrier of the grafted catalyst. The population of grafted sites, as predicted by the population balance model in chapter 2, is used as an input for IL. Our method iteratively trains a ML model on the most important (active) sites. IL requires three orders of magnitude fewer samples than random sampling to converge the site-averaged activation barrier with the same uncertainty. Given an atomistic model of amorphous silica these methods provide, for the first time, a framework to model the synthesis and reactivity of atomically dispersed amorphous catalysts. Furthermore, the 2D model of amorphous silica developed here provides a method development platform for disordered catalysts. This model was used in another study (not described here) to develop an IL method to efficiently calculate the site-averaged turnover frequency for disordered catalysts.[1]

Chapter 4 describes the application of the population balance modeling framework, developed in chapter 2, to a real system. We model the grafting of $TiCl_4$ onto amorphous silica and show how local structure can not only affect grafting kinetics but also the grafting mechanism. The final distribution of grafted Ti sites is shown to be in agreement with experimental observations. Furthermore, our calculations show that the nature of the dominant surface species is kinetically controlled. This demonstrates how kinetic modeling at realistic conditions is important for interpreting the results of grafting experiments. It further demonstrates the importance of considering a distribution of structural environments obtained from realistic silica models instead of single-site models routinely used in the literature.

The computational methods developed in this work use atomistic models of amorphous silica generated by other groups. Different groups have used different protocols

to generate these atomistic models. And all of them have claimed that their models are representative of real amorphous silicas. Chapter 5 compares atomistic models of amorphous silica from different studies. We demonstrate that models generated using different protocols have different structural features. And we discuss how differences in the structure of these models can influence the grafting thermodynamics/kinetics and the catalytic activity of the grafted catalyst. We outline some ideas to validate atomistic models against experimental observables relevant to catalysis. This work motivates future studies to systematically investigate the effect of generation protocols on the structure of the generated models and identify parameters which lead to models consistent with several experimental observables.

Most grafting experiments attempt to qualitatively estimate the nature of the grafted species using information from spectroscopic measurements (IR, EXAFS, NMR, etc.) and mass balance techniques. Chapter 6 presents a site balance algebra to quantify the amounts of different grafted species (monopodal, bipodal, tripodal, etc.) on amorphous supports. We apply this framework to quantify the amounts of monopodal and bipodal species in experiments grafting $TiCl_4$ onto amorphous silica. We further demonstrate how this framework can be used to inspect the consistency of grafting experiments.

7.2 Outlook

Several challenges remain to be solved before we can develop a complete in silico model of amorphous catalysts. Some of these challenges have been outlined throughout this thesis. In this section we present a few extensions of this work which can be immediately implemented to develop more accurate models.

The ML model to learn grafting barriers in chapter 2 was trained on a randomly sampled set of sites. Majority of the sites sampled by this brute-force strategy are near

the peak of the distribution and fewer sites are sampled towards the tails. This causes the ML model to be poorly trained in these sparsely populated regions. Hence, the ML parameterized population balance model makes errors in predicting the population of sites in the tails of the distribution as shown in fig. 2.10. Accurately predicting the population of sites in the tails can be important in cases where sites in sparsely populated regions, once grafted, are responsible for most of the activity. Such cases may require the implementation of biased sampling strategies to learn grafting kinetics of sites in sparsely populated regions of the distribution. One possible solution can be to use an importance learning algorithm which importance samples sites weighted by the inverse density.[2] This strategy will sample sites from a flat histogram, thus making it more probable to sample sites from low density tails of the distribution.

The model of grafting $TiCl_4$ onto amorphous silica, presented in chapter 3, only considered the effect of the dihedral angle on the free energy pathway. Other coordinates might also influence grafting kinetics/thermodynamics. The procedure to identify important features of sites, described in section 2.6.3, can be used to identify coordinates important for describing $TiCl_4$ grafting. Furthermore, we used a minimal model to represent the vicinal site without considering the extended silica matrix. Future studies should use larger models of silica to account for the extended silica matrix.

Population balance modeling and importance learning can be put together to develop a complete in silico model of atomically dispersed catalysts. The population balance modeling framework can be used to predict the kinetics/thermodynamics of grafting organometallic precursors onto amorphous silica models. Importance learning can then be used to calculate site-averaged kinetics (site-averaged activation barrier and turnover frequency) of the grafted catalyst. These calculations can be validated by different experiments. The predicted distribution of sites can be validated by the estimated populations of sites obtained using mass balance experiments and various spectroscopic measurements

(EXFS, IR, XAS, etc.). Grafting kinetics can be validated by measuring the evolution kinetics of the gas phase species eliminated in grafting reactions. The calculated site-averaged kinetic properties can be validated by measuring the kinetics of the catalytic reaction. Some of these experimental measurements are already available for various catalysts in the literature. In particular, many experimental studies have estimated the population of grafted sites[3–6] and measured the reactivity of the grafted catalyst.[7–9] Some examples to consider are: Cr/SiO_2 catalysts for ethylene polymerization,[7] Mo/SiO_2 catalysts for olefin metathesis,[10] and Ti/SiO_2 catalysts for olefin epoxidation.[4] Modeling Ti/SiO_2 catalysts will be the most natural extension of this work as we already have a model of the grafting step (Chapter 4). Such a demonstration will be the first complete in silico model of such catalysts. It will also serve as a tool to validate silica models.

Bibliography

1. Vandervelden, C. A., Khan, S. A. & Peters, B. Importance learning estimator for the site-averaged turnover frequency of a disordered solid catalyst. *The Journal of Chemical Physics* **153,** 244120. https://doi.org/10.1063/5.0037450 (2020).

2. Wang, F. & Landau, D. P. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **86,** 2050. https://doi.org/10.1103/PhysRevLett.86.2050 (2001).

3. Deguns, E. W., Taha, Z., Meitzner, G. D. & Scott, S. L. An X-ray Absorption Study of Two VOCl3-Modified Silicas: Evidence for Chloride-Silica Interactions. *The Journal of Physical Chemistry B* **109,** 5005–5011. https://www.doi.org/10.1021/jp045886y (2005).

4. Rice, G. L. & Scott, S. L. Characterization of Silica-Supported Vanadium(V) Complexes Derived from Molecular Precursors and Their Ligand Exchange Reactions. *Langmuir* **13,** 1545–1551. https://doi.org/10.1021/la960679d (1997).

5. Lwin, S., Li, Y., Frenkel, A. I. & Wachs, I. E. Nature of WOx Sites on SiO_2 and Their Molecular Structure-Reactivity/Selectivity Relationships for Propylene Metathesis. *ACS Catal.* **6,** 3061–3071. https://www.doi.org/10.1021/acscatal.6b00389 (2016).

6. Coperet, C. *et al.* Surface Organometallic and Coordination Chemistry toward Single-Site Heterogeneous Catalysts: Strategies, Methods, Structures, and Activities. *Chem. Rev.* **116,** 323–421. https://www.doi.org/10.1021/acs.chemrev.5b00373 (2016).

7. McDaniel, M. P. A Review of the Phillips Supported Chromium Catalyst and Its Commercial Use for Ethylene Polymerization. English. *Adv. Catal.* **53,** 123–606. https://www.doi.org/https://doi.org/10.1016/S0360-0564(10)53003-7 (2010).

8. Zhong, L. *et al.* Spectroscopic and structural characterization of Cr(II)/SiO_2 active site precursors in model Phillips polymerization catalysts. *J. Catal.* **293,** 1–12. https://www.doi.org/10.1016/j.jcat.2012.05.014 (2012).

9. Samantaray, M. K. *et al.* Surface organometallic chemistry in heterogeneous catalysis. *Chem. Soc. Rev.* **47,** 8403–8437. https://www.doi.org/10.1039/C8CS00356D (2018).

10. Amakawa, K. *et al.* Active Sites in Olefin Metathesis over Supported Molybdena Catalysts. *ChemCatChem* **7,** 4059–4065. https://www.doi.org/10.1002/cctc.201500725 (2015).

Appendix A

Supporting information for Grafting catalysts onto amorphous supports: from elementary steps to site populations via kernel regression

A.1 Parameters in model of grafting barriers and lattice displacements

A.1.1 Grafting temperature and dimensionless precursor concentration (m)

The grafting temperature was chosen to be 298.15 K (room temperature). The ratio of the ML_2 concentration in the gas phase to the reference concentration ($\hat{V}_0^{-1}$), $m = [\mathbf{ML_2}]/\hat{V}_0^{-1}$, was set to the ratio of CrO_2Cl_2 vapor pressure at 298.15 K (20 Torr) to

atmospheric pressure (760 Torr), giving $m = 0.026$.

A.1.2 DFT computational details

$D_{\mathbf{M}-O}$, $D_{\mathbf{M}\cdots O}$, $a_{\mathbf{M}-O}$, $a_{\mathbf{M}\cdots O}$, and $\Delta G^{\ddagger}_{ref}$ were set using density functional theory (DFT) calculations. All DFT calculations were performed with the ωB97X-D functional.[1] The def-2TZVP basis set[2] was used for chromium and the TZVP basis set[3,4] was used for all other atoms. The Berny algorithm, as implemented in Gaussian 16, was used to find minima and transition states.[5] The RMS and maximum forces were required to be less than 3.00×10^{-4} Hartrees/Bohr and 4.50×10^{-4} Hartrees/Bohr, respectively, while the RMS and maximum displacements were required to be less than 1.20×10^{-3} Bohr and 1.80×10^{-3} Bohr respectively. Transition states were required to have one imaginary frequency.

To make a reference site for DFT calculations, the bis(silanolato)chromium(II) cluster was optimized and its peripheral atom positions were held fixed for all subsequent computations to mimic a rigid support, Fig. A.1.[6]

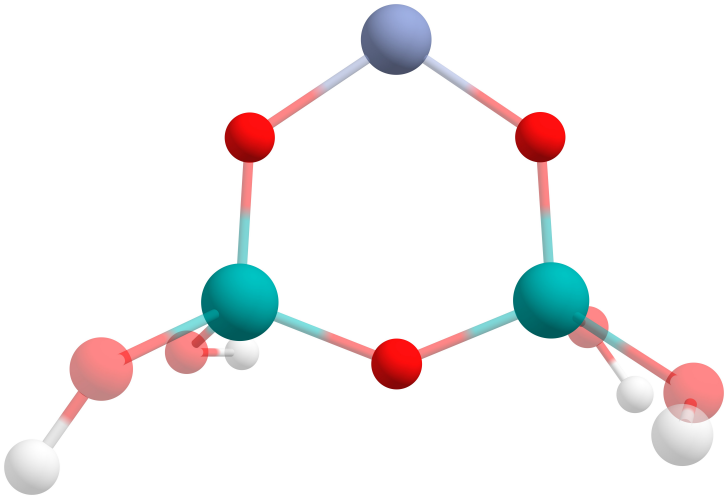


Figure A.1: The optimized bis(silanolato)chromium(II) cluster. Color scheme: oxygen (red), hydrogen (white), silicon (blue), and chromium (purple). Peripheral atoms (fixed) are transparent.

A.1.3 Morse potential parameters

The M-O bond strength ($D_{\mathbf{M}-O}$) was calculated by removing the Cr atom from the bis(silanolato) chromium(II) and performing a single-point energy calculation, Fig. A.2. $D_{\mathbf{M}-O}$ was calculated using

$$D_{\mathbf{M}-O} = E_{II} + E_{Cr} - E_I. \tag{A.1}$$

Here E_{II} is the electronic energy of structure II, E_{Cr} is the electronic energy of a Cr atom, and E_I is the electronic energy of the bis(silanolato)chromium(II) cluster (structure I). We get $D_{\mathbf{M}-O}$=524.4 kJ/mol.

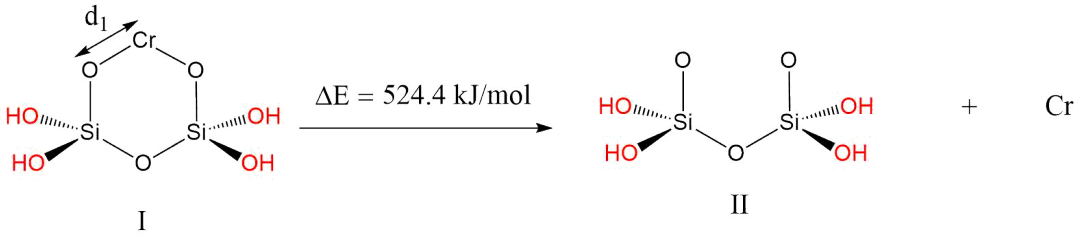


Figure A.2: Calculation of the M-O bond strength. Electronic energies of the optimized Cr(II) cluster (left) and cluster with dissociated Cr (right). Peripheral atoms (fixed) are red. d1 is used as the displacement variable in a Morse potential model (below).

The Morse potential width (a) can be related to the force constant by a second-order Taylor expansion of $V(r)$ around the equilibrium bond length (r_{eq}):

$$V(r) \approx \frac{d^2V}{dr^2}\frac{(r-r_{eq})^2}{2!} = k\frac{(r-r_{eq})^2}{2!} = Da^2(r-r_{eq})^2 \tag{A.2}$$

Here, k is the force consant. The zeroth-order term of the taylor expansion evaluates to 0 by construction, while the first derivative evaluates to 0 because r_{eq} corresponds to the minimum of the potential energy surface. Thus, $a = [k/2D]^{0.5}$. The force constant was computed using DFT by calculating the second derivative of the potential energy with

respect to the Cr-O bond length (d_1). We obtain $k_{\mathbf{M}-O} = 0.2063$ Ha/Bohr and $a_{\mathbf{M}-O} = 1$. Here $a_{\mathbf{M}-O}$ was non-dimensionalized by d_1. The non-dimensionalized equilibrium bond distance for the M-O bond ($r_{\mathbf{M}-O,eq}$) was set to 1.

To calculate the $\mathbf{M}\cdots O$ bond strength ($D_{\mathbf{M}\cdots O}$), a water molecule was adsorbed on the bare Cr cluster, and the cluster was reoptimized while keeping the positions of the peripheral atoms fixed, Fig. A.3. The bond strength of the $\mathrm{Cr}\cdots\mathrm{OH_2}$ bond was calculated using

$$D_{\mathbf{M}\cdots O} = E_I + E_{H_2O} - E_{III} \tag{A.3}$$

Here E_I is the electronic energy of structure I, E_{H_2O} is the electronic energy of the optimized water molecule, and E_{III} is the electronic energy of structure III. We get $D_{\mathbf{M}\cdots O} = 117$ kJ/mol.

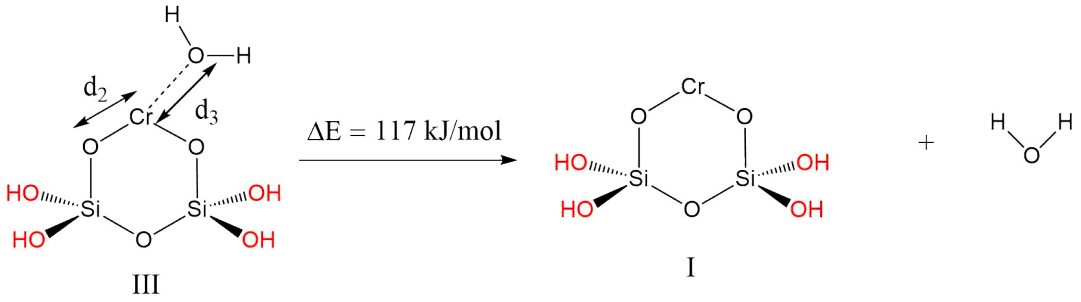


Figure A.3: Calculation of the M···O bond strength. Electronic energies of the Cr cluster with H_2O adsorbed (left) and the bare Cr cluster and H_2O in the gas phase (right). Peripheral atoms (fixed) are red. d_2/d_3 was used to compute $r_{M\cdots O,eq}$ and d_3 was used as a displacement variable in the Morse potential model for M···O.

$k_{\mathbf{M}\cdots O}$ was computed as the second derivative of the energy of structure III with respect to the $M\cdots O$ bond length (d_3). We get $k_{\mathbf{M}\cdots O} = 0.054$ Ha/Bohr and $a_{\mathbf{M}\cdots O} = 2.3$ (eq. (A.2)). Here $a_{\mathbf{M}\cdots O}$ was non-dimensionalized by d_3 (Fig. A.3).

The non-dimensionalized equilibrium $k_{\mathbf{M}\cdots O}$ bond length ($r_{\mathbf{M}\cdots O,eq}$) was set to d_3/d_2. This yields $r_{\mathbf{M}\cdots O,eq} = 1.16$.

A.1.4 Lattice displacements

The lattice points were displaced by drawing dispalcements according to a bivariate Gaussian distribution using the numpy.random.multivariate_normal function in python:[7]

$$p(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{2\pi|\boldsymbol{\Sigma}|^{\frac{1}{2}}} exp(\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})) \tag{A.4}$$

Here Σ is a 2×2 dimensional covariance matrix, $\boldsymbol{\mu} \in \boldsymbol{R}_2$ is the mean, and $\boldsymbol{x} \in \boldsymbol{R}_2$ is a 2D random variable representing displacement of lattice points. The covariance matrix was set equal to a scalar diagonal matrix

$$\boldsymbol{\Sigma} = \sigma^2_{lattice} I. \tag{A.5}$$

Here I is the 2×2 identity matrix. $\sigma^2_{lattice}$ was set to 0.00022 and the value of $\boldsymbol{\mu}$ was set to (0,0).

A.1.5 ϵ_{HL}, V_*, ΔPV, and ΔS^o to compute grafting free energy

From eq. A.4, it follows that the grafting energy for a site on the unperturbed lattice is given by

$$\Delta E_{unperturbed} = 2\epsilon_{\boldsymbol{HL}} - (V_* + 2\epsilon_{\boldsymbol{ML}}) + V_{\boldsymbol{M}*}(\boldsymbol{x}_{unperturbed}). \tag{A.6}$$

Here V_{M*} is obtained by optimizing the metal position in a site on an unpertubed lattice. Similarly, from eq. 2.8 it follows that the grafting free energy on an unperturbed site is given by

$$\Delta G^o_{unperturbed} = 2\epsilon_{\boldsymbol{HL}} - (V_* + 2\epsilon_{\boldsymbol{ML}}) + V_{\boldsymbol{M}^*}(\boldsymbol{x}_{unperturbed}) + \Delta PV - T\Delta S^o. \tag{A.7}$$

Rearranging eq. A.7, we get

$$2\epsilon_{\boldsymbol{HL}} - (V_* + 2\epsilon_{\boldsymbol{ML}}) + \Delta PV - T\Delta S^o = \Delta G^o_{unperturbed} - V_{\boldsymbol{M}^*}(\boldsymbol{x}_{unperturbed}) \tag{A.8}$$

Using Morse potential parameters from Section A.1.3, we obtain $V_{\boldsymbol{M}^*}(\boldsymbol{x}_{unperturbed}) = -1259.57$ kJ/mol. To make grafting favorable for a reference site, $\Delta G^o_{unperturbed}$ was set equal to -30 kJ/mol. This yields $2\epsilon_{HL} - (V_* + 2\epsilon_{ML} + \Delta PV - T\Delta S^o) = 1229.56$ kJ/mol. ϵ_{HL}, V_*, ϵ_{ML}, ΔPV, and ΔS^o always occur together in the combination on the LHS of eq. A.8, therefore they do not need to be determined separately.

A.1.6 Reference free energy barrier and linear free energy relation ($\Delta G^{\ddagger}_{ref}$)

The LFER for an unperturbed site is given by

$$\Delta G^{\ddagger}_{unperturbed}(\boldsymbol{r}) = \Delta G^{\ddagger}_{ref} + \alpha \Delta G^o_{unperturbed}(\boldsymbol{r}). \tag{A.9}$$

$\Delta G^{\ddagger}_{unperturbed}(\boldsymbol{r})$ was set equal to the DFT-computed activation barrier for CrO_2Cl_2 grafting to a vicinal disilanol model site at 1 atm pressure of CrO_2Cl_2 (Fig. A.4).

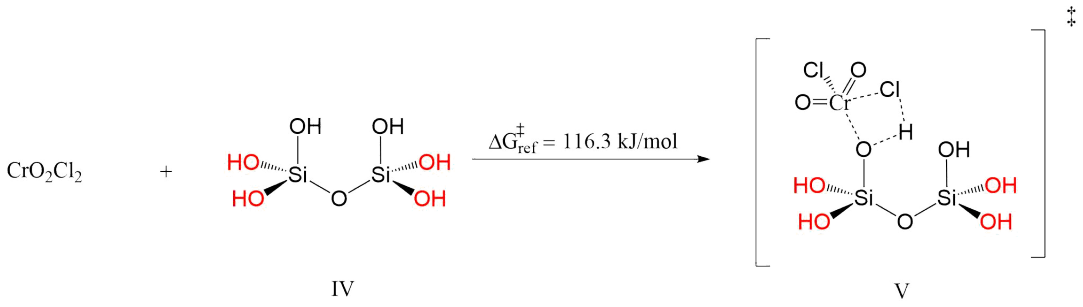


Figure A.4: Calculation of the reference free energy barrier. Free energies of CrO_2Cl_2 in the gas phase with the vicinal silanol site (left) and the transition state for CrO_2Cl_2 grafting to the vicinal silanol site (right). Free energies are at 1 atm CrO_2Cl_2 and 298.15 K. Peripheral atoms (fixed) are red.

Using $\Delta G^{o}_{unperturbed} = -30$ kJ/mol (Section A.1.5) and solving for $\Delta G^{\ddagger}_{ref}$ we get

$$\Delta G^{\ddagger}_{ref} = 131.3 kJ/mol. \tag{A.10}$$

A.2 Effect of training set size on test set error

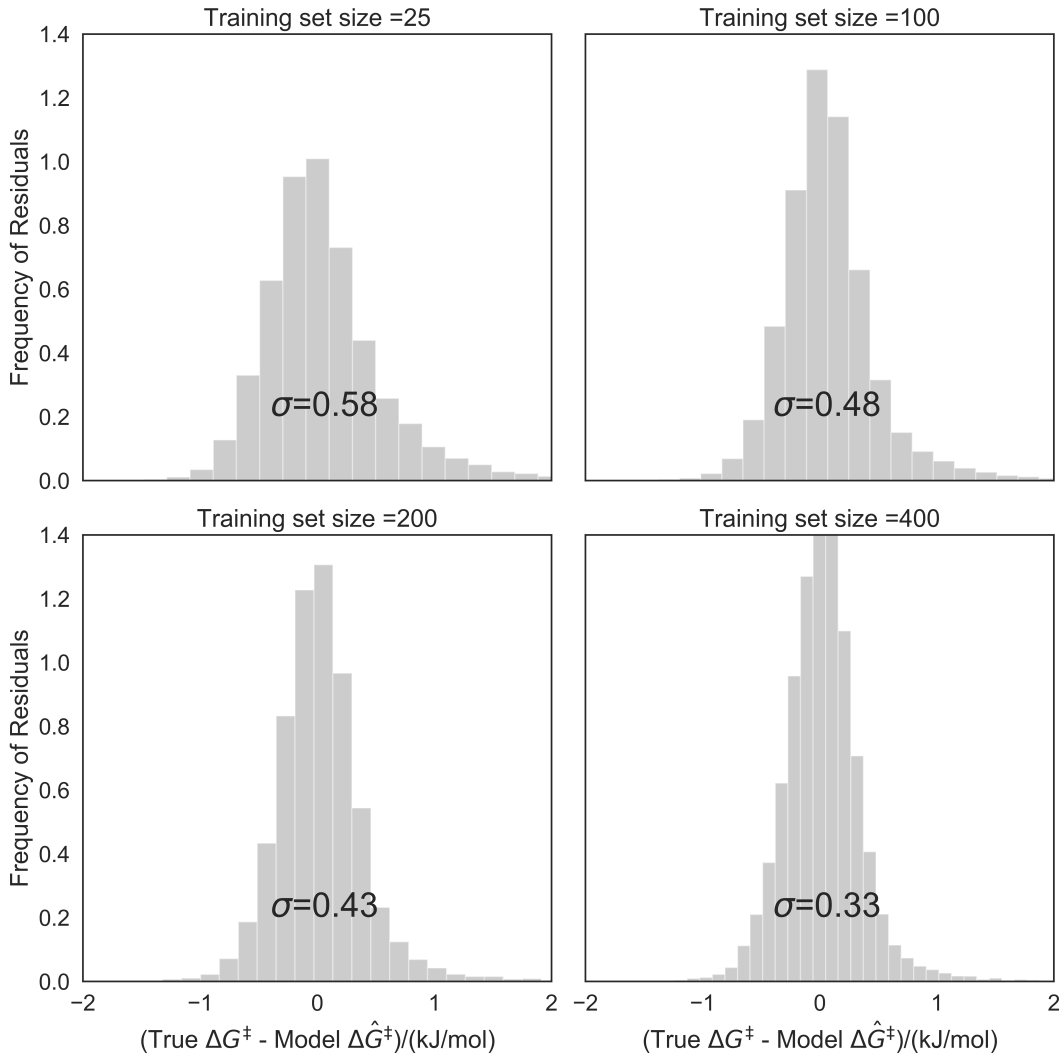


Figure A.5: Residual distributions for predicted grafting barriers as a function of training set size for all $\approx 20,000$ sites. As expected, the width of the residual distribution decreases on increasing the training set size.

Bibliography

1. Chai, J.-D. & Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **10,** 6615–6620. https://www.doi.org/10.1039/B810189B (2008).

2. Feller, D. The role of databases in support of computational chemistry calculations. *J. Comput. Chem.* **17,** 1571–1586. https://www.doi.org/10.1002/(SICI)1096-987X(199610)17:13%3C1571::AID-JCC9%3E3.0.CO;2-P (1996).

3. Schuchardt, K. L. *et al.* Basis Set Exchange: A Community Database for Computational Sciences. *J. Chem. Inf. Model.* **47,** 1045–1052. https://www.doi.org/10.1021/ci600510j (2007).

4. Godbout, N., Salahub, D. R., Andzelm, J. & Wimmer, E. Optimization of Gaussian-type basis sets for local spin density functional calculations. Part I. Boron through neon, optimization technique and validation. *CJC University of Calgary 50th Anniversary Virtual Issue* **01,** 560–571. https://doi.org/10.1139/v92-079 (1992).

5. Frisch, M. J. *et al.* Gaussian16. Gaussian Inc. Wallingford CT 2016.

6. Fong, A., Yuan, Y., Ivry, S. L., Scott, S. L. & Peters, B. Computational Kinetic Discrimination of Ethylene Polymerization Mechanisms for the Phillips (Cr/SiO_2) Catalyst. English. *ACS Catal.* **5,** 3360–3374. https://www.doi.org/10.1021/acscatal.5b00016 (2015).

7. Oliphant, T. E. Guide to NumPy. *CreateSpace Independent Publishing Platform* (2015).

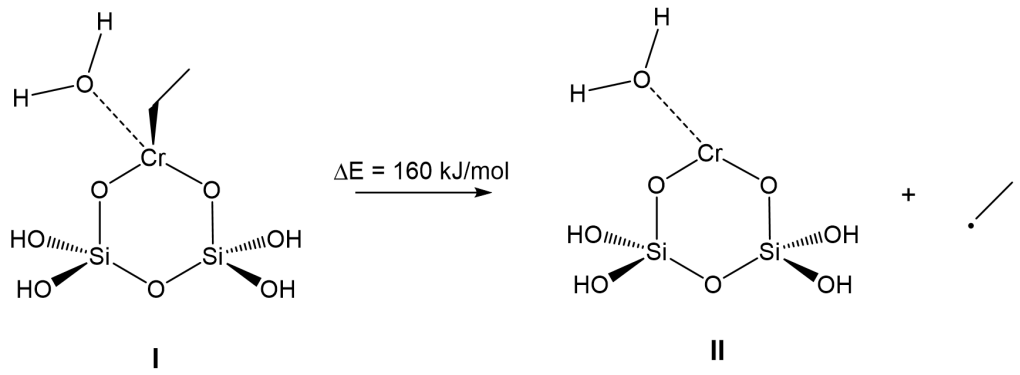
Appendix B

Supporting information for Site-averaged kinetics for catalysts on amorphous supports: An importance learning algorithm

B.1 Strength of M-A bond

The **M-A** bond strength in the quenched-disordered lattice model was chosen to approximately match the Cr-C bond strength for an alkylchromium(III) site on SiO_2. We started from a *bis*(silanolato)chromium(II) cluster model, which has been used in previous studies of Cr/SiO_2 catalysts.[1,2] Labile siloxane coordination was modeled by binding a water molecule. The **M-A** bond strength was calculated according to Scheme S1 and density functional theory calculations, and was computed as $\varepsilon_{\mathbf{M\text{-}A}} = 160$ kJ/mol. We chose a one-electron redox pathway (as opposed to a two-electron redox pathway) to ensure a strong bond energy for the chemisorption step. We stress that the bond strength

was chosen only to ensure a realistic model. The model is not intended make accurate predictions for Cr/SiO_2 olefin polymerization catalysts.



DFT calculations were carried out using Gaussian16.[3] All energies were calculated with the range-separated density functional, ωB97X-D.[4] The def2-TZVP basis set was used for Cr[5] and TZVP was used for C, H, O, and Si atoms.[6] All minima have zero imaginary frequencies. The peripheral OH atoms of the cluster model were also held constrained to model the geometric constraints of an extended silica network. The peripheral atom constraints were found by optimizing the bare Cr(II) cluster. The same peripheral atom constraints were applied to structures **I** and **II**. Cartesian coordinates of the optimized clusters are tabulated below Table S.

Table B.1: Spin contamination before (S2) and after annihilation (S2A) of highest spin contaminant; energies in Hartrees

Species	Energy	S2	S2A
I	-2229.395021	6.0107	6.0000
II	-2308.613831	3.8158	3.7506
C3H7	-157.7869992	0.7544	0.7500

B.2 Derivation of apparent activation energy

In this section, a formula for the apparent activation energy of a site, $E_a(\mathbf{x}_i)$, is derived. The apparent activation energy for site i is given by

$$E_a(\mathbf{x}_i) = -\frac{d \ln \text{𝓇}_i}{d\beta} \tag{B.1}$$

From eqs. 3.2 - 3.5, the turnover frequency (TOF) of a site, 𝓇_i, can be expressed as

$$\text{𝓇}_i = k_2 K(\mathbf{x}_i) c_{\mathbf{A}} = \frac{k_B T}{h} \exp\left[-\frac{\Delta H^{\ddagger} - T\Delta S^{\ddagger}}{k_B T}\right] \exp\left[-\frac{\Delta H(\mathbf{x}_i) - T\Delta S}{k_B T}\right] c_{\mathbf{A}} \tag{B.2}$$

Taking the natural logarithm of the eq. B.2, grouping temperature dependent terms, and simplifying yields

$$\ln \text{𝓇}_i = \ln c_{\mathbf{A}} - \ln h + \frac{\Delta S + \Delta S^{\ddagger}}{k_B} - \beta \Delta H(\mathbf{x}_i, \beta) - \beta \Delta H^{\ddagger} - \ln \beta \tag{B.3}$$

where $\beta = 1/k_B T$. From eq. 3.6, $\Delta H(\mathbf{x}_i)$ is temperature dependent through kBT. Taking the derivative of eq. B.3 gives

$$E_a(\mathbf{x}_i) = \frac{d}{d\beta} \beta \Delta H(\mathbf{x}_i, \beta) + \Delta H^{\ddagger} + k_B T \tag{B.4}$$

Inserting eq. 3.6 into $\Delta H(\mathbf{x}_i)$ to evaluate the derivative gives

$$\begin{aligned} \frac{d}{d\beta} \beta \Delta H(\mathbf{x}_i, \beta) &= \frac{d}{d\beta} [\beta V_{\mathbf{AM}*}(\mathbf{x}_i) - \beta V_{\mathbf{M}*}(\mathbf{x}_i) + 1] \\ &= V_{\mathbf{AM}*}(\mathbf{x}_i) - V_{\mathbf{M}*}(\mathbf{x}_i) \\ &= \Delta H(\mathbf{x}_i) + k_B T \end{aligned} \tag{B.5}$$

Thus, $E_a(\mathbf{x}_i)$ can be written as

$$E_a(\mathbf{x}_i) = \Delta H(\mathbf{x}_\mathrm{i}) + \Delta H^{\ddagger} + 2k_B T \tag{B.6}$$

B.3 Propagation of kernel regression model uncertainty in estimating $\langle E_a \rangle_k$

Site-averaged kinetics are estimated by importance sampling the activation energy distribution with E_a values obtained from the trained kernel regression model. Since errors in the kernel regression model propagate through the $\langle E_a \rangle_k$ calculation (beyond sampling error and error from ab initio calculations), the kernel regression model contributes additional errors. Here, we show that the regression errors, even when unbiased, will systematically bias the $\langle E_a \rangle_k$ estimate toward lower activation energy. We also show how this bias can be quantified and corrected to obtain $\langle E_a \rangle_k$ estimates with only sampling and ab initio calculation errors. Let the distribution of kernel regression activation energies be $\hat{\rho}(\hat{E}_a)$. rhoEaHat can be related to the E_a distribution, $\rho(E_a)$, by

$$\tilde{\rho}(\hat{E}_a) = \int dE_a \tilde{\rho}(E_a) P(\hat{E}_a | E_a) \tag{B.7}$$

Here, $P(\hat{E}_a|E_a)$ is the distribution of the model-predicted activation barriers around the true activation barriers, and the integral is over the all possible E_a values. The site averaged activation energy from $\hat{\rho}(\hat{E}_a)$ is

$$\langle \hat{E}_a \rangle_k = \frac{\int d\hat{E}_a \hat{E}_a e^{-\beta \hat{E}_a} \tilde{\rho}(\hat{E}_a)}{\int d\hat{E}_a e^{-\beta \hat{E}_a} \tilde{\rho}(\hat{E}_a)} \tag{B.8}$$

where $\beta = 1/k_B T$, and T is the operating temperature of the catalyst. Combining eqs. B.7 and B.8 yields

$$\langle \hat{E}_a \rangle_k = \frac{\int \int d\hat{E}_a dE_a \hat{E}_a e^{-\beta \hat{E}_a} P(\hat{E}_a | E_a)}{\int \int d\hat{E}_a dE_a e^{-\beta \hat{E}_a} P(\hat{E}_a | E_a)} \tag{B.9}$$

Assuming $\hat{\rho}(\hat{E}_a)$ is normally distributed around $\rho(E_a)$ with a standard deviation of σ_g gives:

$$\tilde{\rho}(\hat{E}_a) = \int_{-\infty}^{+\infty} dE_a \tilde{\rho}(E_a) \frac{1}{\sqrt{2\pi}\sigma_g} \exp \frac{\left(\hat{E}_a - E_a\right)^2}{2\sigma_g^2}. \tag{B.10}$$

Combining eqs. B.7 and B.10 and simplifying yields:

$$\langle \hat{E}_a \rangle_k = \frac{\int dE_a \tilde{\rho}(E_a) \int_{-\infty}^{+\infty} d\hat{E}_a \hat{E}_a \exp\left[\frac{\left(\hat{E}_a - E_a\right)^2}{2\sigma_g^2} - \beta \hat{E}_a\right]}{\int dE_a \tilde{\rho}(E_a) \int_{-\infty}^{+\infty} d\hat{E}_a \exp\left[\frac{\left(\hat{E}_a - E_a\right)^2}{2\sigma_g^2} - \beta \hat{E}_a\right]}. \tag{B.11}$$

The two integrals in eq. B.11 have closed-form solutions:

$$\int_{-\infty}^{+\infty} d\hat{E}_a \hat{E}_a \exp\left[\frac{\left(\hat{E}_a - E_a\right)^2}{2\sigma_g^2} - \beta \hat{E}_a\right] = \sqrt{2\pi}\sigma_g (E_a - \beta {\sigma_g}^2) \exp\left[\frac{\beta^2 {\sigma_g}^2}{2} - \beta E_a\right] \tag{B.12}$$

and

$$\int_{-\infty}^{+\infty} d\hat{E}_a \exp\left[\frac{\left(\hat{E}_a - E_a\right)^2}{2\sigma_g^2} - \beta \hat{E}_a\right] = \sqrt{2\pi}\sigma_g \exp\left[\frac{\beta^2 {\sigma_g}^2}{2} - \beta E_a\right] \tag{B.13}$$

Introducing eqs. B.12 and B.13 into eq. B.11 and simplifying gives

$$\langle \hat{E}_a \rangle_k = \frac{\int dE_a \tilde{\rho}(E_a)(E_a - \beta {\sigma_g}^2) e^{-\beta E_a}}{\int dE_a \tilde{\rho}(E_a) e^{-\beta E_a}} \tag{B.14}$$

From eq. B.8, it can be seen that

$$\left\langle \hat{E}_a \right\rangle_k = \langle E_a \rangle_k - \beta {\sigma_g}^2 \tag{B.15}$$

Kernel regression errors (σ_g) can be used to estimate the error in the kernel regression model predicted k-weighted activation barrier ($\langle E_a \rangle_k$) using eq. B.15. We can estimate typical size of kernel regression errors using the training set error.

B.4 Test set and training set statistics

The set of randomly sampled sites used to train the kernel regression model should sufficiently sample the main support of $\rho(E_a)$ to properly normalize $\hat{\rho}(\hat{E}_a)$ for predicting kinetic properties. Once the main support of $\rho(E_a)$ is sufficiently sampled, additional sites do not improve the normalization of $\hat{\rho}(\hat{E}_a)$ and require additional, costly structure optimizations. Figure B.1 shows the leave-one-out parity plot of the kernel regression plot trained on 25, 50, 75, and 100 randomly sampled sites.

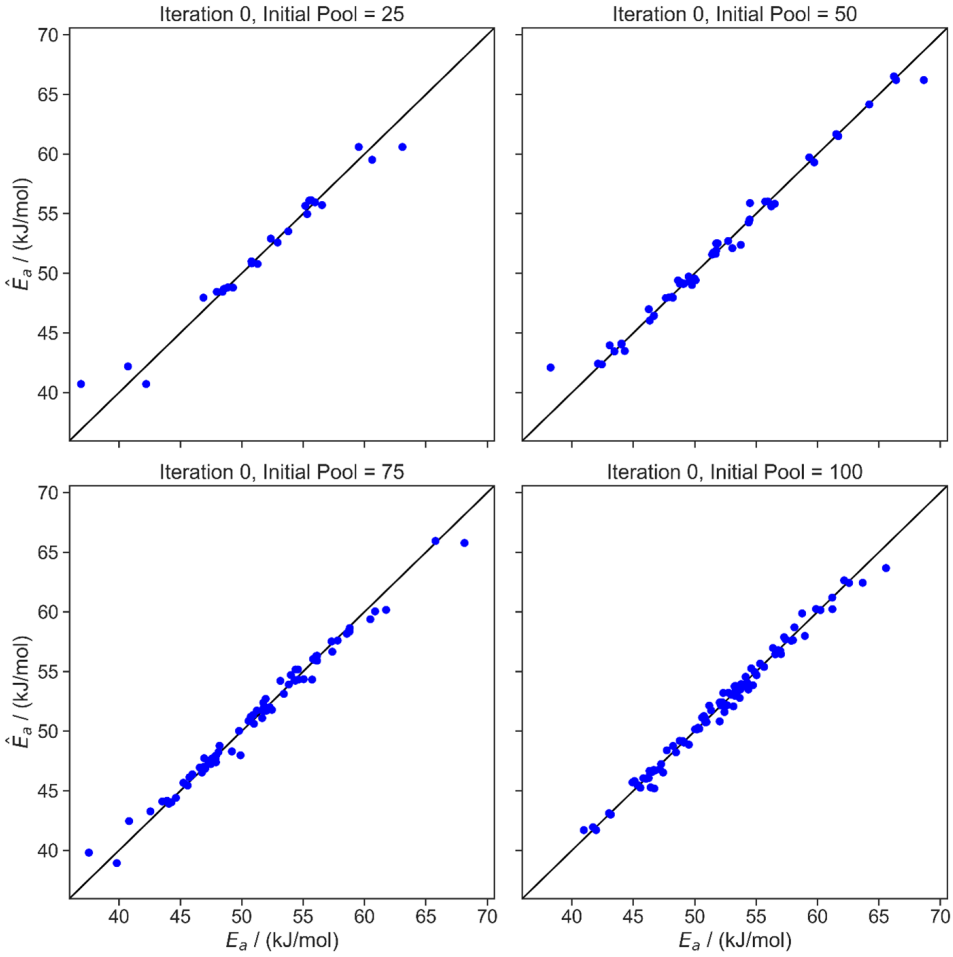


Figure B.1: Parity plot of kernel regression model trained on different initial pool sizes. An initial pool of 50 randomly selected sites samples the main support of $\rho(E_a)$.

Errors in the kernel regression model should be smaller than the width of $\rho(E_a)$ to accurately importance sample $\hat{\rho}(\hat{E}_a)$. Therefore, the initial pool should contain a set of sites with diverse local environments and activation energies to effectively train the kernel regression model. Residual distributions of all ca. 20,000 sites are shown for the kernel regression model trained on 25, 50, 75, and 100 randomly sampled sites in Figure B.2.

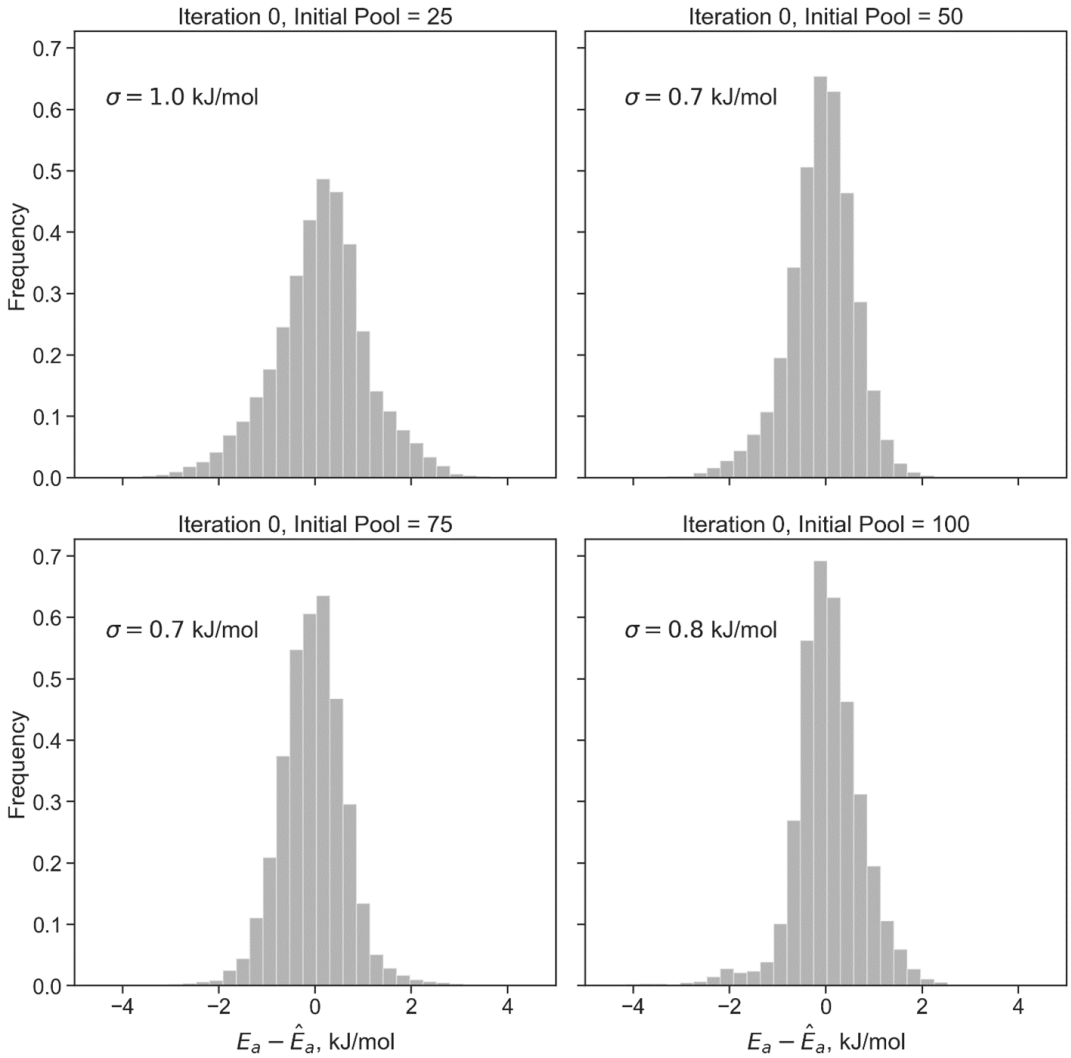


Figure B.2: Kernel regression model residual distribution for all ca. 20,000 sites with different initial pool sizes. For all initial pool sizes, the standard error is within 1.0 kJ/mol which is ca. 40 times smaller than the range of $\hat{\rho}(\hat{E}_a)$. The standard error does not decrease for initial pool sizes greater than 50.

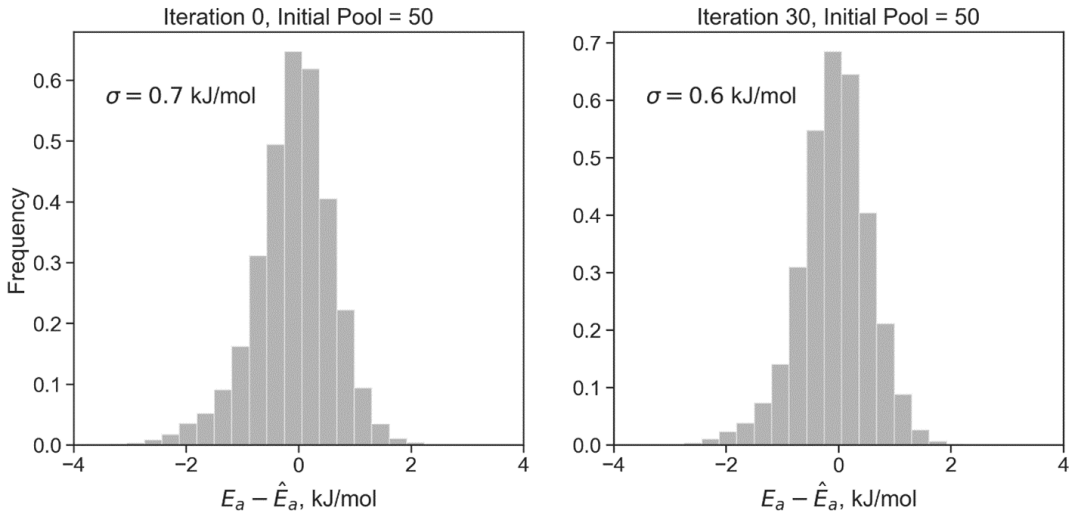


Figure B.3: Distribution of residuals for iterations 0 (left) and 30 (right) of the importance learning algorithm.

B.5 Number of samples required to estimate $\tilde{E}_a$ with the same precision $\overline{E}_a$

The $\overline{E}_a$ estimator from the importance learning algorithm (eq. 3.15) quickly converges to the correct site averaged activation energy because sites are sampled with weights $\rho(\mathbf{x})\exp[-\beta E_a(\mathbf{x})]$. Alternatively, the $\tilde{E}_a$ estimator randomly samples sites with weights $\rho(\mathbf{x})$ and computes a ratio of exponential averages (eq. 3.14). The reweighted estimator will require many more samples to converge to a precise estimate. In this section, the relative variance for the $\tilde{E}_a$ estimator is derived and the number of samples required to estimate $\tilde{E}_a$ with the same level of confidence as $\overline{E}_a$ is computed.

From eq. 3.14, $\tilde{E}_a$ is computed by

$$\begin{aligned}\tilde{E}_a &= \sum_{i=1}^{n} k(\mathbf{x}_i)E_a(\mathbf{x}_i) \Big/ \sum_i k(\mathbf{x}_i) \\ &= \frac{\widetilde{kE}_a}{\tilde{k}}\end{aligned} \tag{B.16}$$

Both $\widetilde{kE}_a$ and $\tilde{k}$ are random variables for a given sample size, so their ratio is also a random variable. Assuming $\tilde{k}$ and $\widetilde{kE}_a$ are independent and uncorrelated, the sample variance of $\tilde{E}_a$ can be approximated by:

$$\sigma^2_{\tilde{E}_a} \approx \left(\frac{\partial \tilde{E}_a}{\partial \widetilde{kE}_a}\right)^2 \sigma^2_{\widetilde{kE}_a} + \left(\frac{\partial \tilde{E}_a}{\partial \tilde{k}}\right)^2 \sigma^2_{\tilde{k}} \tag{B.17}$$

Evaluating the derivatives and dividing by $\tilde{E}_a$ yields the relative sample variance

$$\frac{\sigma^2_{\tilde{E}_a}}{\tilde{E}_a^2} = \frac{\sigma^2_{\widetilde{kE}_a}}{\left(\widetilde{kE}_a\right)^2} + \frac{\sigma^2_{\tilde{k}}}{\tilde{k}^2} \tag{B.18}$$

The relative sample variance can be related to the relative variance by the central limit theorem:[7]

$$\frac{\sigma^2_{\widetilde{kE}_a}}{\left(\widetilde{kE}_a\right)^2} + \frac{\sigma^2_{\tilde{k}}}{\tilde{k}^2} = \frac{1}{N}\left[\frac{\sigma^2_{kE_a}}{\langle kE_a\rangle^2} + \frac{\sigma^2_k}{\langle k\rangle^2}\right]_{\rho(\mathbf{x})} \equiv \left.\frac{\sigma^2_{\langle E_a\rangle_k}}{\langle E_a\rangle_k^{\,2}}\right|_{\rho(\mathbf{x})} \tag{B.19}$$

where N is the number of samples. The right most equality with subscript $\rho(\mathbf{x})$ indicates that B.19 estimates the relative variance in the E_a estimate as computed with a sample from $\rho(\mathbf{x})$. The number of random samples required to match the uncertainty of the $\overline{E}_a$ estimator from the importance learning algorithm is found by equating the relative uncertainties of the two estimators:

$$\left.\frac{\sigma^2_{\langle E_a\rangle_k}}{\langle E_a\rangle_k^{\,2}}\right|_{k(\mathbf{x})\rho(\mathbf{x})} = \left.\frac{\sigma^2_{\langle E_a\rangle_k}}{\langle E_a\rangle_k^{\,2}}\right|_{\rho(\mathbf{x})} \tag{B.20}$$

Inserting eq. B.19 in the right hand side of B.20 and solving for N yields

$$N = \left[\frac{\sigma^2_{\langle E_a\rangle_k}}{\langle E_a\rangle_k^{\,2}}\right]^{-1}_{k(\mathbf{x})\rho(\mathbf{x})} \left[\frac{\sigma^2_{kE_a}}{\langle kE_a\rangle^2} + \frac{\sigma^2_k}{\langle k\rangle^2}\right]_{\rho(\mathbf{x})} \tag{B.21}$$

The relative uncertainty in the $\overline{E}_a$ estimator is (0.75 kJ/mol) / (40.5 kJ/mol) = 1.85 %.

Since $\rho(E_a)$ can be precisely calculated for our simple model, $\sigma^2_{kE_a}/\langle kE_a\rangle^2$ and $\sigma^2_k/\langle k\rangle^2$ can be computed exactly. Evaluating eq. B.21 gives

$$N = (0.0185)^{-2} \times (28.1 + 41.2) \approx 200,000 \tag{B.22}$$

Therefore, the reweighting estimator $\tilde{E}_a$ requires about 200,000 sites for the same level of confidence that the importance learning estimator $\overline{E}_a$ achieved with less than 100 sites.

Bibliography

1. Fong, A., Yuan, Y., Ivry, S. L., Scott, S. L. & Peters, B. Computational Kinetic Discrimination of Ethylene Polymerization Mechanisms for the Phillips (Cr/SiO_2) Catalyst. English. *ACS Catal.* **5,** 3360–3374. https://www.doi.org/10.1021/acscatal.5b00016 (2015).

2. Fong, A., Vandervelden, C., Scott, S. L. & Peters, B. Computational Support for Phillips Catalyst Initiation via Cr-C Bond Homolysis in a Chromacyclopentane Site. *ACS Catal.* **8,** 1728–1733. https://doi.org/10.1021/acscatal.7b03724 (2018).

3. Frisch, M. J. *et al.* Gaussian16. Gaussian Inc. Wallingford CT 2016.

4. Chai, J.-D. & Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **10,** 6615–6620. https://www.doi.org/10.1039/B810189B (2008).

5. Weigend, F. & Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **7,** 3297–3305. https://www.doi.org/10.1039/b508541a (2005).

6. Schäfer, A., Horn, H. & Ahlrichs, R. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* **97,** 2571–2577. https://www.doi.org/10.1063/1.463096 (1992).

7. Logan, J. D. *Applied mathematics* 3rd. **14204421,** 529 (Wiley-Interscience, Hoboken, N.J., 2006).

Appendix C

Supporting information for Grafting $TiCl_4$ onto amorphous silica: modeling effects of silanol heterogeneity

C.1 Rate constants used in the population balance model

Figure C.1 shows the rate constants in the grafting pathway used in the population balance model.

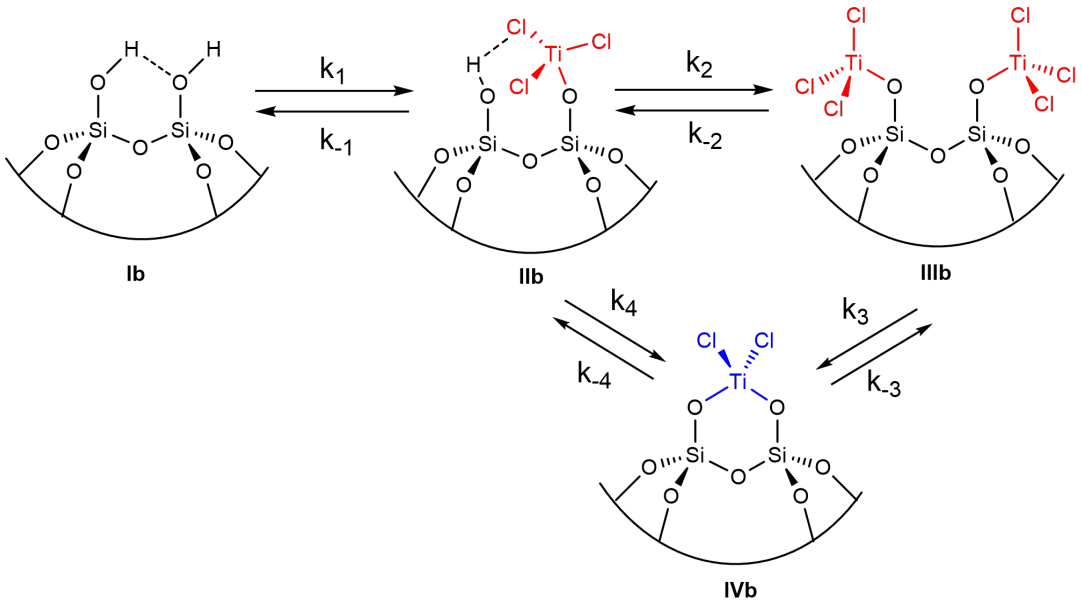


Figure C.1: Rate constants in the grafting pathway.

All rate constants were computed using transition state theory (TST) as

$$k_i(\phi) = \frac{k_B T}{h}\left(\frac{k_B T}{P^\circ}\right)^{v-1} exp\left[\frac{\Delta G_i^\ddagger(\phi)}{k_B T}\right] \tag{C.1}$$

Here, k_B is Boltzmann's constant, h is Planck's constant, $\Delta G_i^\ddagger$ is the free energy barrier of the reaction, ϕ is the vicinal site dihedral angle, P° is the pressure at which $\Delta G_i^\ddagger$ is computed, v is the order of reaction, and T is the temperature. k_3 and k_4 are first order rate constants, while all the other rate constants are second order.

C.2 Free energies as a function of ϕ

Table C.1 shows the free energies of different species calculated on 5 different vicinal sites.

Table C.1: Free energies of species relative to the bare site (**I**) for that dihedral angle.

Species/Dihedral (ϕ)	**0°**	**20°**	**40°**	**56°**	**60°**
I_II TS	77.5	78.4	80.6	80.6	83.6
II	-12.2	-14.9	-14.0	-12.9	-12.1
II_IV TS	67.3	68.3	77.0	96.1	111.0
III	-21.8	-22.0	-22.9	-23.5	-21.4
III_IV TS*	83.6	87.1	-	-	-
IV	-4.9	-2.4	7.9	27.5	41.4
II_III TS	75.5	76.6	77.9	77.1	74.6

*We were not able to find saddle points for **III_IVTS** for $\phi > 20°$. We expect the **III_IVTS** barrier to monotonically increase with ϕ because of increased ring strain. And we anticipate that the **IV**→**III** pathway will become unfavourable at large ϕ and the exact form of the fit function will not be relevant. Hence, we use only two data points to fit a parabolic function.

C.3 Polynomial fits for free energies as a function of the ϕ

Figure C.2 shows polynomial fits for free energies of the different species as a function of ϕ. We only parametrize polynomial models for species whose free energies relative to **Ib** are a strong function of ϕ. We fit a parabolic equation for **III_IVTSb** and a polynomial with a free exponent for all other species (**I_IITSb**, **II_IVTSb**, and **IVb**).

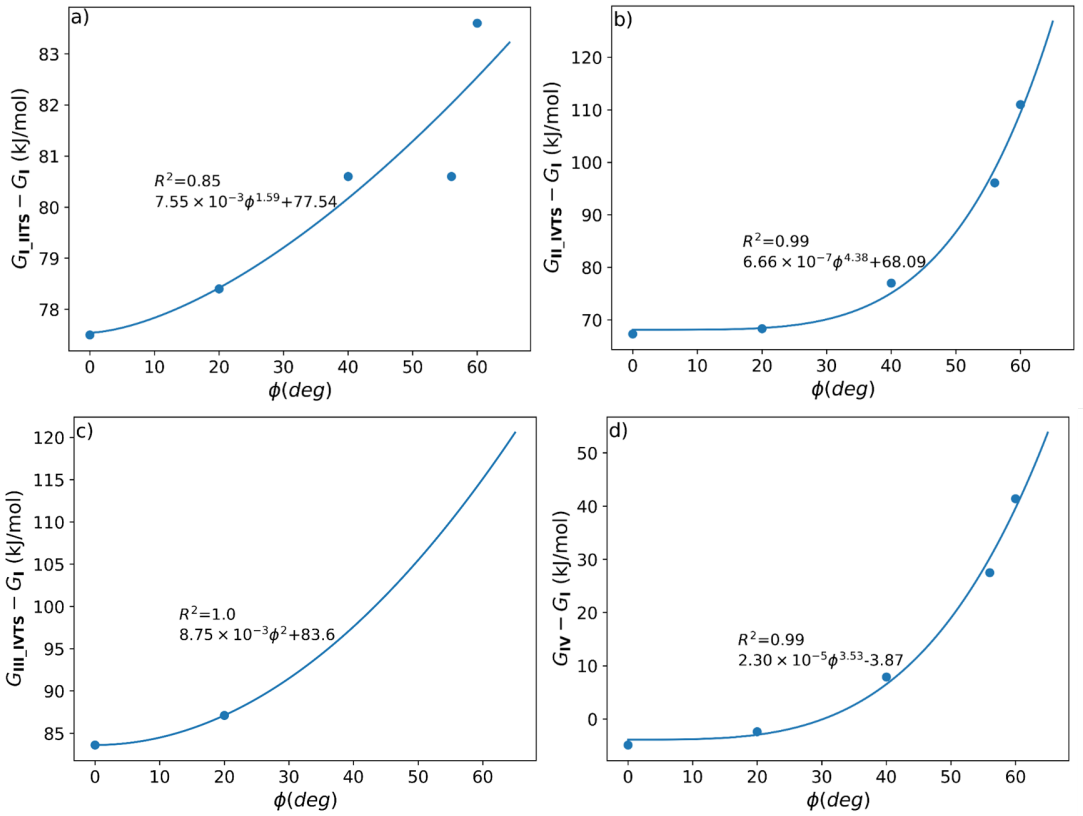


Figure C.2: Polynomial fits of species free energies relative to the bare site (**I**) as a function of ϕ: a) **I_IITSb** b) **II_IVTSb** c) **III_IVTSb** d) **IVb**. The figure also shows the fit polynomial equations and R^2 values of the fit.

C.4 Discretized equation to calculate P_{HCl} at each time step in the simulation

$$P_{HCl} = \frac{\sigma}{2}\frac{amk_BT}{V}\frac{\sum_\psi \rho_{tot}(\psi)[\int_{\psi-w/2}^{\psi+w/2}\theta_{\mathbf{IIb}}(\psi,t)\,d\psi]}{w\sum_\psi \rho_{tot}(\psi)} + \frac{\sigma}{2}\frac{amk_BT}{V}\frac{\sum_\psi \rho_{tot}(\psi)[2\int_{\psi-w/2}^{\psi+w/2}\theta_{\mathbf{IIIb}}(\psi,t)\,d\psi + 2\int_{\psi-w/2}^{\psi+w/2}\theta_{\mathbf{IVb}}(\psi,t)]}{w\sum_\psi \rho_{tot}(\psi)}. \quad \text{(C.2)}$$

Here, ϕ is the vicinal site dihedral angle, t is time, σ is the number density of surface

silanols, a is the area per unit mass of silica, m is the mass of silica, and V is the headspace volume of the reactor. $\theta_{\mathbf{IIb}}$, $\theta_{\mathbf{IIIb}}$, and $\theta_{\mathbf{IVb}}$ are the fractions of **IIb**, **IIIb**, and **IVb**, respectively. $\rho(\psi)$ is the discrete density (non-normalized) of vicinal dihedrals and w is the bin-width used to discretize $\rho(\psi)$.

Appendix D

Supporting information for Are atomistic models of amorphous silica reaslistic?

D.1 Ugliengo slabs

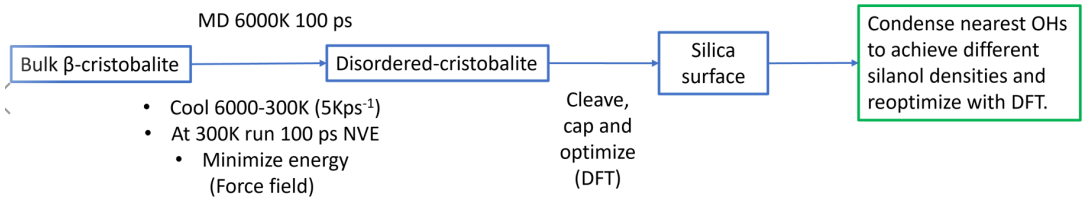


Figure D.1: Schematic outlining generation protocol of Ugliengo slabs.

Table D.1: Types of silanols in models with different silanol densities by Ugliengo

Slab(#silanols/nm^2)	Total	Isolated	Geminal	Vicinal
A(7.2)	28	4	8	17
B(5.4)	20	6	3	9
C(4.5)	16	6	1	6
D(2.4)	8	5	1	2
E(1.5)	4	2	1	0

D.2 Comas-Vives slabs

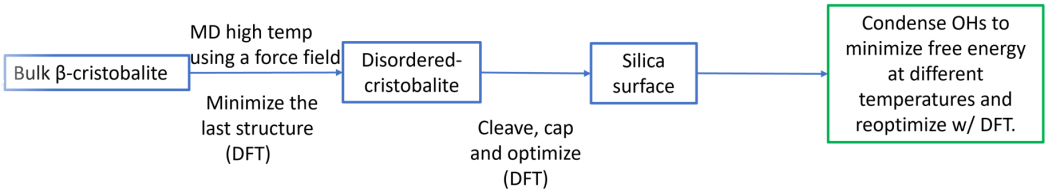


Figure D.2: Schematic outlining generation protocol of Comas-Vives slabs.

Table D.2: Types of silanols in models with different silanol densities by Comas-Vives

Slab(#silanols/nm^2)	Total	Isolated	Geminal	Vicinal
A(7.2)	33	1	11	23
B(5.9)	27	3	7	17
C(4.6)	21	2	3	13
D(3.3)	15	4	1	5
E(2.8)	13	4	0	5
F(2.4)	11	6	0	3
G(2.0)	9	5	0	2
H(1.5)	7	5	0	1
I(1.1)	5	5	0	0

D.3 Tielens slabs

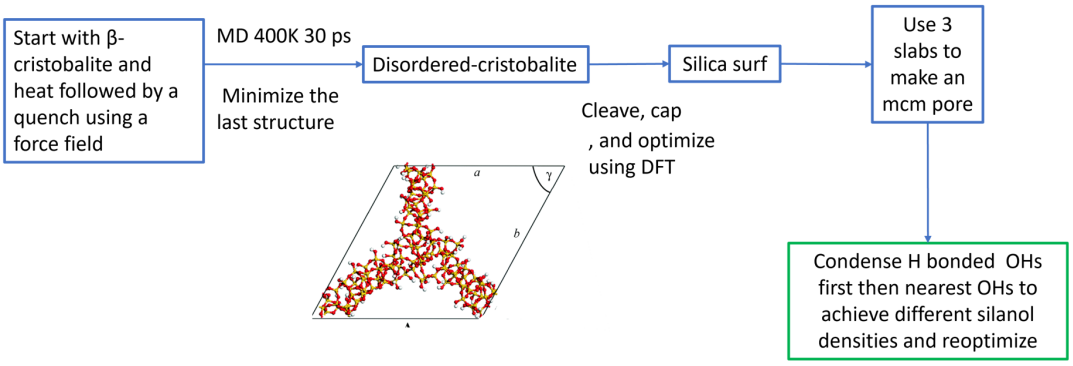


Figure D.3: Schematic outlining generation protocol of Tielens slabs.

Table D.3: Types of silanols in models with different silanol densities by Tielens

Slab(#silanols/nm^2)	**Total**	**Isolated**	**Geminal**	**Vicinal**
A(5.8)	70	7	4	47
C(4.8)	58	9	1	35
D(3.6)	44	16	0	18
E(2.3)	28	13	0	8
F(1.7)	20	16	0	2