

Computationally Guided Searches for Efficient Catalysts Through Chemical/Materials Space: Progress and Outlook

Charles D. Griego,^{†,‡} Alex M. Maldonado,^{†,‡} Lingyan Zhao,[†] Barbaro Zulueta,[†]
Brian M. Gentry,[†] Eli Lipsman,[†] Tae Hoon Choi,[†] and John A. Keith^{*,†}

[†]*Department of Chemical and Petroleum Engineering, University of Pittsburgh, Pittsburgh,
Pennsylvania 15261, United States*

[‡]*Contributed equally to this work*

E-mail: jakeith@pitt.edu

Abstract

Computational quantum chemistry promises to help guide the design of catalysts that are more sustainable and economical. This feature article gives a tutorial overview of how our group accounts for thermodynamics and kinetics of chemical reactions in complex environments. We start with explanations of how to include environmental contributions when modeling homogeneous and heterogeneous catalytic processes. We also provide examples of schemes that use machine learning and alchemical perturbation density functional theory that eschew high computational costs while providing useful insights into chemical reaction mechanisms. With this tool-box of computational methods, we highlight progress in understanding how to reliably model renewable energy catalysis reaction mechanisms that occur in complex environments.

Introduction

Catalysis plays a critical role in society by producing fuels and value-added chemicals,¹ but the continually evolving socioeconomic and environmental landscapes require advances to make catalytic processes more sustainable.²⁻⁴ Science and engineering have traditionally used knowledge, intuition, and ingenuity to guide catalyst design, but we are in an exciting epoch where computational codes and algorithms are becoming faster, more accurate, and better automated. This means that more and more hypothetical catalysts can be computationally screened to find the most promising candidates that warrant consideration in the arduous and costly process of experimental synthesis, testing, and implementation.

In catalysis applications, computational chemistry is wielded to predict quantitative trends in how local chemical bonding and solvent effects influence thermodynamics and kinetics of reaction steps. The required robustness of the modeling ultimately depends on the complexity of the system. The proverbial zoo of multiscale methods, model chemistries (i.e., levels of theory and basis sets) and assorted keywords can make it almost too easy for novices to use a method to obtain a pre-conceived result. But for experts, this zoo represents a gallery of ornate methods that account for different physicochemical interactions within molecules and materials in hierarchical degrees of physical rigor and computational expense.

Depending on the relevant physics of the system, homogeneous reactions might be sufficiently modeled within a vacuum, within a polarizing continuum solvent model, or they may require more complex and locally heterogeneous solvating environment.⁵ Alternatively, heterogeneous mechanisms will occur at an interface of two (or more) different phases, for example, a solid/liquid interface (SEI). If the solid phase here is a conductor, then standard generalized gradient approximation (GGA)-based Kohn-Sham density functional theory (DFT) might be adequate. However, GGA-based functionals also might not be trustworthy, including the case of CO adsorption on transition metal surfaces being imprecisely described by PBE.^{6,7} Additionally, studying the water gas shift reaction (WGSR) on copper with increasingly complex DFT functional can result in different kinetic predictions of WGSR.⁸ In cases

such as these, and if the solid phase in a SEI is a highly correlated semi-conductor, higher level calculations such as DFT+U, other methods from the Jacob’s Ladder of DFT functionals, or one of a variety of different embedded DFT theories may be needed to physically describe the correct state. Additionally, studies of the water gas shift reaction (WGSR) on copper with increasingly complex GGA exchange-correlation functionals can result in varying kinetic predictions of WGSR, including the determination of the surface mechanism. In cases such as these the solid phase in the SEI is a highly correlated semi-conductor, then higher level calculations such as DFT+U,^{9–12} other methods from the Jacob’s Ladder of DFT functionals,¹³ or one of a variety of different embedded DFT theories^{14–22} may be needed to physically describe the correct state. Besides the choice of model chemistry for the electronic structure, the solvent phase could be modeled using one or more well-ordered solvent layers, pseudo-amorphous blobs of solvent molecules (perhaps including ions and counterions), solvent interactions might be treated with implicit models,^{23–26} or neglect solvation altogether. The nature of atoms within the local environment around homogeneous or heterogeneous chemical steps also depend on the outer shell environmental factors: chemical potentials of protons, electrons, electric fields, and any potentially mass-transfer limited species. There admittedly is not a consensus about the ‘correct’ way to model all of these effects. However, whatever the approach chosen, computational models must provide useful insight to advance understanding and guide the design of new and improved technologies.

One of our group’s primary interests is understanding how to best model chemical reactions on a computer and navigate the vastness of chemical and material space for improved design. Even with the advances in computational quantum chemistry (QC) and more efficient calculation platforms,^{27–31} chemical and materials space are so tremendously massive that QC calculations on all possible candidates should be considered intractable for quite some time. With catalysts, the vastness of chemical space arises from different hypothetical atom configurations at active sites, their surrounding ligands and/or degrees of strain, and environmental aspects such as solvation, pH, and ionic strength. Thus, there is recent interest in

developing and applying cost-efficient and approximate methods based on machine learning (ML) and/or other physical theories.³² The validity of these methods can be assessed with straightforward *fundamental application studies* that consider metrics for their reliability in catalysis applications: e.g., predicted binding energies of reaction intermediates, activation energies of chemical reactions, or thermodynamic descriptors such as acidity constants and standard redox potentials. Below, we discuss our progress in understanding how to model these factors.

Materials screening with alchemy

In 2011, the United States launched the Materials Genome Initiative (MGI) to create new infrastructure for rapid computational prediction and screening of novel materials.^{33,34} Catalysis researchers have followed suit and pursued similar efforts with the catalyst genome.³⁵ The main target of interest has been the atomically precise nature of the active site; which is characterized in part by the local coordination of atoms within the surface facet. Useful descriptors—such as the reacting adsorbate binding energy (BE) to a surface site and reaction step energetics—provide a straightforward way to assess kinetics of heterogeneous catalyzed reactions. Indeed, there are concerted efforts for developing databases of these properties such as Citrination,³⁶ Catalysis-Hub,³⁷ and the Material’s Project³³ to name a few.

However, with more promising classes of materials being discovered experimentally and computationally, the community must adopt procedures suitable for systematic studies. GGA methods in general are suitable for many systems, but some require higher levels of theory that bottleneck high-throughput screening. Our group has been investigating computationally efficient alchemical perturbation DFT (APDFT) approaches that leverage a small set of QC calculations to create much more useful data without sacrificing much accuracy.^{38–40}

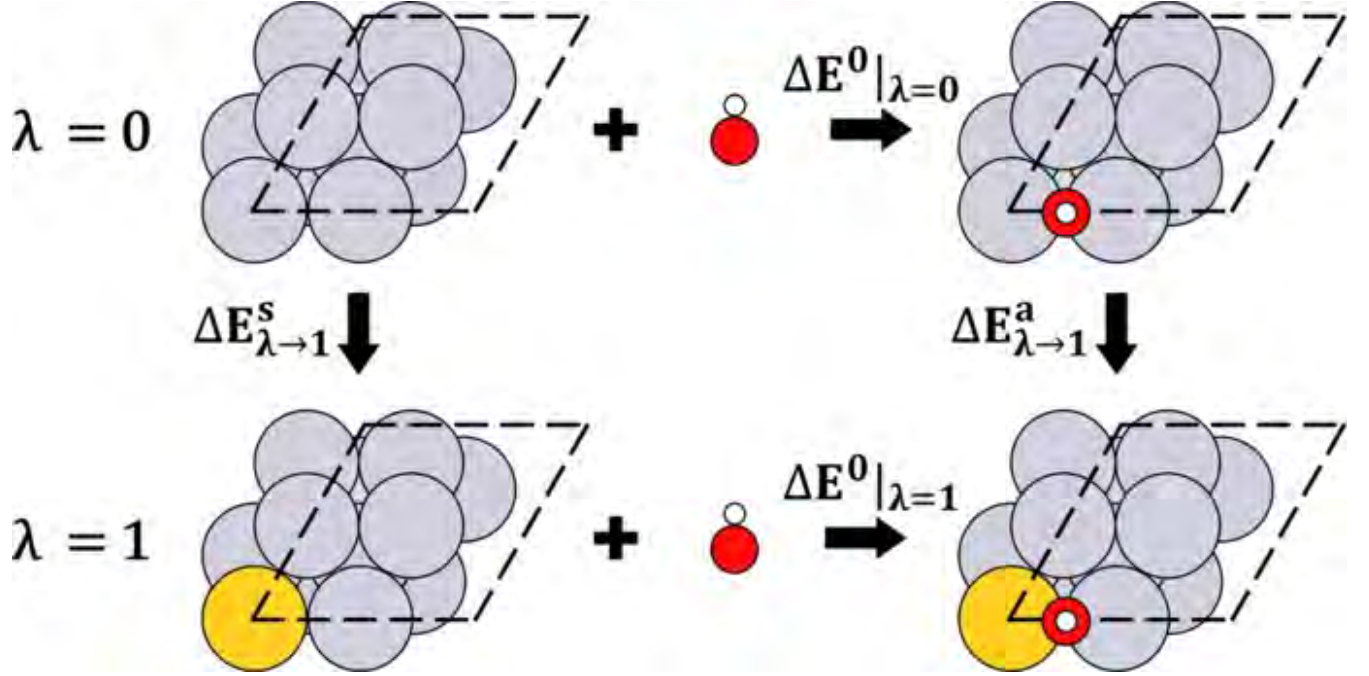


Figure 1: Illustration of a thermodynamic cycle that depicts the energetic pathways of adsorption, atomic transmutations, and how these transmutations impact the adsorption. We commonly refer to this type of cycle when considering APDFT predictions of BEs on alloy surfaces that were hypothesized from a reference material. Pathways are composed of the BEs of an adsorbate on a surface (horizontal legs) and atomic transmutations (vertical legs). $\Delta E^0|_{\lambda=0}$ and $\Delta E^0|_{\lambda=1}$ denote the BEs for the top and bottom horizontal legs, respectively. $\Delta E_{\lambda \rightarrow 1}^s$ and $\Delta E_{\lambda \rightarrow 1}^a$ denote the energy change associated with the atomic transmutation for the left (s = surface) and right (a = ads-site) vertical legs, respectively. **Reprinted with permission from Ref. 41. Copyright 2020 John Wiley & Sons.**

In theory, APDFT provides many adsorbate binding energies by an approximated relationship of how electrostatic potentials in a reference adsorbate-catalyst system and the BE change (ΔBE) upon a compositional change i.e., an alchemical transmutation. This procedure is illustrated in Figure 1, where a thermodynamic cycle depicts the energetic changes of adsorbate binding to a surface after an atomic transmutation. The hypothetical energy contribution arising from this alchemical transmutation is approximated as a Taylor series

$$\Delta E^0|_{\lambda=1} = \Delta E^0|_{\lambda=0} + \partial_{\lambda} \Delta E^0 \Delta \lambda + \frac{1}{2} \partial_{\lambda}^2 \Delta E^0 \Delta \lambda^2 + \dots, \quad (1)$$

where $\Delta E^0|_{\lambda=1}$ is the energy of the adsorbate binding on the hypothetical system result-

ing from the alchemical transmutation. This is equal to the BE on the reference system ($\Delta E^0|_{\lambda=0}$) plus alchemical derivative terms that are written as $\partial_\lambda^n \Delta E^0$, where n is the order of the derivative, and $\Delta\lambda = 1$.

In past applications of APDFT,^{40–56} this expression is typically truncated to just the first order derivative to approximate the change in BE between the two states as

$$\Delta BE = \Delta E^0|_{\lambda=1} - \Delta E^0|_{\lambda=0} = \Delta E_{\lambda \rightarrow 1}^a - \Delta E_{\lambda \rightarrow 1}^s = \partial_\lambda \Delta E^0 \Delta\lambda, \quad (2)$$

where $\Delta E_{\lambda \rightarrow 1}^s$ and $\Delta E_{\lambda \rightarrow 1}^a$ are the energy changes associated with an alchemical transmutation done to a bare surface model (s) and a surface model with an adsorbate (a), respectively. When the alchemical transmutation is made, the nuclear charge of an atom (N_I) is altered by an integer amount (ΔZ), resulting in an energy change equal to the energy gradients of nuclear chemical potential ($\Delta\mu_{nI}$) with respect to this variation in N_I . When alchemical transmutations are done isoelectronically (number of electrons in the system is conserved) and the atomic positions remain the same, ΔBE is equal to a simplified first order derivative:

$$\Delta BE = \partial_\lambda \Delta E^0 \Delta\lambda = \sum_I \Delta\mu_{nI} \partial_\lambda N_I. \quad (3)$$

With this approximation, simple arithmetic manipulations involving electrostatic potentials of a reference catalyst model are used to predict BEs of hypothetical materials with minimal computational cost. For a more detailed explanation of first order approximations, we direct the reader to our recent article, which offers hands-on resources that allow users to perform APDFT analyses with Jupyter Notebooks.⁴¹ Here, we overview our recent work with APDFT for heterogeneous catalyst and offer our perspective on future implications of this method.

In our first published work on APDFT, we showed that first order approximations are quite accurate for high-throughput binding energy predictions for doped active sites in transition metal surfaces.⁵⁴ Using oxygen reduction reaction intermediates binding on hypothetical alloys of Pt, Pd, and Ni, we benchmarked APDFT predicted binding energies against DFT

predicted values and found that these estimates agreed within 0.1 eV.

In a study where we applied APDFT to BE predictions on carbides, nitrides, and oxides, we found that APDFT agrees with DFT predictions within 0.33 eV for rocksalt TiC(111), TiN(100), and TiO(100) materials, which do not exhibit a bandgap.⁵⁵ Conversely, we found that APDFT has significant shortcomings with BE predictions on materials based on semi-conductors like rutile TiO₂(110), rutile SnO₂(110), and rocksalt ZnO(100). Our hypothesis was that first order corrections using APDFT were benefited by error cancellation present in conductive systems that are subjected to electronic screening. We tested this hypothesis adding Pt dopants to surface layers in TiO₂, which decreased the reference material’s band gap and we found higher accuracy was achieved with those APDFT predictions. We continue to search for a more precise physical explanations for why APDFT is challenged by these systems so that it can be used more generally in semi-conducting and insulating systems.

We are also interested in understanding the kinetics of reactions on surfaces that are dictated by the energy barrier (E_a) between two reaction steps. A standard approach for predicting E_a for a surface-bound reaction is to employ the nudged elastic band (NEB) algorithm⁵⁷ that interpolates images between the initial and final states of a reaction step. This procedure does not require a Hessian, but the electronic energies and forces on all images are calculated, and thus making it moderately computationally expensive. The expense for each NEB calculation makes it challenging to use this approach for determining many hypothetical E_a s for different elementary processes or for systems involving different surface atoms. Just as when calculating a BE, the number of required calculations for each NEB pathway will linearly scale with the number of barriers we want to predict.

We tackled this issue by using one NEB calculation to establish an elementary reaction pathway as a set of reference data and then APDFT to generate approximate many NEB pathways based on this reference.⁴¹ Using the same thermodynamic cycle scheme to approximate the binding energy change following an alchemical transmutation (shown in Figure 1), we approximated the energy change for alchemically transmuted target systems based

on each image from a traditional NEB calculation for CH_4 dehydrogenation on a surface of Pt(111) that contained 10 images. From this, we used APDFT to generate approximate energy profiles of this pathway on 32 transmuted variations of the surface (Figure 2). With these profiles we also predicted E_a for all 32 pathways that agreed with DFT calculations within 0.3 eV.

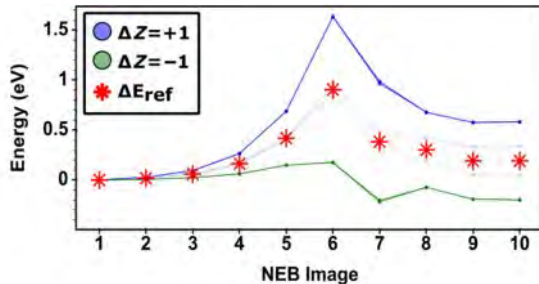


Figure 2: Energy profiles for the CH_4 dehydrogenation mechanism on hypothetical alloys of Pt. Here we compare the reaction energy pathway on the original Pt(111) reference surface, which was calculated with DFT, and the APDFT predicted reaction energies of the same pathway taken over 32 transmuted variations of the reference surface. The reference pathway on pure Pt is denoted with red asterisks. Pathways with the most significant effect from a transmutation with a nuclear charge change $\Delta Z = +1$ is shown in dark blue, pathways with the most significant effect from a transmutation with a nuclear charge change $\Delta Z = -1$ is shown in dark green, and other reaction pathways computed with APDFT are shown in a less visible light blue/green. Due to energy profiles being very similar, there is significant overlap for many alloys. Reprinted with permission from Ref. 41. Copyright 2020 John Wiley & Sons.

We have also investigated procedures to buttress APDFT models using Δ -ML procedures,⁵⁸ showing that the computational effort in developing ML models is justified when they are valid for data sets magnitudes larger than data sets used for training.⁵⁶ By ranking the relative importance of the input features to our models, we identified the variables that contributed most to low accuracy APDFT BE predictions on certain target alloys, the size of the transmutation (ΔZ), and the number of transmutions. Based on the work of von Lilienfeld and coworkers demonstrating accurate screening of deprotonation energies using APDFT with third order corrections,⁵⁹ we expect that shortcomings we observe with first order corrections would likely be treatable with higher order corrections and thus allow accurate screening of more target alloys made with multiple transmutions of larger ΔZ .

Using third order alchemical derivatives, a symmetric finite differences procedure that requires $1 + 2N + N^2$ single point energy calculations, where N is the number of sites that could be transmuted. For a catalyst surface slab model with 16 transmutable sites, one would only need 289 single point energies to accurately screen thousands of adsorbate BE on hypothetical alloys. We also anticipate the possibility of higher order corrections treating the shortcomings we observed with semiconducting systems. With second and third order energy derivatives, systems that are not subjected to screening effects may be better described, and these derivatives may be able to accurately point us in the direction of more complex materials with higher catalytic activity without a large reliance on other more expensive methods like DFT+U or using more complex exchange correlation functionals. Finally, in other ongoing work, we are investigating how APDFT procedures would be related to other traditional theoretical models for understanding catalyst descriptors such as Newns-Anderson Hamiltonian methods.⁶⁰

Since APDFT brings high computational efficiency, we foresee it making a transformative change in how to alleviate the necessity of running many calculations to understanding stable states of catalysts under ambient reaction conditions across chemical space. Catalyst descriptors are normally modeled assuming ideal conditions on an idealized surface facet, but in reality, materials may succumb to defects, alloy segregations,^{61,62} reconstruction, or non-innocent phase reconstruction when subjected to a reaction environment. In future work, we will tackle these challenges with APDFT schemes coupled with other methods, described in the following section, that our group has applied to model ambient catalyst states.

Modeling catalysts under ambient conditions

Combining the computational hydrogen electrode model⁶³ with atomistic thermodynamics⁶⁴⁻⁶⁸ allows one to construct electrochemical phase diagrams (including Pourbaix diagrams) that identify stable resting states of catalysts under ambient reaction conditions.

We showed that purported catalysts for CO₂ reduction have different and complementary accessible states that would facilitate energetically efficient shuttling of multiple protons and electrons.^{67,69} The procedures outlined there would be useful for identifying and understanding bio-inspired orthogonal hydride transfers that have been identified as a promising catalyst design strategy.⁷⁰

We also showed, in a collaboration with Snyder’s group at Drexel University, that accounting for multiple thermodynamically accessible states can help deconvolute complex experimental observations.⁷¹ Nanoporous PdX skin alloys (X = Co, Ni, Cu, and Ag) were presented as electrocatalysts that produced formate from CO₂ with high selectivity and avoided deactivation from CO poisoning. Among these alloys, Pd-skin/Pd₃Co was found to be most promising, which our group theoretically explained by evaluating stable configurations of CO and H binding on the surface with DFT. Moreover, we had to computationally reconcile both the destabilization of CO (poisoning tolerance) and H (facilitating CO₂ hydrogenation to formate) on these surfaces. BE calculations with co-adsorbed species were run (shown in Figure 3), and results showed that adsorbed CO and H were both the most destabilized on Pd-skin/Pd₃Co, regardless of the co-adsorbed species. Again, simple calculation models for understanding catalyst activity would be significantly amplified by using accurate APDFT modeling.

In other work, we used electrochemical phase diagrams to understand which molecular and material states would be relevant under different electrochemical reaction conditions. In one study,⁶⁷ electrochemical phase diagrams illustrated the stability of different intermediates on SnO₂ (110) surface as a function of the thermodynamic driving force for CO₂ binding to a surface (i.e., $\Delta\mu_{\text{CO}_2}$) and electrochemical potential (i.e., U). We plotted and overlaid the theoretical Pourbaix diagram over the experimental Pourbaix diagram for CO₂ intermediates in aqueous phase to show the proximity of boundaries between them. This led us to propose reaction intermediates that would explain experimentally observed overpotentials potentials for CO₂ reduction at maximum Faradaic efficiency.

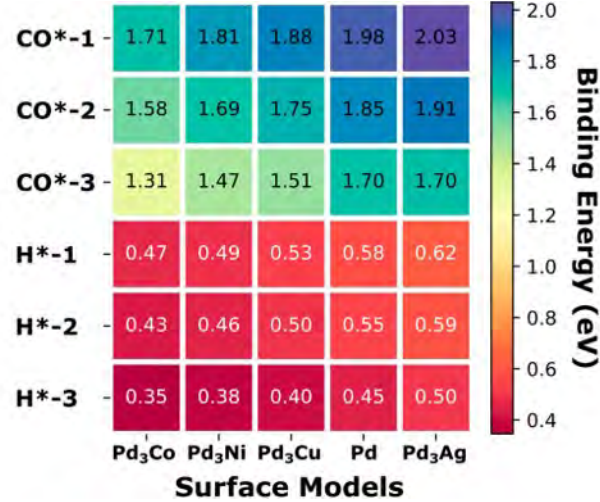
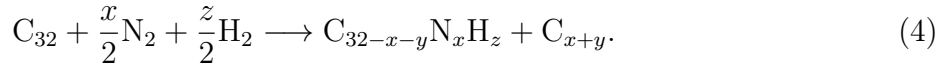


Figure 3: DFT calculations considered the interplay of co-adsorbed species to rationalize experimental observations of nanoporous PdX skin alloys as CO poisoning tolerant electrocatalysts for CO₂ reduction to formate. Tabulated above are binding energies (in eV) on Pd-skinned Pd₃X alloys (X = Co, Ni, Cu, Ag) for CO binding to a clean surface (CO*-1), CO binding to a surface with H* (CO*-2), CO binding to a surface with 2 H* (CO*-3), H binding to a clean surface (H*-1), H binding to a surface with H* (H*-2), and H binding to a surface with CO* (H*-3). Reprinted with permission from Ref. 71. Copyright 2019 American Chemical Society.

To create these diagrams, we first calculated the free energy change for different reactions in the system. We refer back to our work on the electroreduction of CO₂ on N-doped graphene⁶⁹ as an example derivation. Considering the following chemical formula for reducing a graphene basal planes (GBP) in aqueous environment according to atomistic thermodynamics follows:



(Note that C₃₂ refers to the clean GBP model.) We then calculated the Gibbs energy ($G = E^{\text{DFT}} + G^{\text{vib}}$) of each species in the equation with the electronic energy from DFT (E^{DFT}) and G^{vib} , which is comprised of the energy correction from zero point energy (E_{ZPE}) and vibrational entropy (TS_{vib}). The free energy change of the reaction as shown in Equation 4 is written below, where the chemical potentials (μ) of atomic species are included:

$$\Delta G = G(C_{32-x-y}N_xH_z) - G(C_{32}) + (x+y)\mu(C) - x\mu(N) - z\mu(H). \quad (5)$$

Next, we could reference this energy with the Standard Hydrogen Electrode (SHE), and finally, we get the free energy expression for the GBP reaction as a function of pH and applied potential (U) below, where $\Delta\mu_N$ is the thermodynamic driving force for N binding:

$$\begin{aligned} \Delta G = & E_{C_{32-x-y}N_xH_z}^{\text{DFT}} - E_{C_{32}}^{\text{DFT}} + \Delta G^{\text{vib}} - x \left(\frac{1}{2} E_{N_2} + \Delta\mu_N \right) \\ & + (x+y) E_C - z \left(\frac{1}{2} \mu(H_2) - 2.303 k_B T \text{pH} - U \right) \end{aligned} \quad (6)$$

To plot an electrochemical phase diagram, the free energy of each intermediate is calculated with Equation 6 assuming a constant pH environment, and this gives a relation between ΔG and $\Delta\mu_N$ for each intermediate. Next, ΔG is converted into a reduction potential E^{red} with $E^{\text{red}} = -\Delta G / (nF)$ to have a relation between reduction potential and thermodynamic driving force, where n is the number of electrons transferred and F is Faraday's constant. Finally, we could plot the boundary lines of each intermediate and find the most thermodynamically stable state at different voltages and $\Delta\mu_N$. Plotting a Pourbaix diagram can be done with a similar procedure, where we set $\Delta\mu_N$ as zero and find out the relationship between U and pH for each intermediates. With that, we determine the most stable intermediate at different U and pH to get the final Pourbaix diagram.

We believe these atomistic thermodynamics schemes should be and will be used more often for theoretically complete catalyst screening studies. For example, there are excellent opportunities for combining schemes for identifying thermodynamically relevant structures using high-throughput generations of microkinetic mechanism parameters,^{72,73} active site ensembles of metastable states,^{74,75} and models based on active site coordination numbers.⁷⁶⁻⁷⁸ This is not done now because it would require very large numbers of QC calculations, but this would also be alleviated with new advances in APDFT and ML approaches.

Toward mechanistic understanding

Any given chemical species may undergo countless reaction mechanisms. Hydrogenations are one such example, and Figure 4 shows 23 hypothetical steps (consisting of covalent hydrogen atom transfers, stepwise or coupled proton and electron transfers, and formal hydride transfers) that might warrant consideration when computationally analyzing an arbitrary reduction starting with species A (or an oxidation ending with species A).

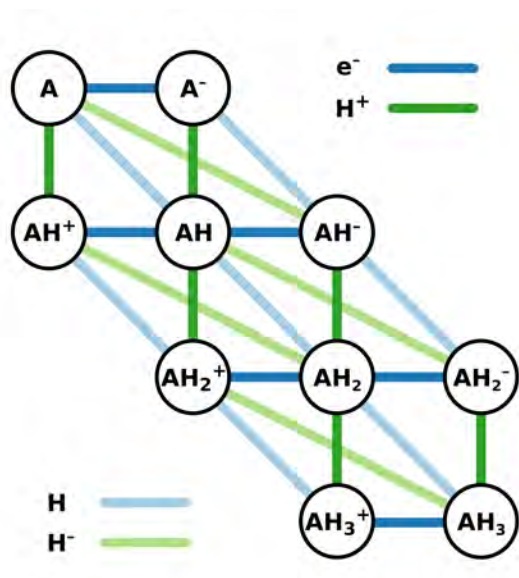


Figure 4: A general hydrogenation of A can be considered as a series of elementary electron and proton transfers, proton coupled electron transfers, or formal hydride transfers, and energetics of each pathway may vary depending on the environment.

For example, suppose a catalyst were experimentally found to reduce CO₂ (A in Figure 4), but its actual mechanism was unclear, and computational theory was needed for insight. From standard QC calculations, the most favorable state of CO₂ and the catalyst with respect to pH and applied potential (a Pourbaix diagram) can be determined as shown in Figure 5. The boundary lines of the calculated Pourbaix diagrams show at which conditions of applied potential and pH species on opposite sides of the boundary have the same chemical potential (i.e., when a reaction from one to the other would bring $\Delta G = 0$). This is an important point of reference for understanding reaction mechanisms, but reaction barriers would also

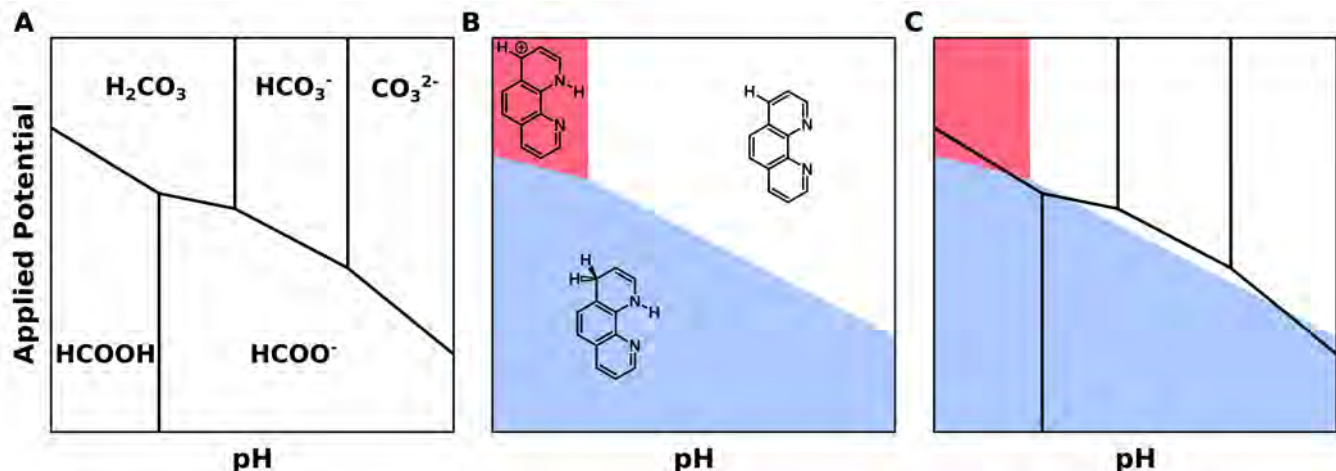


Figure 5: (A) Pourbaix diagram showing stable states of the reactant, CO_2 ; (B) Pourbaix diagram showing stable states of a hypothetical molecular catalyst, 1,10-phenanthroline; (C) overlaid Pourbaix diagrams from (A) and (B) showing similar boundaries for hydrogen shuttling and CO_2 reduction. Vertical lines represent $\text{p}K_a$ s, the horizontal lines represent the pH-independent standard redox potentials and the diagonal lines represent the pH-dependent proton-coupled electron transfer steps. Adapted with permission from Ref. 79. Copyright 2019 Wiley Periodicals, Inc.

need to be introduced for modeling reaction kinetics. Since these reactions very likely can involve the participation of solvent molecules, we now turn to how our group approaches solvated systems.

Modeling solvent environments

We consider three main classifications of solvent models: implicit (or continuum), mixed implicit/explicit (or cluster-continuum), and explicit. All bring different strengths and weaknesses, but useful insights can be learned by comparing results from different models. Implicit models is a broad category, but generally refers to any solvent models that does not explicitly include solvent molecules as shown in Figure 6A. Many are parameterized to predict solvation energies based on a homogeneous dielectric medium interacting with the solute.^{80,81} Other implicit models use more complicated solute-solvent and/or ensemble descriptions. The conductor-like screening model for real solvents (COSMO-RS), for example, uses QC

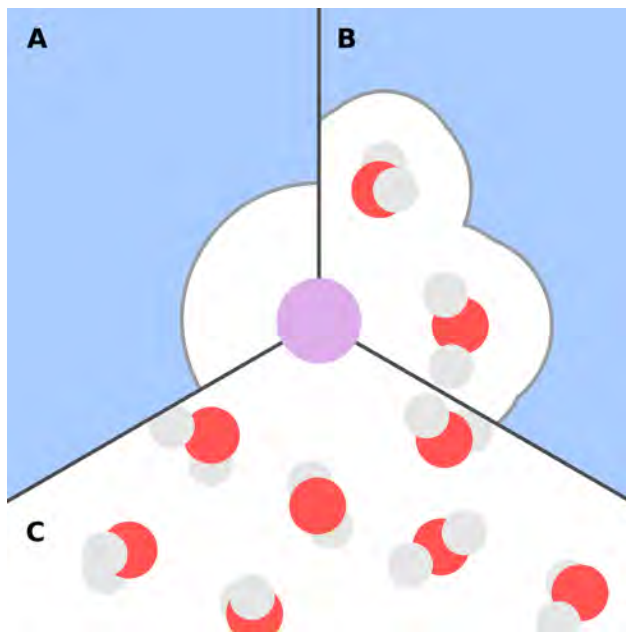


Figure 6: Various solvent modeling techniques are illustrated on a fictitious solute (purple) solvated in water. **(A)** Local and bulk solvent interactions can be treated with an implicit solvent model; a fictitious solute cavity is shown. **(B)** Local solvent effects are captured with explicit solvent molecules. The less important bulk contributions are efficiently described with an implicit model. **(C)** The entire solvent is modeled explicitly.

to describe the solute polarization of a conductor’s cavity surface and statistical thermodynamics to compute the molecular chemical potential.⁸² The reference interaction site model (RISM) models correlation functions of solute and solvent molecular sites to compute the solvent distributions and thermodynamics.⁸³ **This is not an exhaustive list, and there are many more implicit models and variations that each have their own approximations, formulations, and applications.**

The well-known problem of implicit models is their respective approximations and parameterizations that dictate their accuracy, reliability, and transferability. By including some explicit solvent molecules with the solute in QC calculations the solvent is essentially separated into local (inner) and bulk (outer) contributions.⁸⁴ This technique is commonly called mixed implicit/explicit or cluster–continuum modeling (Figure 6B). Quasi-chemical theory (QCT) is a physically rigorous way to separate solvation free energies into statistical contributions and is thoroughly explained elsewhere.^{85–89} Essentially, the excess chemical

potential (i.e., molar solvation free energy) of some arbitrary solute, X, in a pure solvent, L, is expressed in QCT as

$$\mu_X^{(\text{ex})} = -k_B T \ln(K_n^{(0)} \rho^n) + k_B T \ln p_X(n) + \left[\mu_{L_n X}^{(\text{ex})} - n \mu_L^{(\text{ex})} \right]. \quad (7)$$

The terms in Equation 7 can be conceptually described as desolvating n individual solvent molecules ($-n \mu_L^{(\text{ex})}$), associating the solute and solvent molecules ($nL + X \rightleftharpoons (L)_n X$) into a cavity with a solvent density ρ and equilibrium constant $K_n^{(0)}$ in the ideal-gas phase ($-k_B T \ln K_n^{(0)} \rho^n$), solvating the associated solute-solvent cluster ($\mu_{L_n X}^{(\text{ex})}$), then releasing the geometric constraint inside the cavity ($k_B T \ln p_X(n)$) where $p_X(n)$ is the probability of observing n ligands in solution within a predefined region.

Various contributions in Equation 7, specifically $\ln p_X(n)$, require *a priori* information of solvent coordination numbers or dynamic simulations to explicitly quantify them. Alternatively, one could use preformed solvent clusters as proposed by Bryantsev et al.⁹⁰ This cluster thermodynamic cycle can be thought of as applying QCT to both the solvent and solute-solvent clusters to provide error cancellation allowing one to forgo *a priori* information requirements and dynamic simulations. The challenge with these approaches, however, is determining the quantity and configurations of solvent molecules to be used. This would involve molecular simulations or comparisons to experiment.

To tackle this challenge, we employed an unsupervised ML procedure to identify solute-solvent clusters that result in single-ion solvation energies that appear to converge toward values in reasonable agreement to experimental data.⁹¹ The smooth overlap of atomic positions (SOAP)⁹² representation with sketch-map⁹³ dimensionality reduction was used to quantitatively compare structures from an automated, multi-step filtering procedure of candidate solute-solvent clusters of different sizes. We then compared the closeness, or distance between points, to determine the local solvent environment similarity of the solute-solvent structures. Once larger solute-solvent clusters overlap with smaller ones on the sketch-map

one can assume the additional solvent molecules are far enough away to have minimal impact on the local solvent environment.

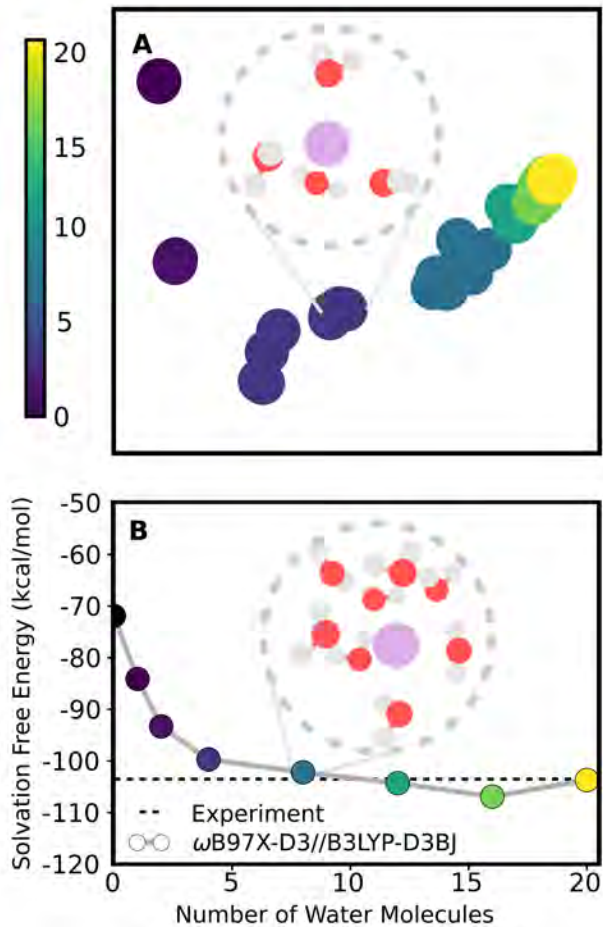


Figure 7: (A) SOAP/sketch-map representations of solute-solvent clusters containing Na^+ with various numbers of water molecules. An example of a cluster with four water molecules is shown. The color bar represents the number of water molecules in the cluster. (B) Boltzmann-weighted average of solvation free energies of Na^+ with the variable number of water ligands from B3LYP-D3BJ/def2-SVP geometries and $\omega\text{B97X-D3/def2-TZVP}$ energies. An example of a cluster with eight water molecules is shown. Adapted with permission from Ref. 5. Copyright 2020 AIP Publishing.

Figure 7 shows the case of Na^+ hydration. A global optimization code, ABCluster,⁹⁴ was used to generate hundreds of solvent and solute-solvent clusters. The five lowest-energy structures were then optimized using a relatively inexpensive QC method: BP86-D3BJ/def2-SVP, and the solute-solvent clusters were mapped using SOAP/sketch-map as shown in Figure 7A. One dot represents a single solute-solvent structure (an example case is given for four water

molecules). A Boltzmann-weighted average of the clusters were used to calculate the solvation free energy of Na^+ in water. Based on our procedure, the most reliable structures came from 12-water clusters, and data closely agreed with the experimental solvation free energy. Our approach has also been successfully demonstrated on ion solvation free energies spanning 2- to 2+ charges.⁹¹ Reiher and coworkers have recently developed a similar computational implementation that focuses on rigorously achieving statistically relevant ensembles of local solvent molecules in an automated manner.⁹⁵ Their approach is an improved implementation compared to what we published previously, but it does not make an explicit connection to QCT, which is a useful way to reduce computational expense by leveraging implicit solvent models. Even though their approach has not yet been tested for single ion solvation energy predictions, the similarities of our and their approaches indicates that theirs should be useful and accurate as well.

A key point to reiterate is that cluster-continuum modeling, when done correctly, can be useful and cost-effective calculation scheme for property predictions and mechanistic studies in solvents.^{84,96,97} However, the dynamics of the solvent are often crucial for accurate predictions⁹⁸ and when implicit or mixed implicit/explicit models are not sufficient, they should be modeled with explicit solvent modeling (Figure 6C) such as Born-Oppenheimer molecular dynamics (BOMD)⁹⁹⁻¹⁰¹ or quantum mechanics/molecular mechanics (QM/MM).¹⁰²⁻¹⁰⁴ Both methods offer unparalleled assessment of the vast configurational space observable in reactions.

Accounting for radial and spatial distribution functions (SDFs) in explicit solvation can provide molecular insights into solvent effects of reactions in electrolyte solutions. Figure 8 shows SDFs of our group’s BOMD simulations in two different solvent environments and the effects of high base concentrations. In pure H_2O (Figure 8a), the oxygen distribution in the first solvation shell forms a spherical cage around BH_4^- . Moreover, hydrogen has some distributions inside this oxygen shell that results in the formation of dihydrogen bonds between BH_4^- and H_2O . However, in 7 mol/L NaOH, the oxygen distribution has no sig-

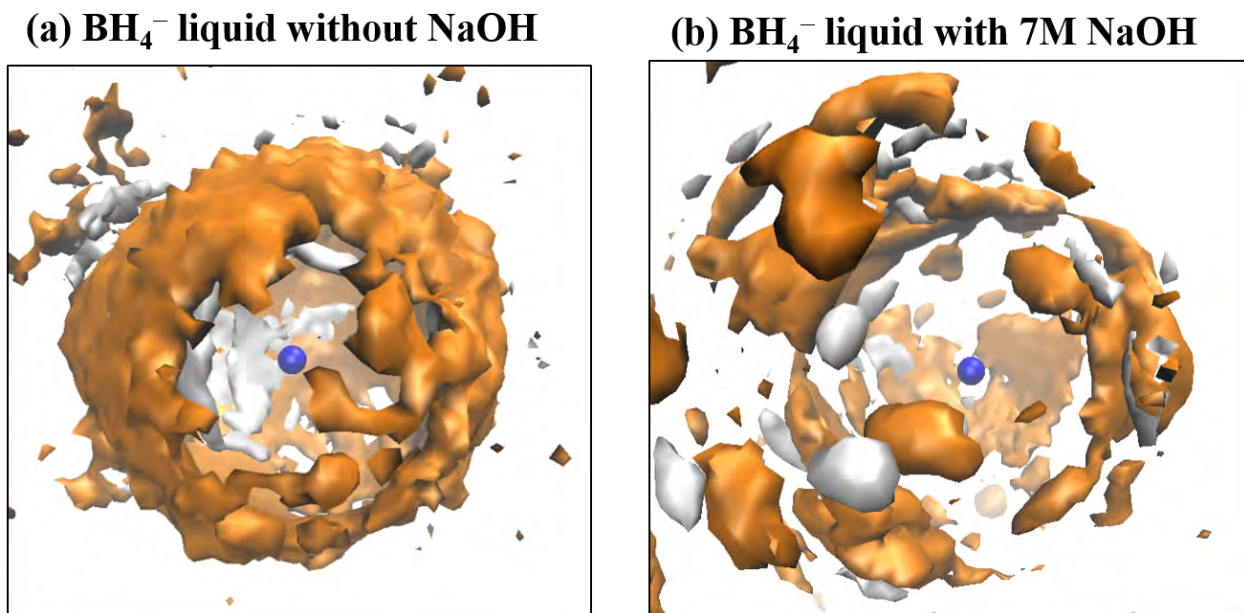


Figure 8: Computational insight into the local solvent environment of BH_4^- liquid simulations of (a) pure H_2O and (b) 7 mol/L NaOH (isovalue of 48 nm^{-3}). Oxygen and hydrogen distributions in the three-dimensional spatial distribution functions (SDFs) are described with orange and white colors, respectively. The central blue sphere represents the boron atom.

nificant spherical shape. Furthermore, the hydrogen distribution inside of the oxygen shell has a smaller distribution range compared to that in the BH_4^- liquid simulations. These observations indicate that increased concentrations of NaOH result in weaker interactions between H_2O and BH_4^- , and this may have ramifications in reaction mechanism and kinetics studies in realistic electrolyte environments.

Solvated reaction mechanisms

Consider again the hydrogenation reaction network in Figure 4. While a local environment can influence which states are thermodynamically favorable, each fundamental reaction step may have a significant energy barrier to consider. Experimental observations (or strong chemical intuition) would help clarify which pathways would be likely; however, in the absence of these, QC explorations can rigorously discern pathways to find which would be most likely.

Next, we will overview our approach and general guidelines for modeling solvated reaction mechanisms with minimal user bias.

BOMD simulations have proven to be a powerful tool for analyzing solvated reaction mechanisms,^{105,106} but such methods are usually costly and make them used less often. Modeling using mixed implicit/explicit modeling with chain-of-states methods such as nudged elastic band (NEB) or growing string method (GSM) can be done on smaller systems and these will be more amenable to higher levels of theory. On the other hand, this benefit of lower computational cost comes with its own complexities. For example, we computationally investigated aqueous sodium borohydride (NaBH_4) reduction of carbon dioxide (CO_2) to formate (HCOO^-) via a hydride (H^-) transfer using several different modeling approaches.^{107,108} Hydride transfers are charge migrations, and one might expect these to be sensitive to the local solvent environment. As such, it might be expected to be computationally demanding to modeling this kind of system using explicit solvent molecules. The reaction pathway energetics using this modeling scheme is shown in Figure 9 (labeled “Explicit PMF”).¹⁰⁸ Since this approach involves the least empiricism, this is the reaction pathway all other less expensive computational schemes would ideally reproduce.

Using the same collective variable as the explicit PMF we identified a static pathway using the generalized solid-state NEB (g-SSNEB)¹⁰⁹ method while keeping the explicit solvent modeled with QC.¹⁰⁷ This pathway, “Explicit NEB” in Figure 9, is essentially the same as the PMF without any free energy (entropic) contributions. The quantitative similarity in energy profiles shows that entropies along this pathway do not significantly impact the barrier for this reaction. This is good to know since avoiding dynamics simulations would significantly lower the computational costs for mechanisms studies. For example, instead of full dynamics in an explicit solvent, one might treat the bulk contributions (i.e., everything but the first solvent shell) using an implicit model. The resulting reaction pathways are shown in Figure 9. We found that cluster-continuum modeling was actually suitable for qualitatively capturing the metastable intermediate, second transition state, and the prod-

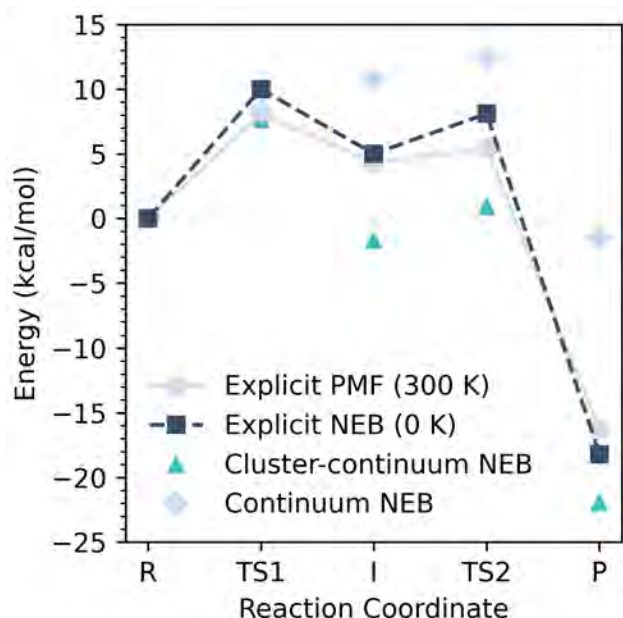


Figure 9: Free energy differences and comparison of continuum and cluster-continuum solvent modeling of a g-SSNEB pathway of sodium borohydride reduction of carbon dioxide. Different subsystems were crafted from an explicit g-SSNEB pathway containing 70 water molecules with an implicit solvent model replacing the removed explicit water molecules.¹⁰⁷ The continuum solvent model was unable to capture the qualitative trend of the g-SSNEB pathway unless the counterion and first solvent shell was included. A BOMD PMF calculation at 300 K quantified the free energy differences not included in the g-SSNEB energies.¹⁰⁸

uct. The counterion was also crucial for describing the reaction mechanisms with the static calculations. As expected, static calculations benefit from cluster-continuum modeling when a full solvent shell is modeled.

Modeling multi-step reactions brings additional complexities as well. Plata and Singleton demonstrated that implicit solvent models often fall short of reliably capturing local solvent effects of the five-step Morita Baylis–Hillman (MBH) reaction.¹¹⁰ Liu et al. illustrated that explicit methods with free energy perturbation solvation treatments on top of high-level wave function methods (i.e., DLPNO-CCSD(T)) can accurately reproduce reaction energetics from kinetics experiments.¹¹¹ Their work showed that predictions of complicated reactions are possible with significant computational resources; something that is not always available nor recommended before properly vetting collective variables using less intensive methods.

Cluster-continuum modeling of the borohydride mechanism above relied on taking struc-

tures out of a periodic, explicitly solvated chain-of-state calculation derived from QM BOMD simulations. The multiple steps of the MBH mechanism make it challenging to employ this method of determining the solvent shells. Instead, BP86-D3BJ/def2-SVP optimized structures from ABCluster were used to describe the local solvent environment of each MBH intermediate.¹¹² Solute-solvent clusters containing between zero to ten explicit methanol molecules were considered. We unexpectedly observed a lack of correlation between more methanol molecules and accuracy; showing that adding more explicit solvent molecules does not guarantee better predictions due to intrinsic errors in calculation procedures.

Reactive atomistic modeling and machine learning

We have stressed that different modeling schemes can be used depending on the complexity of the systems and computational resources available. Still, computational insight is generally most reliably derived when modeling as much of the actual reaction environment as possible. Reactive force fields, such as ReaxFF,^{113,114} are a framework to provide very useful insights even if these methods might not always be as accurate or transferable as QC methods they are trained on.

As an example, we previously studied catalytic reaction mechanisms using the computational hydrogen electrode model to understand how doping TiO_2 might *deoptimize* reduction kinetics and therefore result in an improved anticorrosion coating.¹¹⁵ The actual system involved Ti materials that formed native oxides that were believed to have largely amorphous structures, but QC analyses could not be carried out until atomic scale amorphous structures were established. Using ReaxFF parameters from another study,¹¹⁶ we used computationally efficient molecular dynamics to anneal crystalline TiO_2 structures into less ordered structures following analogous procedures used by Johnson and coworkers for studies of amorphous silica.¹¹⁷ The amorphous TiO_2 structures we obtained did not well represent experimental structures for small nanoparticles,¹¹⁸ but further geometry optimizations using DFT resulted

in structures that were in good agreement with experimental data. Furthermore, our analysis of atomic dopants predicted that aluminum and vanadium would be useful for suppressing oxygen reduction activity, and this was later confirmed by experiment in our publication. Thus, in this case ReaxFF was still very useful for identifying salient structures for reactions.

However, when we turned to analyze similar reaction mechanisms on more complicated TiAl_2O_5 materials, we sought to determine whether these materials could form stable amorphous structures. Attempts to train a ReaxFF model using automated procedures were unsuccessful (which could be due to any number of factors), so alternative automatably trained forcefield methods were sought. Ideally, BOMD with on-the-fly electronic structure calculations would be useful, but computational costs are severely limiting. ML is now a popular alternative to reactive force fields used in molecular simulation.^{119–121} Neural networks (NNs) have recently become popular for learning energies and forces across chemical space, one example being the Behler–Parrinello NN (BPNN) approach.¹²² To understand how BPNN methods compare to ReaxFF methods trained using the same DFT data sets (largely based off of earlier data sets),¹²³ we collaborated with Kitchin’s group to study Au bulk, surfaces, and clusters totaling 9,972 Kohn-Sham density functional calculations.¹²⁴ The optimally trained BPNN potential was trained on 9,734 calculations; while ReaxFF only required 848 data points. BPNN outperformed ReaxFF in all cases, but the computational cost for training the BPNN was substantially higher. Since existing ReaxFF parameters were not sufficiently accurate and BOMD simulations were too expensive, we turned to generate accurate BPNN potentials using Khorshidi and Peterson’s AMP code.¹²⁵ We then could predict that amorphous structures of TiAl_2O_5 were unlikely to form even though TiO_2 systems did, but analyses of dopants on TiAl_2O_5 systems were inconclusive with respect to experimental results from our collaborators at the Naval Research Lab.¹²⁶ This showed that agreement between experiment and computation on these more complicated systems are still an open question. In particular, the next step for this work might be to account for solvation interactions at the TiAl_2O_5 /electrolyte interface.

In principle, different variety of ML potentials could be used here to account for solvated interfaces, but we are intrigued at the possibilities of kernel-based symmetric gradient domain machine learning (sGDML) methods¹²⁷ that can be trained against potential energy surfaces of medium-sized molecules with far fewer data points than standard BPNN potentials (sGDML methods usually require only a few hundred training points whereas BPNNs can requires many thousands). The need for less data again allows one to focus computational costs on obtaining much higher quality data, for instance *ab initio* molecular dynamics energies and forces using correlated wave function methods. A key current limitation with sGDML force fields though is their reliance on a few molecular representations as well as their inability to be transferable to systems with different numbers of atoms than is used in their training set, and this currently limits there use to simulations on just single molecules or clusters. Current efforts by our group in collaboration with the Tkatchenko group have been toward exploring ways to overcome these limitations. Once overcome, we see opportunities for automated and efficient developments of ML models suitable for studies of mixed solvent systems,⁵ and using those in studies of electrochemical interfaces, e.g. in the context of Refs.:^{128–131}

Conclusions

We have given an overview and outlook for how our group uses computational chemistry to examine the thermodynamic and kinetic properties of hypothetical catalysts in complex environments. APDFT, especially using higher order corrections, has tremendous promise for accelerating computational screening efforts while relying on only relatively small sets of QC calculations. For example, APDFT can be used on a single set of images along a reaction pathway to generate insightful data for many other related processes. As it becomes more tested, it will become more likely to transform how conventional computational modeling workflows are used when exploring chemical space for new catalysts that are subjected to

conditions in their local environment under ambient conditions that include applied potential and pH effects. Finally, while less expensive modeling schemes can be effective, we envision comprehensive investigations of explicitly solvated reaction mechanisms will become possible with continued developments of ML force fields. All combined, one can envision elaborate workflows that would be suitable for microkinetic predictions based on massive search spaces for hypothetical candidates across chemical and materials space.

Acknowledgement

We acknowledge support from the R. K. Mellon Foundation and the U.S. National Science Foundation (Grant Nos. CBET-1653392, CBET-1705592, and CHE-1856460). C. D. G. acknowledges support from National Science Foundation Graduate Research Fellowship under Grant No. 1747452. Computational resources and technical support were provided by the University of Pittsburgh Center for Research Computing.

References

- (1) Schlögl, R. Heterogeneous Catalysis. *Angew. Chem. Int. Ed.* **2015**, *54*, 3465–3520.
- (2) Montoya, J. H.; Seitz, L. C.; Chakthranont, P.; Vojvodic, A.; Jaramillo, T. F.; Nørskov, J. K. Materials for solar fuels and chemicals. *Nat. Mater.* **2017**, *16*, 70–81.
- (3) Seh, Z. W.; Kibsgaard, J.; Dickens, C. F.; Chorkendorff, I.; Nørskov, J. K.; Jaramillo, T. F. Combining theory and experiment in electrocatalysis: Insights into materials design. *Science* **2017**, *355*, eaad4998.
- (4) Bullock, R. M.; Chen, J. G.; Gagliardi, L.; Chirik, P. J.; Farha, O. K.; Hendon, C. H.; Jones, C. W.; Keith, J. A.; Klosin, J.; Minteer, S. D., et al. Using nature’s blueprint to expand catalysis with Earth-abundant metals. *Science* **2020**, *369*, eabc3183.

- (5) Maldonado, A. M.; Basdogan, Y.; Berryman, J. T.; Rempe, S. B.; Keith, J. A. First-principles modeling of chemistry in mixed solvents: Where to go from here? *J. Chem. Phys.* **2020**, *152*, 130902.
- (6) Feibelman, P. J.; Hammer, B.; Nørskov, J. K.; Wagner, F.; Scheffler, M.; Stumpf, R.; Watwe, R.; Dumesic, J. The CO/Pt(111) Puzzle. *The Journal of Physical Chemistry B* **2001**, *105*, 4018–4025.
- (7) Stroppa, A.; Termentzidis, K.; Paier, J.; Kresse, G.; Hafner, J. CO adsorption on metal surfaces: A hybrid functional study with plane-wave basis set. *Phys. Rev. B* **2007**, *76*, 195440.
- (8) Tameh, M. S.; Huang, C. Large Impact of Approximate Exchange-Correlation Functionals on Modeling the Water Gas Shift Reaction on Copper. *J. Phys. Chem. C* **2020**, *124*, 22506–22520.
- (9) Liechtenstein, A. I.; Anisimov, V. I.; Zaanen, J. Density-functional theory and strong interactions: Orbital ordering in Mott-Hubbard insulators. *Phys. Rev. B* **1995**, *52*, R5467–R5470.
- (10) Anisimov, V. I.; Aryasetiawan, F.; Liechtenstein, A. I. First-principles calculations of the electronic structure and spectra of strongly correlated systems: the LDA + U method. *J. Phys. Condens. Matter* **1997**, *9*, 767–808.
- (11) Dudarev, S. L.; Liechtenstein, A. I.; Castell, M. R.; Briggs, G. A. D.; Sutton, A. P. Surface states on NiO(100) and the origin of the contrast reversal in atomically resolved scanning tunneling microscope images. *Phys. Rev. B* **1997**, *56*, 4900–4908.
- (12) Dudarev, S. L.; Botton, G. A.; Savrasov, S. Y.; Humphreys, C. J.; Sutton, A. P. Electron-energy-loss spectra and the structural stability of nickel oxide: An LSDA+U study. *Phys. Rev. B* **1998**, *57*, 1505–1509.

- (13) Perdew, J. P.; Schmidt, K. Jacob’s ladder of density functional approximations for the exchange-correlation energy. *AIP Conf. Proc.* **2001**, *577*, 1–20.
- (14) Libisch, F.; Huang, C.; Carter, E. A. Embedded correlated wavefunction schemes: Theory and applications. *Acc. Chem. Res.* **2014**, *47*, 2768–2775.
- (15) Huang, C.; Pavone, M.; Carter, E. A. Quantum mechanical embedding theory based on a unique embedding potential. *J. Chem. Phys.* **2011**, *134*, 154110.
- (16) Manby, F. R.; Stella, M.; Goodpaster, J. D.; Miller, T. F. A simple, exact density-functional-theory embedding scheme. *J. Chem. Theory Comput.* **2012**, *8*, 2564–2568.
- (17) Goodpaster, J. D.; Ananth, N.; Manby, F. R.; Miller, T. F. Exact nonadditive kinetic potentials for embedded density functional theory. *J. Chem. Phys.* **2010**, *133*, 084103.
- (18) Goodpaster, J. D.; Barnes, T. A.; Miller, T. F. Embedded density functional theory for covalently bonded and strongly interacting subsystems. *J. Chem. Phys.* **2011**, *134*, 164108.
- (19) Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller, T. F. Density functional theory embedding for correlated wavefunctions: Improved methods for open-shell systems and transition metal complexes. *J. Chem. Phys.* **2012**, *137*, 224113.
- (20) Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller, T. F. Accurate and systematically improvable density functional theory embedding for correlated wavefunctions. *J. Chem. Phys.* **2014**, *140*, 18A507.
- (21) Knizia, G.; Chan, G. K.-L. Density matrix embedding: A simple alternative to dynamical mean-field theory. *Phys. Rev. Lett.* **2012**, *109*, 186404.
- (22) Knizia, G.; Chan, G. K.-L. Density matrix embedding: A strong-coupling quantum embedding theory. *J. Chem. Theory Comput.* **2013**, *9*, 1428–1432.

- (23) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
- (24) Barone, V.; Cossi, M. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.
- (25) Klamt, A.; Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799–805.
- (26) Sundararaman, R.; Goddard III, W. A. The charge-asymmetric nonlocally determined local-electric (CANDLE) solvation model. *J. Chem. Phys.* **2015**, *142*, 064107.
- (27) Prentice, J. C. A.; Aarons, J.; Womack, J. C.; Allen, A. E. A.; Andrinopoulos, L.; Anton, L.; Bell, R. A.; Bhandari, A.; Bramley, G. A.; Charlton, R. J., et al. The ONETEP linear-scaling density functional theory program. *J. Chem. Phys.* **2020**, *152*, 174111.
- (28) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. cp2k: atomistic simulations of condensed matter systems. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 15–25.
- (29) Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. Ab initio molecular simulations with numeric atom-centered orbitals. *Comput. Phys. Commun* **2009**, *180*, 2175–2196.
- (30) Neese, F. The ORCA program system. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 73–78.

- (31) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54*, 11169–11186.
- (32) von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Exploring chemical compound space with quantum-based machine learning. *Nat. Rev. Chem.* **2020**, *4*, 347–358.
- (33) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G., et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002.
- (34) de Pablo, J. J.; Jackson, N. E.; Webb, M. A.; Chen, L.-Q.; Moore, J. E.; Morgan, D.; Jacobs, R.; Pollock, T.; Schlom, D. G.; Toberer, E. S., et al. New frontiers for the materials genome initiative. *npj Comput. Mater.* **2019**, *5*, 41.
- (35) Nørskov, J. K.; Bligaard, T. The Catalyst Genome. *Angew. Chem. Int. Ed.* **2013**, *52*, 776–777.
- (36) O’Mara, J.; Meredig, B.; Michel, K. Materials data infrastructure: A case study of the citrination platform to examine data import, storage, and access. *JOM* **2016**, *68*, 2031–2034.
- (37) Winther, K. T.; Hoffmann, M. J.; Boes, J. R.; Mamun, O.; Bajdich, M.; Bligaard, T. Catalysis-Hub.org, an open electronic structure database for surface reactions. *Sci. Data* **2019**, *6*, 75.
- (38) von Lilienfeld, O. A.; Lins, R. D.; Rothlisberger, U. Variational particle number approach for rational compound design. *Phys. Rev. Lett.* **2005**, *95*, 153002.
- (39) von Lilienfeld, O. A.; Tuckerman, M. E. Molecular grand-canonical ensemble density functional theory and exploration of chemical space. *J. Chem. Phys.* **2006**, *125*, 154104.

- (40) to Baben, M.; Achenbach, J. O.; von Lilienfeld, O. A. Guiding ab initio calculations by alchemical derivatives. *J. Chem. Phys.* **2016**, *144*, 104103.
- (41) Griego, C. D.; Kitchin, J. R.; Keith, J. A. Acceleration of catalyst discovery with easy, fast, and reproducible computational alchemy. *Int. J. Quantum Chem.* **2020**, *121*, e26380.
- (42) Alfè, D.; Gillan, M. J.; Price, G. D. Constraints on the composition of the Earth’s core from ab initio calculations. *Nature* **2000**, *405*, 172–175.
- (43) Beste, A.; Harrison, R. J.; Yanai, T. Direct computation of general chemical energy differences: Application to ionization potentials, excitation, and bond energies. *J. Chem. Phys.* **2006**, *125*, 074101.
- (44) Solovyeva, A.; von Lilienfeld, O. A. Alchemical screening of ionic crystals. *Phys. Chem. Chem. Phys.* **2016**, *18*, 31078–31091.
- (45) Chang, K. Y. S.; Fias, S.; Ramakrishnan, R.; von Lilienfeld, O. A. Fast and accurate predictions of covalent bonds in chemical space. *J. Chem. Phys.* **2016**, *144*, 174110.
- (46) Chang, K. Y. S.; von Lilienfeld, O. A. $\text{Al}_x\text{Ga}_{1-x}\text{As}$ crystals with direct 2 eV band gaps from computational alchemy. *Phys. Rev. Mater.* **2018**, *2*, 073802.
- (47) Sheppard, D.; Henkelman, G.; von Lilienfeld, O. A. Alchemical derivatives of reaction energetics. *J. Chem. Phys.* **2010**, *133*, 084104.
- (48) Weigend, F.; Schrod, C.; Ahlrichs, R. Atom distributions in binary atom clusters: A perturbational approach and its validation in a case study. *J. Chem. Phys.* **2004**, *121*, 10380–10384.
- (49) Weigend, F. Extending DFT-based genetic algorithms by atom-to-place re-assignment via perturbation theory: A systematic and unbiased approach to structures of mixed-metallic clusters. *J. Chem. Phys.* **2014**, *141*, 134103.

- (50) Al-Hamdani, Y. S.; Michaelides, A.; von Lilienfeld, O. A. Exploring dissociative water adsorption on isoelectronically BN doped graphene using alchemical derivatives. *J. Chem. Phys.* **2017**, *147*, 164113.
- (51) Balawender, R.; Welearegay, M. A.; Lesiuk, M.; De Proft, F.; Geerlings, P. Exploring Chemical Space with the Alchemical Derivatives. *J. Chem. Theory Comput.* **2013**, *9*, 5327–5340.
- (52) Balawender, R.; Lesiuk, M.; De Proft, F.; Geerlings, P. Exploring Chemical Space with Alchemical Derivatives: BN -Simultaneous Substitution Patterns in C 60. *J. Chem. Theory Comput.* **2018**, *14*, 1154–1168.
- (53) Fias, S.; Heidar-Zadeh, F.; Geerlings, P.; Ayers, P. W. Chemical transferability of functional groups follows from the nearsightedness of electronic matter. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 11633–11638.
- (54) Saravanan, K.; Kitchin, J. R.; von Lilienfeld, O. A.; Keith, J. A. Alchemical predictions for computational catalysis: Potential and limitations. *J. Phys. Chem. Lett.* **2017**, *8*, 5002–5007.
- (55) Griego, C. D.; Saravanan, K.; Keith, J. A. Benchmarking computational alchemy for carbide, nitride, and oxide catalysts. *Adv. Theory Simul.* **2019**, *2*, 1800142.
- (56) Griego, C. D.; Zhao, L.; Saravanan, K.; Keith, J. A. Machine learning corrected alchemical perturbation density functional theory for catalysis applications. *AIChE J.* **2020**, *66*, e17041.
- (57) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **2000**, *113*, 9901–9904.

- (58) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; Von Lilienfeld, O. A. Big data meets quantum chemistry approximations: The Δ -machine learning approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087–2096.
- (59) von Rudorff, G. F.; von Lilienfeld, O. A. Rapid and accurate molecular deprotonation energies from quantum alchemy. *Phys. Chem. Chem. Phys.* **2020**, *22*, 10519–10525.
- (60) Wang, S.; Pillai, H. S.; Xin, H. Bayesian learning of chemisorption for bridging complexities of electronic descriptors. *Nat. Commun.* **2020**, *11*, 6132.
- (61) Boes, J. R.; Groenenboom, M. C.; Keith, J. A.; Kitchin, J. R. Neural network and ReaxFF comparison for Au properties. *Int. J. Quantum Chem.* **2016**, *116*, 979–987.
- (62) Boes, J. R.; Kitchin, J. R. Neural network predictions of oxygen interactions on a dynamic Pd surface. *Mol. Simul.* **2017**, *43*, 346–354.
- (63) Nørskov, J. K.; Rossmeisl, J.; Logadottir, A.; Lindqvist, L.; Kitchin, J. R.; Bligaard, T.; Jónsson, H. Origin of the overpotential for oxygen reduction at a fuel-cell cathode. *J. Phys. Chem. B* **2004**, *108*, 17886–17892.
- (64) Reuter, K.; Scheffler, M. First-principles atomistic thermodynamics for oxidation catalysis: surface phase diagrams and catalytically interesting regions. *Phys. Rev. Lett.* **2003**, *90*, 046103.
- (65) Exner, K. S.; Anton, J.; Jacob, T.; Over, H. Chlorine evolution reaction on RuO₂(110): Ab initio atomistic thermodynamics study - pourbaix diagrams. *Electrochim. Acta* **2014**, *120*, 460–466.
- (66) Marjolin, A.; Keith, J. A. Thermodynamic descriptors for molecules that catalyze efficient CO₂ electroreductions. *ACS Catal.* **2015**, *5*, 1123–1130.
- (67) Saravanan, K.; Basdogan, Y.; Dean, J.; Keith, J. A. Computational investigation of

- CO₂ electroreduction on tin oxide and predictions of Ti, V, Nb and Zr dopants for improved catalysis. *J. Mater. Chem. A* **2017**, *5*, 11756–11763.
- (68) Jacob, T. Theoretical investigations on the potential-induced formation of Pt-oxide surfaces. *J. Electroanal. Chem.* **2007**, *607*, 158 – 166, Theoretical and Computational Electrochemistry.
- (69) Saravanan, K.; Gottlieb, E.; Keith, J. A. Nitrogen-doped nanocarbon materials under electroreduction operating conditions and implications for electrocatalysis of CO₂. *Carbon* **2017**, *111*, 859–866.
- (70) Yang, J. Y.; Kerr, T. A.; Wang, X. S.; Barlow, J. M. Reducing CO₂ to HCO₂⁻ at Mild Potentials: Lessons from Formate Dehydrogenase. *J. Am. Chem. Soc.* **2020**, *142*, 19438–19445.
- (71) Chatterjee, S.; Griego, C.; Hart, J. L.; Li, Y.; Taheri, M. L.; Keith, J.; Snyder, J. D. Free standing nanoporous palladium alloys as CO poisoning tolerant electrocatalysts for the electrochemical reduction of CO₂ to formate. *ACS Catal.* **2019**, *9*, 5290–5301.
- (72) Goldsmith, C. F.; West, R. H. Automatic Generation of Microkinetic Mechanisms for Heterogeneous Catalysis. *J. Phys. Chem. C* **2017**, *121*, 9970–9981.
- (73) Blondal, K.; Jelic, J.; Mazeau, E.; Studt, F.; West, R. H.; Goldsmith, C. F. Computer-generated kinetics for coupled heterogeneous/homogeneous systems: A case study in catalytic combustion of methane on platinum. *Ind. Eng. Chem. Res.* **2019**, *58*, 17682–17691.
- (74) Zhang, Z.; Zandkarimi, B.; Alexandrova, A. N. Ensembles of metastable states govern heterogeneous catalysis on dynamic interfaces. *Acc. Chem. Res.* **2020**, *53*, 447–458.
- (75) Cheula, R.; Maestri, M.; Mpourmpakis, G. Modeling morphology and catalytic activ-

- ity of nanoparticle ensembles under reaction conditions. *ACS Catal.* **2020**, *10*, 6149–6158.
- (76) Calle-Vallejo, F.; Loffreda, D.; Koper, M. T. M.; Sautet, P. Introducing structural sensitivity into adsorption–energy scaling relations by means of coordination numbers. *Nat. Chem.* **2015**, *7*, 403–410.
- (77) Yan, Z.; Taylor, M. G.; Mascareno, A.; Mpourmpakis, G. Size-, shape-, and composition-dependent model for metal nanoparticle stability prediction. *Nano Lett.* **2018**, *18*, 2696–2704.
- (78) Roling, L. T.; Choksi, T. S.; Abild-Pedersen, F. A coordination-based model for transition metal alloy nanoparticles. *Nanoscale* **2019**, *11*, 4438–4452.
- (79) Basdogan, Y.; Maldonado, A. M.; Keith, J. A. Advances and challenges in modeling solvated reaction mechanisms for renewable fuels and chemicals. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2020**, *10*, e1446.
- (80) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999–3094.
- (81) Tomasi, J.; Persico, M. Molecular interactions in solution: an overview of methods based on continuous distributions of the solvent. *Chem. Rev.* **1994**, *94*, 2027–2094.
- (82) Klamt, A. *COSMO-RS: from quantum chemistry to fluid phase thermodynamics and drug design*; Elsevier, 2005.
- (83) Hirata, F. *Molecular theory of solvation*; Springer, Dordrecht, 2003; Vol. 24.
- (84) Pliego Jr, J. R.; Riveros, J. M. Hybrid discrete-continuum solvation methods. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2020**, *10*, e1440.
- (85) Chaudhari, M. I.; Vanegas, J. M.; Pratt, L.; Muralidharan, A.; Rempe, S. B. Hydration mimicry by membrane ion channels. *Annu. Rev. Phys. Chem.* **2020**, *71*, 461–484.

- (86) Asthagiri, D.; Dixit, P.; Merchant, S.; Paulaitis, M.; Pratt, L.; Rempe, S. B.; Varma, S. Ion selectivity from local configurations of ligands in solutions and ion channels. *Chem. Phys. Lett.* **2010**, *485*, 1–7.
- (87) Pratt, L. R.; Asthagiri, D. In *Free energy calculations: Theory and applications in chemistry and biology*; Chipot, C., Pohorille, A., Eds.; Springer, 2007; pp 323–351.
- (88) Beck, T. L.; Paulaitis, M. E.; Pratt, L. R. *The potential distribution theorem and models of molecular solutions*; Cambridge University Press, 2006.
- (89) Pratt, L. R.; Rempe, S. B. Quasi-chemical theory and implicit solvent models for simulations. *AIP Conf. Proc.* **1999**, *492*, 172–201.
- (90) Bryantsev, V. S.; Diallo, M. S.; Goddard III, W. A. Computational study of copper (II) complexation and hydrolysis in aqueous solutions using mixed cluster/continuum models. *J. Phys. Chem. A* **2009**, *113*, 9559–9567.
- (91) Basdogan, Y.; Groenenboom, M. C.; Henderson, E.; De, S.; Rempe, S. B.; Keith, J. A. Machine learning-guided approach for studying solvation environments. *J. Chem. Theory Comput.* **2019**, *16*, 633–642.
- (92) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B* **2013**, *87*, 184115.
- (93) Ceriotti, M.; Tribello, G. A.; Parrinello, M. Simplifying the representation of complex free-energy landscapes using sketch-map. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 13023–13028.
- (94) Zhang, J.; Dolg, M. ABCcluster: the artificial bee colony algorithm for cluster global optimization. *Phys. Chem. Chem. Phys.* **2015**, *17*, 24173–24181.
- (95) Simm, G. N.; Türtcher, P. L.; Reiher, M. Systematic microsolvation approach with

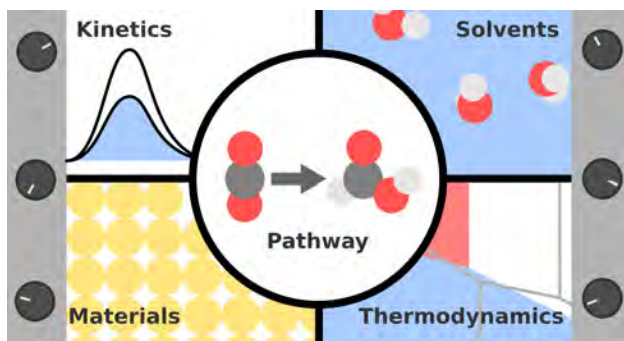
- a cluster-continuum scheme and conformational sampling. *J. Comput. Chem.* **2020**, *41*, 1144–1155.
- (96) Rogers, D. M.; Jiao, D.; Pratt, L. R.; Rempe, S. B. In *Annual Reports in Computational Chemistry*; Wheeler, R. A., Ed.; Elsevier, 2012; Vol. 8; pp 71–127.
- (97) Sunoj, R. B.; Anand, M. Microsolvated transition state models for improved insight into chemical properties and reaction mechanisms. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12715–12736.
- (98) Roytman, V. A.; Singleton, D. A. Solvation dynamics and the nature of reaction barriers and ion-pair intermediates in carbocation reactions. *J. Am. Chem. Soc.* **2020**, *142*, 12865–12877.
- (99) Hassanali, A. A.; Cuny, J.; Verdolino, V.; Parrinello, M. Aqueous solutions: state of the art in ab initio molecular dynamics. *Philos. Trans. R. Soc. A* **2014**, *372*, 20120482.
- (100) Iftimie, R.; Minary, P.; Tuckerman, M. E. Ab initio molecular dynamics: Concepts, recent developments, and future trends. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6654–6659.
- (101) Tse, J. S. Ab initio molecular dynamics with density functional theory. *Annu. Rev. Phys. Chem.* **2002**, *53*, 249–290.
- (102) Lu, X.; Fang, D.; Ito, S.; Okamoto, Y.; Ovchinnikov, V.; Cui, Q. QM/MM free energy simulations: Recent progress and challenges. *Mol. Simul.* **2016**, *42*, 1056–1078.
- (103) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.
- (104) Lin, H.; Truhlar, D. G. QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* **2007**, *117*, 185.

- (105) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition Path Sampling: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291–318.
- (106) Agarwal, V.; Dauenhauer, P. J.; Huber, G. W.; Auerbach, S. M. Ab initio dynamics of cellulose pyrolysis: Nascent decomposition pathways at 327 and 600 C. *J. Am. Chem. Soc.* **2012**, *134*, 14958–14972.
- (107) Groenenboom, M. C.; Keith, J. A. Explicitly unraveling the roles of counterions, solvent molecules, and electron correlation in solution phase reaction pathways. *J. Phys. Chem. B* **2016**, *120*, 10797–10807.
- (108) Groenenboom, M. C.; Keith, J. A. Quantum chemical analyses of BH_4^- and BH_3OH^- hydride transfers to CO_2 in aqueous solution with potentials of mean force. *ChemPhysChem* **2017**, *18*, 3148–3152.
- (109) Sheppard, D.; Xiao, P.; Chemelewski, W.; Johnson, D. D.; Henkelman, G. A generalized solid-state nudged elastic band method. *J. Chem. Phys.* **2012**, *136*, 074103.
- (110) Plata, R. E.; Singleton, D. A. A case study of the mechanism of alcohol-mediated Morita Baylis–Hillman reactions. The importance of experimental observations. *J. Am. Chem. Soc.* **2015**, *137*, 3811–3826.
- (111) Liu, Z.; Patel, C.; Harvey, J. N.; Sunoj, R. B. Mechanism and reactivity in the Morita–Baylis–Hillman reaction: the challenge of accurate computations. *Phys. Chem. Chem. Phys.* **2017**, *19*, 30647–30657.
- (112) Basdogan, Y.; Keith, J. A. A parametric treatment for modeling explicitly solvated chemical reaction mechanisms. *Chem. Sci.* **2018**, *9*, 5341–5346.
- (113) Van Duin, A. C.; Dasgupta, S.; Lorant, F.; Goddard III, W. A. ReaxFF: a reactive force field for hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396–9409.

- (114) The ReaxFF reactive force-field: development, applications and future directions. *npj Comput. Mat.* **2016**, *2*, 15011.
- (115) Groenenboom, M.; Anderson, R.; Horton, D.; Basdogan, Y.; Roeper, D.; Policastro, S.; Keith, J. Doped Amorphous Ti Oxides to Deoptimize Oxygen Reduction Reaction Catalysis. *J. Phys. Chem. C* **2017**, *121*, 16825–16830.
- (116) Kim, S.-Y.; Kumar, N.; Persson, P.; Sofo, J.; van Duin, A. C. T.; Kubicki, J. D. Development of a ReaxFF Reactive Force Field for Titanium Dioxide/Water Systems. *Langmuir* **2013**, *29*, 7838–7846.
- (117) Ewing, C. S.; Bhavsar, S.; Vesper, G.; McCarthy, J. J.; Johnson, J. K. Accurate Amorphous Silica Surface Models from First-Principles Thermodynamics of Surface Dehydroxylation. *Langmuir* **2014**, *30*, 5133–5141.
- (118) Petkov, V.; Holzhüter, G.; Tröge, U.; Gerber, T.; Himmel, B. Atomic-scale structure of amorphous TiO₂ by electron, X-ray diffraction and reverse Monte Carlo simulations. *J. Non-Cryst. Solids* **1998**, *231*, 17–30.
- (119) Schütt, K. T.; Chmiela, S.; von Lilienfeld, O. A.; Tkatchenko, A.; Tsuda, K.; Müller, K.-R. *Machine Learning Meets Quantum Physics*; Springer, 2020.
- (120) Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **2016**, *145*, 170901.
- (121) von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Exploring chemical compound space with quantum-based machine learning. *Nat. Rev. Chem.* **2020**, *4*, 347–358.
- (122) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.
- (123) Keith, J. A.; Fantauzzi, D.; Jacob, T.; Van Duin, A. C. Reactive forcefield for simulating gold surfaces and nanoparticles. *Phys. Rev. B* **2010**, *81*, 235404.

- (124) Boes, J. R.; Groenenboom, M. C.; Keith, J. A.; Kitchin, J. R. Neural network and ReaxFF comparison for Au properties. *Int. J. Quantum Chem.* **2016**, *116*, 979–987.
- (125) Khorshidi, A.; Peterson, A. A. Amp: A modular approach to machine learning in atomistic simulations. *Comput. Phys. Commun.* **2016**, *207*, 310–324.
- (126) Groenenboom, M. C.; Anderson, R. M.; Wollmershauser, J. A.; Horton, D. J.; Policastro, S. A.; Keith, J. A. Combined Neural Network Potential and Density Functional Theory Study of TiAl_2O_5 Surface Morphology and Oxygen Reduction Reaction Overpotentials. *J. Phys. Chem. C* **2020**, *124*, 15171–15179.
- (127) Chmiela, S.; Sauceda, H. E.; Müller, K.-R.; Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.* **2018**, *9*, 3887.
- (128) Sakong, S.; Mahlberg, D.; Roman, T.; Li, M.; Pandey, M.; Groß, A. Influence of Local Inhomogeneities and the Electrochemical Environment on the Oxygen Reduction Reaction on Pt-Based Electrodes: A DFT Study. *J. Phys. Chem. C* **2020**, *124*, 27604–27613.
- (129) Rendón-Calle, A.; Builes, S.; Calle-Vallejo, F. Substantial improvement of electrocatalytic predictions by systematic assessment of solvent effects on adsorption energies. *Appl. Catal. B: Environ.* **2020**, *276*, 119147.
- (130) Kastlunger, G.; Lindgren, P.; Peterson, A. A. Controlled-Potential Simulation of Elementary Electrochemical Reactions: Proton Discharge on Metal Surfaces. *J. Phys. Chem. C* **2018**, *122*, 12771–12781.
- (131) Heenen, H. H.; Gauthier, J. A.; Kristoffersen, H. H.; Ludwig, T.; Chan, K. Solvation at metal/water interfaces: An ab initio molecular dynamics benchmark of common computational approaches. *J. Chem. Phys.* **2020**, *152*, 144703.

TOC Graphic



Author Biographies

Charles D. Griego

Charles D. Griego received his bachelor's degree in chemical engineering at the New Mexico Institute of Mining and Technology in 2017. He is currently an NSF Graduate Research Fellow and fourth-year Ph.D. student at the University of Pittsburgh. His research focuses on the development and application of alchemical perturbation density functional theory to heterogeneous catalyst systems.

Alex M. Maldonado

Alex M. Maldonado received his bachelor's in chemical engineering at Western Michigan University in 2018 and was a MI-LSAMP scholar. He is currently a Pitt STRIVE scholar, a R. K. Mellon Graduate Fellow, and a third-year Ph.D. student at the University of Pittsburgh. His research focuses on the development and application of computational chemistry methods to elucidate solvated reaction mechanisms.

Lingyan Zhao

Lingyan Zhao received her bachelor's degrees in Materials science at Beijing University of Chemical Technology in 2018 and master's degree at Carnegie Mellon University in 2019.

She is currently a second-year Ph.D. student at the University of Pittsburgh. Her research focuses on modeling electrocatalytic homogeneous and heterogeneous reaction mechanisms.

Barbaro Zulueta

Barbaro Zulueta obtained his bachelor's degrees in Physics and Chemistry from Temple University in 2019. He is a first-year Ph.D. student at the University of Pittsburgh. His research focuses on development and applications of alchemical perturbation density functional theory.

Brian M. Gentry

Brian Gentry completed his bachelor's degree in mechanical engineering at the University of Pittsburgh in 2020 where his research focused on modeling molecular chelants and gelator molecules. He is now a graduate student and an NSF Graduate Research Fellow at Carnegie Mellon University pursuing a joint Ph.D. in mechanical engineering and engineering and public policy.

Eli Lipsman

Eli Lipsman was a undergraduate researcher who worked on implementing descriptors for machine learning force fields while finishing his bachelor's degree in chemical engineering.

Tae Hoon Choi

Tae Hoon Choi earned his Ph.D. in chemistry from the University of Pittsburgh, and did post-doctoral research at the University of Chicago. After six years of being a professor in Chungnam National University in Korea, he returned to the University of Pittsburgh in 2017 to be a research assistant professor. His research focuses on implementing multi-scale simulations for metal-ligand binding and solvated reactions.

John A. Keith

John A. Keith earned a bachelors from Wesleyan University and a Ph.D. from Caltech. After an Alexander von Humboldt postdoctoral fellowship at the University of Ulm (Germany), he was an Associate Research Scholar at Princeton University. He began his independent career in 2013 at the University of Pittsburgh in the Department of Chemical and Petroleum Engineering as an inaugural R.K. Mellon Faculty Fellow in Energy. He received an NSF-CAREER award in 2017 and was promoted to associate professor with tenure in 2019. His research develops and applied computational chemistry methods for insights into the improved design of molecules and materials.