Understanding Solvation Environments in Chemical Systems

by

Yasemin Basdogan

B.S. in Chemical and Biological Engineering, KOC University,

Turkey, 2015

Submitted to the Graduate Faculty of

the Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Yasemin Basdogan

It was defended on

January 3, 2020

and approved by

John A. Keith, Ph.D., Associate Professor, Department of Chemical and Petroleum

Engineering

J. Karl Johnson, Ph.D., Professor, Department of Chemical and Petroleum Engineering

Kenneth D. Jordan, Ph.D., Professor, Department of Chemistry

Christopher E. Wilmer, Ph.D., Assistant Professor, Department of Chemical and

Petroleum Engineering

Dissertation Director: John A. Keith, Ph.D., Associate Professor, Department of Chemical

and Petroleum Engineering

Understanding Solvation Environments in Chemical Systems

Yasemin Basdogan, PhD

University of Pittsburgh, 2020

Molecular level understanding and characterization of solvation environments is often needed across chemistry, biology, and engineering. In many cases, the explicit interactions between molecules with nearby solvents are crucial for molecular-scale understanding. Toward practical modelling of local solvation effects of any solute in any solvent, we developed a general, all-QM, cluster-continuum approach. This approach uses a global optimization procedure to identify low energy molecular clusters with different numbers of explicit solvent molecules and then employs a machine learning algorithm with the help of the Smooth Overlap of Atomic Positions (SOAP) kernel to quantify the similarity between different low-energy solvent environments. From these data, we use a sketch-map non-linear dimensionality reduction technique to obtain a visual representation of the similarity between solvent environments in differently sized microsolvated clusters. After studying the evolution of the local solvation environment around the molecules, we systematically explore reaction pathways using Growing String Method. Without needing either dynamics simulations or an *a priori* knowledge of the local solvation structure, this procedure was used to calculate reaction energies, solvation free energies and barrier heights in solvated systems. We now use this approach to model reaction mechanisms in more complicated reaction environments that are relevant for renewable fuels and chemicals. We reliably predict CO_2 hydrogenation pathways and calculate barrier heights under electrochemical environments. This approach can be used to study physically significant solvation environments in any solvated system where the solvent molecules affects the quantum level nature of reaction mechanisms.

Table of Contents

List of Tables

List of Figures

Preface

First of all, I would like to thank my PhD advisor Dr. John Keith. He changed my life by hiring me into his research group. Over the past five years, I have learned a lot from him, not just about research but about life. He always believed in me and supported me in every step of my career. Even though we had our differences in the beginning, when I think about it now, this has been an amazing fit and we made a great team.

I would like to thank my committee members Dr. Karl Johnson, Dr. Ken Jordan, and Dr. Chris Wilmer for their time and useful feedback on my PhD thesis. I met with Chris six years ago at a AIChE conference in Atlanta and he changed my life my convincing me to joining Pitt. Dr. Johnson has been a second mentor to me over the past five years. He welcomed me into his family for numerous occasions when I first moved to Pittsburgh.

I would also like to thank Dr. Michele Ceriotti on collaborating with us on machine learning and who welcomed me in his group for a summer internship. Michele and his students helped me understand machine learning when I spent a month at their research group in Lausanne. I am also very grateful to Dr. Susan Rempe for her scientific advice, knowledge and many insightful discussions and suggestions.

I feel like I have been very fortunate when it came to lab members. I would like to thank all past and present members of the Keith Group for creating a joyful working environment. I would like to acknowledge Dr. Mitchell Groenenboom and Dr. Karthikeyan Saravanan, two senior graduate students in the Keith Group when I first started. I have learned a lot from them about research and PhD studies. Next, I would like to give my special thanks to Alex Maldonado and Chasz Griego, my current lab members. They have made this job very easy for me by making the office a friendly environment. The emotional support they have showed over the past couple of months made my life easier while I was trying to finish-up my PhD research.

1.0   Advances and Challenges in Modeling Solvated Reaction Mechanisms

This work has been published as Basdogan, Y., Maldonado, A. M., & Keith, J. A. (2020). Advances and challenges in modeling solvated reaction mechanisms for renewable fuels and chemicals. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 10(2), e1446.

1.1   Introduction

Developing sources of renewable energy is paramount to long-term human sustainability. [3, 4, 5] For instance, CO_2 emissions correlate with severe weather patterns[6] and global climate change,[7] but more than 78% of the world's energy consumption through the year 2040 is expected to come from fossil fuels.[8] Thus, many are interested in recycling anthropogenic CO_2 into fuels and chemicals[9, 10, 11, 12, 13] as well as sustainably producing ammonia[14, 15] and/or hydrogen.[16, 17] Unfortunately, most of these are currently unfeasible on large scales due to low conversion efficiency and/or high electrochemical overpotentials. For CO_2 electroreductions into fuels and chemicals, selectivity and energy efficiency remain as major challenges for proton and electron transfers.[18] These challenges are also present in other fundamental transformations such as N_2 reduction for ammonia synthesis[19] and H_2O oxidation for H_2 generation.[20]

Computational quantum chemistry modeling can help interpret and guide experimental work in this area by providing insights into chemical reaction mechanisms. Advances in algorithms and hardware make it easier to computationally model larger scale systems with higher accuracy, but the central challenges of understanding *what* processes to model and *how* to physically model them in a reliable way still remain. Indeed, many chemical reactions have intermediate states that are stabilized by different degrees of solvating environments, and neglecting or incorrectly modeling these environments can significantly impact the quality of predictions from computational modeling. We begin this section by summarizing how one can use computational modeling to explore the chemical and materials space of renewable

energy catalysis through the lens of identifying energetically efficient hydrogenation pathways for CO_2 reduction catalysis. We will then summarize different approaches to model solvating environments in reaction mechanism studies while also reviewing knowns and unknowns from recent literature.

1.2 The Chemical Space of Hydrogenation Reaction Mechanisms

At a fundamental level, any hydrogenation (or dehydrogenation) process for any reaction might occur as:

1. One or more covalent hydrogen atom (H·) transfers (e.g. with thermal heterogeneous catalytic processes).[21]

2. Stepwise or coupled proton and electron transfers that originate from different sites within the system (e.g. with electrochemical processes).[22]

3. Formal hydride (H$^-$) transfers that may also be coupled with a proton transfer (e.g. with biomimetic processes).[23]

Analogous classifications have been used by many others to distinguish different modes for hydrogenation,[24, 25, 26, 27, 28, 29, 30] and each class has been studied in different contexts of homogeneous,[31] heterogeneous,[32] or biological catalysis.[33] Clearly, the local chemical environment (especially a solvating environment) will play a role in determining the nature of the hydrogenation mechanism.

To understand how environmental conditions can influence multistep processes, we can start by defining a map of elementary electrochemical processes using a 'square-scheme' or 'schemes of squares'[34] and draw analogies to moves on a chessboard (see Figure 1). Here, a generic molecule A can undergo elementary steps to form a new reduced, hydrogenated state (AH$_n$). Individual proton transfer steps are normally represented as horizontal steps, individual electron transfer steps are then represented as vertical steps, and proton-coupled electron transfer steps are diagonal steps, i.e. all possible moves that a king piece is allowed to make in a chess game. Alternatively, an elementary hydride transfer would be an 'L'-
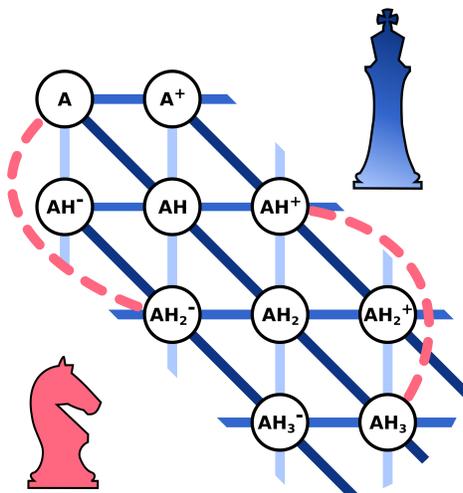
Figure 1: A "square-scheme" of hypothetical pathways for a multistep (de)hydrogenation process.

shaped step that involves two electron transfers and one proton transfer, i.e. a possible move that a knight piece can make. Before going too far, some aspects warrant mention. First, it is usually rare to find a reaction intermediate having a charge with an absolute magnitude of two or more unless there is a polarizing solvent and/or counterions nearby.[35] Thus, it is usually not likely (though not impossible) to move two or more steps away from the diagonal line depicting neutral intermediate states. Second, this square-scheme shows that several different pathways may exist for any multistep process, just like there are multiple paths a chess piece might move from one corner of a board to another. It requires confirmations from experiment and reliable computational modeling to assess which pathways are relevant under specific conditions. We will now describe how computational quantum chemistry can be leveraged to accelerate the discovery of energetically efficient reaction steps.

1.2.1 Theoretical Phase Diagrams

If hypothetical reaction intermediates can be identified, one can then use computational quantum chemistry to calculate the absolute free energies of each species using the standard ideal gas, rigid rotor, and harmonic oscillator approximations. From these data, one can

then make phase diagrams that are functions of parameters (e.g. solution pH, an electrode potential, and/or partial pressures of molecular species) that can be used to navigate chemical and/or materials space.[36] For instance, one can define a generic reaction that refers to intermediates from Figure 1 using Equation 1.1:

$$A + \frac{n}{2}H_2 \rightarrow AH_n \tag{1.1}$$

The corresponding free energy for this reaction at an arbitrary standard state (°) is simply the difference of the free energies of the individual products and reactants:

$$\Delta G_{\mathrm{rxn}}^{\circ} = G_{\mathrm{AH}_n}^{\circ} - G_{\mathrm{A}}^{\circ} - \frac{n}{2}G_{\mathrm{H}_2}^{\circ} \tag{1.2}$$

Note that the free energy of H_2 is also related to the definition for the standard hydrogen electrode (SHE) potential,

$$H^+ + e^- \rightleftharpoons \frac{1}{2}H_2 \qquad E = 0 \text{ V vs. SHE} \tag{1.3}$$
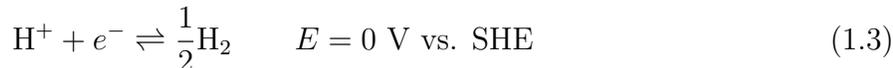
while the free energy for protons, electrons, or other species such as A or AH_n can be expressed as linear functions of a local environmental parameters such as pH, applied potential ϕ, or the relative difference in chemical potential from its standard state $\Delta\mu_X$, respectively. Note that SHE is a commonly used reference electrode which is a hypothetical electrode immersed in a 1 M aqueous solution of proton with unit activity and no ionic interactions. Other reference electrode systems (SCE, NHE, RHE, etc.) can be computationally modeled as well, and some discussion is found in the perspective paper by Marenich et al.[37] Using the SHE reference electrode model, one could define the reaction free energy from Equation 1.2 in an expanded form of several different species, each having a corresponding parameter (all expressed in eV units):

$$\Delta G_{\mathrm{rxn}} = \left(G_{\mathrm{AH}_n}^{\circ} + \Delta\mu_{\mathrm{AH}_n}\right) - \left(G_{\mathrm{A}}^{\circ} + \Delta\mu_{\mathrm{A}}\right) - \left(G_{\mathrm{H}^+}^{\circ} - 0.059 \text{ pH}\right) - \left(G_{e^-}^{\circ} - eU_{\mathrm{SHE}}\right) \tag{1.4}$$

Note that values such as $G_{\mathrm{AH}_n}^{\circ}$ and G_{A}° can be straightforwardly calculated using quantum chemistry codes. $G_{\mathrm{H}^+}^{\circ}$ and $G_{e^-}^{\circ}$ correspond to absolute free energies of a proton and electron in some environment and can be referenced from the literature.[37] The remaining $\Delta\mu_X$ terms

are treated as linear variables that describe environmental factors, e.g. partial pressure of a specific species, a solution pH, or an applied potential.

Considering large numbers of hypothetical reactions and determining the most favorable state at any given set of environmental conditions in this general framework begets "*ab initio*" atomistic thermodynamics phase diagrams that would show any ΔG for any hypothetical reaction at a specified set of conditions. For instance, if pH (x-axis) and ϕ (y-axis) were used as parameters, one would create a Pourbaix diagram, i.e. a phase diagram that depicts the thermodynamically most stable state for a system at a given pH and ϕ.[38] A representative set of Pourbaix diagrams is given in Figure 2.
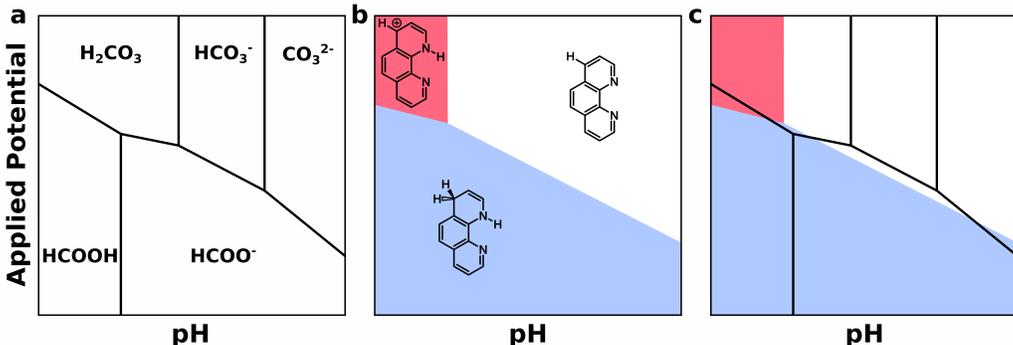
Figure 2: a) Pourbaix diagram showing stable states of the reactant, CO_2; b) Pourbaix diagram showing stable states of a hypothetical molecular catalyst, 1,10-phenanthroline; c) overlaid Pourbaix diagrams from a) and b) showing similar boundaries for hydrogen shuttling and CO_2 reduction. Vertical lines represent pK_as, the horizontal lines represent the pH-independent standard redox potentials and the diagonal lines represent the pH-dependent proton-coupled electron transfer steps.

While Pourbaix diagrams only provide insights into the thermodynamics of different intermediate states, they are still quite useful. First, they are a convenient representation of pK_as, pH-independent standard redox potentials, and pH-dependent proton-coupled electron transfer steps by separating the regions of the Pourbaix diagram with vertical, horizontal, and diagonal boundary lines, respectively. These properties can be useful thermodynamic descriptors for catalysis. Second, the boundaries between different regions of a Pourbaix

diagram define theoretical electrochemical conditions where free energies of reaction for a (de)hydrogenation step are zero, and thus at those electrochemical conditions the process should be highly reversible and thus energetically efficient. *Pourbaix diagram boundary lines therefore show theoretical electrochemical conditions that a species would facilitate energetically efficient shuttling of protons and electrons.* One step further, if one considers a Pourbaix diagram for a reactant such as CO_2 (Figure 2a) and another Pourbaix diagram for a hypothetical catalyst (Figure 2b), one could then overlay the one on top of the other (Figure 2c). Regions where boundaries of the two Pourbaix diagrams overlap signify electrochemical conditions where one species (i.e. a hypothetical catalyst) would facilitate shuttling of protons and electrons to another species (i.e. a reactant). This can be thought as an extension to the Sabatier principle of catalysis, where optimal catalyst activity is achieved when the substrate binds strongly enough to be activated but also weakly enough that it can still be removed and not poison the catalyst. Thus, Pourbaix diagram analyses allow one to search for the *catalyst state under specific electrochemical environments* that would provide the lowest hypothetical overpotential.

Our group has used Pourbaix diagram analyses to study a variety of homogeneous and heterogeneous catalysis systems for CO_2 reduction. Interestingly, we have predicted that reaction conditions for several CO_2 electroreduction processes ranging from homogeneous pyridinium[39, 40] and homogeneous ruthenium-complexes[41] as well as heterogeneous N-doped nanocarbons[42] and partially reduced SnO_2 oxides.[43] All coincidentally share a similar characteristic, all have Pourbaix boundary lines showing the formation of a new intermediate state near the conditions where CO_2 electrocatalysis has been reported. Experimentally validating these computational predictions has been difficult, in part due to difficulties reproducing experimental data that has been reported in the literature.[44, 45] However, other experimental studies have implicated transiently formed hydride-containing species in CO_2 reduction that are intermediates predicted to be thermodynamically stable by Pourbaix diagram analyses.[46, 47] We see opportunities to use computational modeling to discover new catalysts in chemical and materials space and synergistically guide experimental design with high-throughput screening. However, numerous hurdles pertaining to modeling reaction mechanism under solvating reaction conditions must be overcome first.

6

1.2.2 Challenges of Modeling Electrochemical Reaction Mechanisms

As stated earlier, Pourbaix analyses require that all the salient reaction intermediate states to be correctly identified. When modeling catalytic reactions on surfaces, especially gas phase reactions on conducting surfaces, standard Kohn-Sham density functional theory (DFT) is normally suitable for reliably modeling charge neutral reaction intermediates. Additionally, modeling electrochemical reactions using the computational hydrogen electrode model[48] (i.e. modeling electrochemical proton and electron transfers as a $1/2$ H_2 transfer coupled to a linear potential correction) can bring helpful and testable insights into electrocatalysis. However, as illustrated by Exner and Over[49] as well as Janik and Asthagiri,[50] modeling reaction mechanisms without accounting for barriers provides an incomplete picture and can result in qualitatively different outcomes that might be wrong and/or misguide future research efforts.

Carrying out thorough computational investigations is easier said than done. Calculating barrier heights requires substantial computational effort, and these efforts would all be for nothing if an unphysical model system were used. First, simplistic models are never guaranteed to represent the actual atomistic environment, though understanding model systems can provide useful insight into which pathways are feasible and which are unlikely. Adding to this complexity, it is well known that commonly used DFT approaches have self-interaction errors that make them sometimes unphysically model charged intermediate states and/or highly correlated systems,[51, 52, 53] and so higher-level theories are required. Today, we see most development and applications in this area are using models that 1) enable enhanced sampling of reaction mechanisms to identify meaningful reaction pathways;[54, 55, 56] 2) enable physical modeling of electrochemical (i.e. potential dependent) reaction mechanisms;[57, 58, 59, 60, 61, 62, 63, 64] and 3) improve the quality of continuum solvation energies of static systems.[65, 66, 67, 68] There is a growing understanding that solvation is important not just in homogeneous catalysis but also heterogeneous catalysis.[69] Also, solvation modeling treatments are sometimes revealed to not be as reliable as generally believed.[1] While some computational studies are starting to explicitly account for potential-dependent reaction mechanisms in different forms, there has been little

consensus of the best practices for doing so. All of these challenges are important, and the pathway to addressing them will likely be coupled. To better understand these challenges through the lens of solvation, we briefly summarize different solvation modeling techniques.

1.3 Implicit Solvation

Continuum solvation models have been used for many decades and there are many detailed reviews in the literature explaining the theory and the applications.[70, 71, 72, 73, 74, 75] We only briefly overview how continuum solvation models work and how they are used to describe renewable energy catalysis. Figure 3 shows a cartoon model representation of implicit solvation of a methanol molecule with cluster and surface calculations.
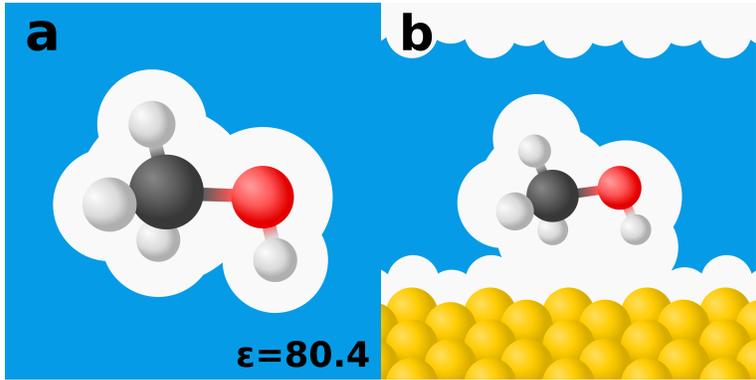
Figure 3: a) Illustration of a methanol molecule modeled within a cavity of a non-periodic continuum solvation model. b) Illustration of a methanol molecule modeled at a surface within a cavity of a periodic continuum solvation model.

Continuum solvation models were first developed for non-periodic systems of small and neutral molecules, and most treat the solvent as a structure-less, homogeneous medium using a polarizable dielectric described by a dielectric constant ε. In the most commonly used methods, a solute cavity is created around a solute to represent a boundary surface that allows a semiempirical calculation of a solvation energy based on the electronic structure of the system of interest. The subtle differences in defining the cavity, the theoretical foun-

8

dations, and the boundary conditions are what gives rise to the various implicit solvation models.[76, 77, 78, 79, 80, 81, 82, 83, 84] For instance, one of the first and most widely used continuum solvation model is the conductor-like screening model (COSMO),[85] which differs from other models by employing a scaled conductor instead of exact dielectric boundary condition, and this approximation considerably simplifies the mathematics. Also, COSMO uses a Green's function as the dielectric operator, and that enabled it to be the first continuum solvation model that was implemented with analytical gradients and used a real cavity shape.[86]

Expanding the applicability of implicit models to periodic systems requires treatment of the ionic response of charged species and interfaces.[87] Fattebert and Gygi were the first to make an isodensity continuum model adaptable to periodic systems that would be appropriate for modeling solvation on surfaces.[88] The simplest way to treat the ionic response is by using Poisson-Boltzmann (PB) theory which considers ions as point particles with mean-field interactions.[87] There are many other ways to treat a solvent implicitly on surfaces and still account for ionic responses. For example, joint DFT was developed to combine typical electronic DFT with a classical DFT description of the liquid environment in order to reduce computational costs of large periodic systems.[89] This was first described by a modified polarizable continuum model (PCM) that has a linear dielectric response for the solvent (linearPCM).[90] The linear dielectric response approximation tends to fail with systems containing strong electric fields like ionic surfaces and electrochemical systems, so Gunceler et al. developed an improved model by using a nonlinear dielectric response (nonlinearPCM).[90] Alternatively, the self-consistent continuum solvation (SCCS) was developed to extend the utility of implicit solvation to plane–wave codes with improved robustness.[91, 92, 93] More recently, the CANDLE method was explicitly developed to handle charged species because it takes into account the charge asymmetry in the solvation structure. In this method the cavity is defined by a nonlocal functional of the solute electron density and potential that enables modeling the system's asymmetric solvent charge.[94] Additionally, separate field–aware approaches are being developed for cavity descriptions that can account for charged species without the need of continued modulation of cavity definitions to improve experimental fitting.[95]

There are several open challenges associated with applying implicit solvation models for periodic systems. In particular, it remains challenging to reliably determine electrochemical interfacial structures as well as reaction energetics. For example, some models cannot capture the local field variations from cations and in some cases default parametrization can place the ionic countercharge unphysically close to the surface.[96] Recently, there has been a number of exciting developments in implicit solvent modeling by improving numerical stability and reducing unphysical artifacts of cavities to better describe the electrochemical environments. For example, Fisicaro et al. used a continuous permittivity to model complex dielectric environments or electrolytes that should be accurate for neutral and charged systems.[97] Also, Andreussi et al. have developed an improved continuum solvation model that eliminated unphysical cavity "pockets" by smoothly varying solute cavities.[98]

Overall, one of the main purposes for an implicit solvation model is to avoid the complexity and computational cost of explicitly modeling solvent molecules. The computational expense for these systems is low and thus these methods are among the most used in applied studies of reaction mechanisms. Continuum solvation models such as COSMO, PCM, and the more recent solvation model based on density (SMD)[84] are highly cited because they are often used in diverse applications including reaction mechanism studies. We now will discuss a few applied studies in detail, but mention several others studies that have employed implicit solvation models to study aqueous CO_2 reduction.[99, 100, 101, 102, 103, 104, 105, 106, 107, 108]

Note that modeling extended surfaces are more physically representative of an actual surface, but being able to model surfaces as clusters can sometimes make it easier to introduce high level theory. However, finite clusters can also have complicated spin states that need to be accounted for (e.g. Ref. [109]) while periodic analogs to these systems may not have significant spin polarization. To understand the extent that continuum solvation models can and should be used in applications of surface cluster models, Gray and co-workers computationally modeled adsorbate binding energies under the presence of continuum solvation on both periodic slab and large cluster models.[110] They modeled the Pt(111) surface with a variety of adsorbates: H*, O*, and OH* at different binding sites. It was found that sufficiently large model clusters captured similar gas phase binding energies as those ob-

tained using periodic calculations and having relatively low surface coverages. Since the two fundamentally different models gave similar gas phase binding energies, it showed promise for future work modeling heterogeneous catalyst sites using modern QM-in-QM embedding models.[111, 112]

Once the gas phase energies were benchmarked for these systems, the energy contributions from continuum solvent methods could then be accounted for. Interestingly, using the COSMO model on the finite cluster resulted in calculated solvation energies ranging from -0.6 eV to -0.9 eV, and these were quite different in magnitude compared to the VASPsol energy contributions using the periodic systems that ranged from $+0.1$ eV to -0.35 eV. This should not be surprising since the surface cluster model had unphysical corners and edges that were being solvated while the periodic slab model had no unphysical corners or edges. The net effect of this was significantly different solvated adsorbate binding energies even though the gas phase adsorbate binding energies between the two models had been found to be similar. However, we also found that the relative solvated adsorbate binding energies were similar across different sites for both the surface cluster and periodic slab models. Since the relative energetics were similar, we concluded that reaction mechanism studies using continuum solvated surface cluster models probably will give similar insights as studies using continuum solvated periodic slab models. The salient point is that if reaction mechanism studies necessitate the use of solvated surface cluster models, it will likely be the case that continuum solvation energies will be less physically relevant, but error cancellations can be leveraged to give useful insights. However, when an intermediate state is being modeled that is different from the rest, the results from a continuum solvation model should be considered with more suspicion and thus warrant additional care to ensure that the solvation model is appropriate for that case.

Another important aspect with continuum solvation models is their cavity definitions. Programs such as GAUSSIAN allow the user to select different cavities based on different empirical radii, and Yang compared some of these models on homogeneous metal complexes for CO_2 hydrogenation.[113] Yang modeled PNP-ligated metal pincer complexes for formation of formic acid from CO_2 and H_2. To model solvation effects the integral equation formalism polarizable continuum model (IEFPCM) was used with van der Waals (i.e. Bondi

11

radii[114]) atomic radii–for geometric optimizations–and United Atom Topological Model (UAKS)–for electronic energy corrections–to describe the cavity. UAKS is based on a model where hydrogen atoms are always enveloped within the molecular cavity while hydrogen atoms from Bondi radii cavities will appear in the cavity surface. Yang compared solvation energies of small ions and found that solvation energies using the UAKS radii were more accurate than energies using Bondi atomic radii. It was found that UAKS cavity data were within 5 kcal/mol of experimental data while Bondi radii cavity data had an error of 16 kcal/mol. While UAKS radii have been shown to be useful in many applications, for instance when predicting pK_a values,[115] most benchmarking has been done for assessments of thermodynamic properties and reaction energies, but much less work has been done in understanding their applicability for determining kinetic barriers. In the cases of modeling (de)hydrogenation processes, it is not yet understood whether one should use a solvent model that explicitly accounts for hydrogen atoms or not. What is understood is that highly parameterized continuum solvation models are clearly very sensitive to cavity definitions, and tuning any specific radii for any specific application should be avoided.

Koper has also studied numerous mechanisms for CO_2 and N_2 reduction.[116, 117, 118, 119, 120] For example, his group has studied CO_2 reduction mechanisms involving cobalt porphyrins,[121] and they identified CO as being the main product from this reaction mechanism and CO_2^- as the key intermediate. Co(P) guided the formation of CO through decoupled proton and electron transfers; however, additional concerted proton–coupled electron transfers involving CO resulted in minor CH_4 formation. This work was made possible using the COSMO implicit solvation model to account for solvation effects. One complex modeled during the reaction, [Co(P)—(CO2)]$^-$, was only stable when solvation treatments were included; however, another complex was still not stable when implicit solvation was included in the calculations ([Co(P)—(CO$_2$)]^0). An analogous observation was also seen in work by Carter,[35] who modeled an anionic complex, [Re(bpy)(CO)$_3$-(CO$_2$)]$^-$, and found it was only stable with an explicit counter ion or under the presence of a continuum solvent method. Thus, continuum solvation models have been and will likely continue to be used to assess metastable (and potentially zwitterionic) reaction intermediates in homogeneous reaction mechanisms.

With the success of implicit models in previous studies, many researchers are attempting to apply these techniques to reactions involving solvent mixtures. Garza et al. studied a tetraaz $[Co^{II}N_4H]^{2+}$ catalyst to understand the selective reduction of CO_2 to CO.[122] Those authors used PCM to include the solvent effects. They used pure acetonitrile in their calculations although the experimental contributions used a wet (10 M water) acetonitrile environment. Mixed solvents present a challenge for computational modeling since only a few models such as COSMO-RS can be used to model mixed solvents, and this model has not yet been as extensively used for mechanistic investigations as the conventional COSMO approach.[123] Garza et al. modeled both pure H_2O and acetonitrile systems and noted that their calculated reaction energies do not differ significantly between these two solvents. Those authors then inferred that mixed solvents would also not be significantly different even though experimental data has shown that mixed solvents can bring peculiar and non-intuitive solvation energies depending on the solute and the mixed solvent composition.[124, 125] From our perspective, since continuum solvation models generally cannot be trusted to recognize the significance of an explicitly bonded solvent molecule, they should not be assumed to be a physical model for any mixed solvent in an arbitrary solvent composition. It is true, however that any errors arising from an insufficient solvation treatment of any one intermediate might cancel out with errors from a different intermediate, and thus the relative energy difference between the two would be reasonably accurate due to fortuitous error cancellation.

Another study by Cao et al. considered Ir(III) pincer dihydrides as electrocatalysts for CO_2 reduction to formate (or formic acid) in acetonitrile/H_2O mixtures.[126] They used IEFPCM with UAKS radii and cavity-dispersion-solvent-structure terms from the SMD solvation model to describe the solvation effects using the GAUSSIAN code. Experiments show that the reaction does not happen in anhydrous acetonitrile and that a water concentration of 5% or more is needed. As with the study by Garza et al, these authors used continuum solvation models to gain insights into chemical reactivity in pure H_2O and acetonitrile solvents. They mainly discuss reaction pathways under acetonitrile because the experimental conditions had a higher percentage of acetonitrile; however, almost all of the calculated barriers are very similar in magnitude compared to calculated barriers in pure H_2O. The barrier for formation of the formate anion appears to have lower energy when it is modeled in water,

which indicates that water explicitly plays an important effect in this reaction mechanism by forming hydrogen bonds with the formate.

To summarize this section, we note that continuum solvation models are very useful, but they are sometimes unreliable and thus should be used cautiously when making predictions. Users should be aware that modeling and comparing different solvents, such as water and acetonitrile, generally only involve a slightly different cavity definition and dielectric constant that may result in a relatively small solvation energy difference. As a result, it should not be surprising when a continuum solvation model gives similar solvation energies for different solvent systems. However, mixed solvent systems are known to exhibit non-linear effects as a function of solvent composition, and standard continuum solvation models have not yet reproduced this behavior.[124, 125]

1.4   Mixed Implicit/Explicit Solvation
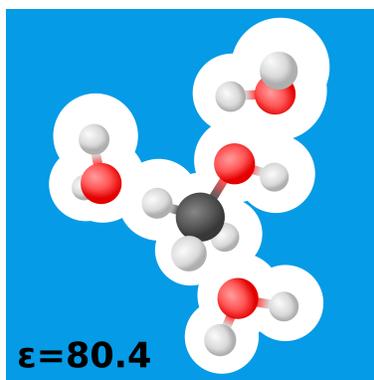
Figure 4: A model cluster with three explicit solvent molecules and implicit solvation.

One technique to improve the performance of continuum solvation models is with so-called mixed implicit/explicit or cluster–continuum solvation modeling, which has been used in practice in an ad hoc manner for decades.[127] Instead of a lone solute being considered, some number of explicit solvent molecules are added to the system, and the resulting cluster of molecules is placed into the dielectric medium. In periodic systems of face-centered cubic metals, explicit solvent molecules are generally added as one or more layers of solvent

molecules and then an implicit solvation model can be used on top of that. On other surfaces one or more solvent molecules need to be added to the system in an ad hoc manner to build up an interfacial solvation structure. Mixed implicit/explicit solvation approaches usually used in calculations where an implicit solvation model is not sufficient to model a system of interest. For instance, mixed implicit/explicit solvation is used to predict energy calculations of ions and/or small molecules,[128, 129] though it is also used for studying reaction mechanisms that involve the participation of the solvent molecule. A model cluster is shown in Figure 4 with three explicit solvent molecules and implicit solvent.

The main challenge of mixed implicit/explicit solvation modeling is to know how many solvent molecules are required to capture the crucial solvation effects and where to place those solvent molecules in a meaningful way. The most commonly used way to overcome this challenge is to place solvent molecules according to chemical intuition and/or with trial and error attempts. This requires *a priori* knowledge of the reaction mechanism and the active sites that need to be stabilized. Even if one can place the solvent molecules with chemical intuition, there is still the open question of how many solvent molecules are needed. Furthermore, one should keep in mind that an entropic penalty would be expected when forming solvent clusters, and that might play an important role in interpreting calculated energies.

Different research groups show different preferences about determining how many solvent molecules are needed for an accurate calculation. Some will only add a single solvent molecule at the site of interest while others may add more solvent molecules until a desired result is achieved. Ahlquist studied CO_2 hydrogenation with a homogeneous iridium catalyst using two explicit water molecules together with Poisson-Boltzmann self-consistent reaction field as defined in the Jaguar simulation package.[130] Ahlquist reported agreement with the experimental values only when both implicit solvation and two water molecules are present in the system. Groenenboom et al. modeled thermodynamic descriptors for a large set of aromatic N-heterocycle molecular catalysts for electrochemical CO_2 reduction.[40] Across 27 different molecular catalysts, using one explicit water molecule located at the relevant hydrogen bonding site for each molecule improved direct pK_a calculations to reasonably low errors of about 1 pK_a unit.

For reaction mechanisms, including explicit solvent molecules plays an important role as well. Lim et al. studied hydride transfer pathways from dihydropyridine to CO_2 by including one or two explicit water molecules together with CPCM model in their system.[131] Those authors found that this was an adequate treatment of the solvent because the resulting polar transition state structure was substantially stabilized by explicit solvent molecules that also facilitated a proton shuttle mechanism. Those authors also looked at a similar system where they used pyridine to catalyze CO_2 reduction by using different degrees of solvation. In this study they considered up to three solvent molecules as participating in their reaction mechanism as well as up to ten more solvent molecules to further solvate the reaction-relevant molecules, and the entire cluster was then embedded in CPCM implicit solvation model. The authors reported good agreement with experimental values when they used three solvent molecules in the active reaction mechanism and ten solvent molecules to solvate the core structure (calculated: 13.6 kcal/mol; experimental: 16.5 kcal/mol). While the computational results may or may not reflect the actual mechanism, they do highlight the important role of proton shuttling networks that standard continuum solvent models (as well as explicit solvent molecule using classical force fields) would not be able to physically model.[132]

Lim et al. also revisited the pteridine molecule [133] that had been proposed as a potential CO_2 reduction catalyst with some controversy.[44, 134] The authors' model system included seven water molecules and the entire cluster embedded in the CPCM model. They then benchmarked results from this implicit/explicit solvation modeling treatment to QM/MM simulations (*vida infra*) where the seven water molecules were kept in a QM region and the rest of this cluster was explicitly solvated with 200 water molecules treated using a classical force field. The authors found that the two solvation treatments resulted in very similar energies, and they also found that the reaction barrier was consistently too high to be valid for a reaction that would be expected to occur at room temperature (QM/MM: 29.7 kcal/mol, QM: 30.9 kcal/mol). Savéant has commented that QM calculations were not necessary to rule out some pathways,[44] but Lim et al.'s work is nevertheless useful because it demonstrates that simpler cluster continuum models can provide similar results as far more computationally intensive QM/MM simulations and thus suggesting other means

forward for modeling these systems besides computationally costly QM/MM simulations. The important role of solvent molecules in reaction mechanisms is not only limited to just H_2O. Rohmann et al. studied CO_2 reduction to formate with a homogeneous ruthenium complex. They modeled their system in DMSO solvent using 10 explicit solvent molecules together with SMD solvation model. They show DMSO solvent molecules are vital for the mechanistic study because the hydrogen bonding between the formate (the end product) and the solvent results with a thermodynamic driving factor for desirable concentrations of the products.[135]

There are far fewer studies on mixed implicit/explicit solvation on periodic surfaces. Carter has studied CO_2 reduction on GaP (110) surface by modeling it as a cluster that can be straightforwardly solvated with a non-periodic solvation model,[136] similar to the work by Gray et al. mentioned previously. Their treatment used structures arising from a full monolayer of half-dissociated water molecules together with the SMD solvation model. They identified 2-pyridinide as an active intermediate in Py-cocatalyzed CO_2 reduction at p-GaP photoelectrodes.

As stated before, there is no easy way to determine how many solvent molecules are needed for an accurate and reliable treatment of mixed implicit/explicit solvation. As a test to deconvolute the relative energy contributions of electronic correlation, explicit solvation, as well as the presence of a counter ion in a reaction mechanism, Groenenboom and Keith followed work by Johnson[137] who studied borohydride hydrolysis using a procedure involving high temperature Born–Oppenheimer molecular dynamics (BOMD) simulations to observe an elementary hydrogenation process and then characterized that pathway using nudged elastic band methods.[138] Groenenboom and Keith used a similar procedure to model CO_2 reduction by $NaBH_4$ and $NaBH_3OH$. Molecular clusters from the NEB calculations were then used with different analyses using high-level single point energy calculations and implicit solvation. In general, it was found that the full first solvation shell along with COSMO implicit solvation resulted in an energy profile almost identical to the fully explicit solvated case. Somewhat surprisingly, a range of different levels of theories found calculated barriers differing by only 0.1 eV while using a continuum solvation model without the first solvation shell resulted in differences as large as 1 eV. This study points out the importance

17

of the solvation treatment however using BOMD simulations together with NEB calculations can become very computationally expensive. It would be especially interesting if there were ways to sufficiently solvate reaction intermediates without the need for dynamics or even fully explicit solvation models.

1.5   Explicit Solvation

Many research groups explicitly solvate their systems to gain detailed information not available from implicit methods. Studies typically use Monte Carlo (MC) or molecular dynamics (MD) to treat the entire solvent box as shown in Figure 5. Still, complications could arise when studying polarizing systems or significant electron density changes. Born-Oppenheimer molecular dynamics (BOMD) and its variant Car-Parrinello Molecular Dynamics (CPMD)[139] have been critical in broadening the scope of systems we could study explicitly. Both use real-time electronic structure calculations to describe the system's behavior instead of parameterized force fields or potentials; however, they are only meaningful if the run time is long enough for the system to visit all energetically relevant configurations. For complicated systems, large energy barriers could separate chemically relevant configurations and severely limit sampling.

Currently there are a couple of ways to avoid the high computational costs of BOMD. First is to use simulation schemes that are computationally faster. These methods often depend on reducing the frequency of full electronic structure calculations or simply reducing the region being treated quantum mechanically and employing a classical treatment for the remaining system. The latter solution is referred to as quantum mechanics/molecular mechanics (QM/MM) which is a hybrid method that combines QM and MM frameworks to make simulations faster than BOMD and more accurate than MM. In QM/MM simulations, the system is divided into primary and secondary subsystems.[140] The primary system is the QM region which contains the reaction-relevant molecules under investigation. The secondary subsystem is the environmental zone where the other solvent molecules are modeled with forcefields to capture the bulk solvation effects.
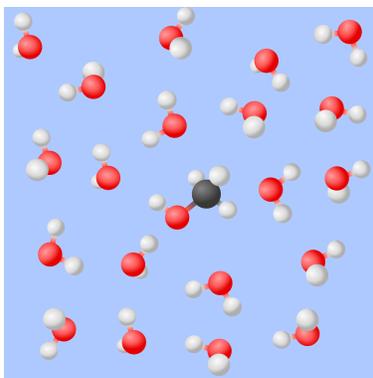
Figure 5: A methanol molecule being explicitly solvated by water.

It is common practice to include solvent molecules from the first solvation shell in the QM region to capture the crucial solvation effects using a higher level of theory. Although difficulties can arise when trying to keep the simulation as physically realistic as possible. Solvent molecules, in real solutions, will migrate towards and away from solute regions. This poses a problem in garnering expensive and highly accurate data on short-ranged solvation effects when a solvent molecule drifts away. Researchers sometimes employ constrained QM/MM; in which a bias is applied to keep solvent molecules from leaving the predefined QM region.[141, 142, 143] While this provides reasonable accuracy, the fundamental issue with this type of modeling is its unphysical treatment of an essentially frozen solvent shell. Alternatively, a method of switching the subsystem designation (QM or MM) of solvent molecules based on the proximity to the solute in real time can be used and is common practice today.[140, 144, 145] This adaptive QM/MM scheme is very useful, but it could still benefit from a reduction of spatial artifacts that affect multiscale modeling.[146] We expect to see substantially more applications of these methods in the coming years as they can allow higher levels of QM theory for improved insights into catalytic reactions.[147, 148]

1.5.1 Sampling Techniques

QM/MM free energy simulations are commonly used to sample free energy surfaces. In renewable energy catalysis, a reaction often needs to be modeled by bond breaking or forming. In order to model such catalysis one needs to treat the system with quantum chemistry. However, the calculations will become very expensive if the entire system is treated with quantum chemistry, i.e. using some variant of BOMD. To overcome this challenge, algorithms are applied to enhance the sampling of reaction-relevant areas of free energy surfaces. These algorithms can vaguely be distinguished into two categories as, methods that introduce additional degrees of freedom along which the free energy is calculated (metadynamics) or methods that sample the system in equilibrium (umbrella sampling). In the following sections we will broadly introduce one technique from each category.

1.5.1.1 Metadynamics   Metadynamics is a sampling technique that is based on adding an additional bias potential that acts on a selected number of collective variables (CV). For reactive systems, bond breaking or bond forming are two examples of widely used collective variables.[149, 150, 151, 152, 153] To accomplish this, Gaussian potentials are placed on the free energy surface in order to flatten the energy wells and reduce oversampling of local minima. A very simplified representation is shown in Figure 6. It is an accelerated sampling technique of rare events that is based on pushing the system away from the local minima. Metadynamics is generally used to explore new reaction pathways without *a priori* knowledge of the free energy surface. However, one must be careful to identify a set of CVs appropriate for describing complex processes.[154] CVs should be a function of the microscopic coordinates of the system and should distinguish between the initial and final states while also describing relevant intermediates. If one can come up with CVs that meet all the requirements then metadynamics should work effectively to model free energy surfaces.[150]

There are handful of examples where *ab initio* metadynamics is used to study CO_2 reduction or any reaction mechanism. Urakawa et al. was exploring a ruthenium dihydride catalyst and its ability to hydrogenate CO_2.[155] Their work demonstrated that a *trans* isomer route
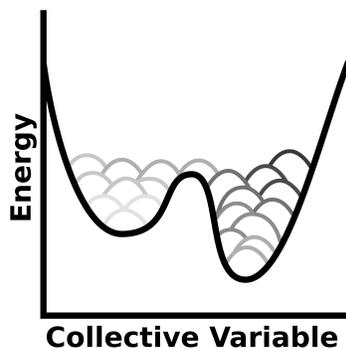
Figure 6: A model metadynamics simulation profile. Gaussian functions are placed on the free energy surface to flatten the energy wells over time during the simulation (lighter to darker curves). This is used to reduce oversampling of the local minima and pushes the system away from it.

was more energetically favorable (higher stability intermediates) while the rate-limiting step was the insertion of H_2 into formate, but there were no explicit solvent molecules included in this study that would account for their role in the reaction mechanism. Ghoussoub et al. studied the effect of temperature on frustrated Lewis pairs on nanoparticles for heterogeneous catalytic reduction of CO_2.[156] They concluded that at higher temperatures, CO_2 adsorbed more easily on the surface which suggests an adsorptive reaction mechanism may be relevant. This study also did not consider how solvation can affect the reaction mechanism, but they investigated H_2O adsorption on the surface at different temperatures. Gallet et al. used metadynamics to simulate the reaction of CO_2 with one, two, or three explicit solvent molecules in the gas phase.[157] This work provides a useful and thorough protocol to study relatively small systems. Future advances of computation resources will continue to allow more extensive studies to be carried out.

There are few studies on CO_2 chemistry that used metadynamics with fully explicit solvation models. Stirling studied the free energy barriers of reversible bicarbonate formation in water at high pH.[158] It was determined that the free energy barrier of $CO_2 + OH^- \longrightarrow HCO_3^-$ was 13.8 kcal/mol, which coincides with the 11.5 kcal/mol experimental value. Interestingly, the forward reaction free energy barrier was mostly entropic while the reverse barrier

was mostly enthalpic. This conclusion was only possible because extensive metadynamics simulations had been performed with explicit solvent. This study outlines an accurate way to calculate free energy barriers of other processes in solvated systems; however, the number of reacting atoms that need to be considered will be a limiting factor. Galib et al. also examined the mechanistic and energetic effects of solvent cluster size on the decomposition of H_2CO_3.[159] They selected atoms to form two small (6 and 9) and two large (20 and 45) water clusters around a H_2CO_3 molecule in a Car–Parrinello molecular dynamics simulation. Metadynamics then allowed sufficient sampling to demonstrate that the small and large clusters led to a concerted and stepwise mechanism, respectively. Thus, H_2CO_3 decomposition likely follows a stepwise mechanism in bulk-like water, but it might be different in other environments like an air/water interface. Goddard and co-workers have investigated multiple aspects of CO reduction on copper surfaces and copper nanoparticles with explicit water layers at different pH levels.[59, 160, 161, 162, 163] Their studies of solvated systems were carried out using reactive force fields which significantly decrease the computational time required. However, even well-parameterized reactive potentials should be assumed to be less accurate than the QM calculation, and thus interpretations based on predictions from these model warrant more caution than all-QM methodologies.

1.5.1.2 Umbrella Sampling Umbrella sampling is another technique to calculate the free energy profile of reaction mechanisms.[164] The main idea behind umbrella sampling relies heavily on splitting the reaction pathway into windows and sampling each window individually. However sampling a full momentum space is difficult, and that is why a bias potential is introduced as an additional term to the energy expression as shown in Equation 1.5.

$$E^b(r) = E^u(r) + \omega_i(\xi) \tag{1.5}$$

This additional term ensures efficient sampling along the reaction pathway by allowing the reaction variable to vary along a biased potential (restrain) and not limiting the variable to a constant value (constrain). The most commonly used biased potential is the harmonic potential as shown in Equation 1.6.

$$\omega_i(\xi) = K/2 * (\xi - \xi_i^{ref})^2 \tag{1.6}$$

Harmonic potentials are appealing because they contain only few parameters: K (spring constant), the number of images (i), and a reference point of the respective window i (ξ_i^{ref}). One needs to decide on the K value before starting the simulations, and make sure it is large enough to drive the system over the energy barrier.[165, 166] This is important because if K is too large there will be too narrow sampling and thus sufficient overlap between the windows will not be achieved. An example of overlapping windows is shown in Figure 7. Having adequate overlap is required to analyze umbrella sampling with weighted histogram analysis (WHAM) or umbrella integration which depends less on overlap but is still advantageous.[167, 168, 169]

Umbrella sampling is widely used for physical transformations from ion solvation to protein folding with force fields;[170] however, modeling chemical reactions is more computationally extensive since it generally requires BOMD simulations. Leung et al. computationally examined a cobalt porphyrin catalyst for CO_2 reduction to CO in water.[171] First they used DFT calculations with implicit solvation and then validated their results with BOMD simulations with an explicit aqueous environment. These simulations demonstrated that the water molecules stabilized the reaction intermediates. With the use of potential of mean force (PMF) calculations they were able to identify the rate limiting step as the transfer of electrons between the polymerized catalyst and the gas diffusion electrode. This study is a good example of how to use umbrella sampling to calculate free energy barriers, however it is very computationally expensive and it limits the number of reactions that can be studied.
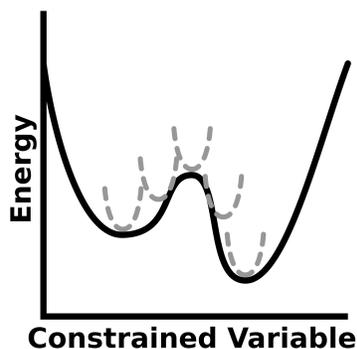
Figure 7: Simplistic view of umbrella sampling along a hypothetical constrained variable.

Several studies have been dedicated to understanding the hydrophobicity of aqueous CO_2; however, many employ classical force field methods which demonstrate the sensitivity to Lennard-Jones parameters.[172] To reduce parameter dependence, Leung et al. performed BOMD simulations to investigate the solvation shell of CO_2 and other dissolution species in water.[173] Ultimately their computations supported the previously observed hydrophobic nature of CO_2 in water. Furthermore, they calculated the free energy change of bicarbonate formation from CO_2 and H_2O to be -9.8 kcal/mol which agrees with the -9.4 kcal/mol experimental value.

In work related to the previously mentioned CO_2 reduction with sodium borohydride, Groenenboom and Keith calculated reaction energy barriers from NEB calculations at 0 K and compared them to free energy barriers obtained at 300 K using PMF calculations from umbrella sampling simulations. They show two different free energy barriers with NEB and PMF calculations which suggests both temperature effects and solvent molecules would play an important role in this reaction mechanism. The NEB pathway obtained at 0 K only slightly differed from the PMF pathway, but energies along the two pathways were found to vary by as much as 0.25 eV. The overall barrier heights from the 0 K NEB calculations and the 298 K PMF calculations for three different elementary steps were quite similar as well. However, the overall reaction energies from the NEB and PMF calculations differed by as much as 0.6 eV when the NEB pathway was based on local minima and the PMF calculations sampled lower energy states.[174] Thus, PMF calculations based on umbrella sampling simulations appear to be more reliable for insights than NEB calculations alone, but PMF calculations are also far more costly.

1.6   Conclusion

We have given a review of recent and legacy approaches that are used to model reaction mechanisms under solvating environments. It will be critical to integrate solvation energy contributions and other environmental parameters into future high-throughput screening approaches, and so we give an overview of implicit, mixed implicit/explicit, and explicit solvation modeling approaches that would be needed to do so. Though already widely used, continuum solvation models still have room for improvement. Notably, few if any can reliably treat explicit solute-solvent bonding or solvation effects, and they should not be used to glean insights into systems involving solvent mixtures. There are still paths forward for computational modeling to use more robust (though computationally cumbersome) techniques that incorporate explicit solvation at least in part. In the absence of accurate forcefield parameters and/or computational resources to run lengthy BOMD simulations, mixed implicit/explicit procedures are a promising route for studying reaction mechanisms in complex environments. Future directions continue to point toward more mixed implicit/explicit modeling as well as the development of more accurate and physical continuum solvation models and explicit solvation models. These advances will help improve the quality of computational predictions that would guide the development of technologies for renewable fuels and chemicals.

2.0 A Paramedic Treatment for Modeling Explicitly Solvated Chemical Reaction Mechanisms

This work has been published as Basdogan, Y., and Keith, J. A. (2018). A paramedic treatment for modeling explicitly solvated chemical reaction mechanisms. *Chemical science*, 9(24), 5341-5346.

2.1 Introduction

Computationally modeling atomic scale chemical reaction mechanisms in solvents is often not trivial. Plata and Singleton's detailed study of the Morita-Baylis-Hillman (MBH) reaction[1] has underscored poor performances of CSM-based quantum chemistry modeling without explicit solvation. Harvey and Sunoj have since evaluated various quantum chemistry modeling schemes and assembled a mechanistic picture that agrees well with Plata and Singleton's reported mechanism.[2] Their calculations used the high-level correlated wavefunction method DLPNO-CCSD(T)[175, 176, 177, 178] for electronic energies and usually an explicit solvation treatment with molecular mechanics. These two important studies have explained the elementary steps of the acid catalyzed MBH reaction mechanism, demonstrated the importance of critically evaluating computational theory to experiment, and discussed the extent that computational modeling can be predictive.

Building from those studies, we show how one can model such a mechanism with an automatable and paramedic modeling procedure that is enhanced with chemical intuition but also lessens the need for it. The paramedic and static quantum chemistry procedure will be more computationally demanding than static studies using a CSM with no explicit solvent, but it can also be expected to require less computational effort than many dynamics-based schemes. We formulated the procedure by calibrating to previously reported studies on the MBH reaction in order to understand how to navigate modeling pitfalls that face static models for reaction mechanisms in solvents.

We first assumed that high-level DLPNO-CCSD(T) theory with a relatively large triple-zeta basis set should provide fairly accurate gas phase reaction energies. Thus, any apparent errors larger than a few kcal/mol in any reaction step would indicate significant errors in solvation energies. We note that interpreting results from CSMs is not trivial, and some have explained that special care is needed.[179] A standard remedy for inaccurate CSM calculations is the modeler to add one or more explicit solvent molecules to the modeled system to more physically describe charge densities and solute solvent interactions.[128, 127, 180] Unfortunately, knowing how many and where solvent molecules should be added is also not trivial unless one makes a priori assumptions about local solvent environments. Below, we show that the paramedic model is an automatable way to overcome these challenges.

We hypothesized that solvation energies from CSM models would improve if we systematically added explicit methanol molecules around each solute while taking special care to ensure that each microsolvated state was a reasonable approximation of a thermodynamically low energy structure. To test this, we modeled each intermediate from Plata and Singleton's MBH reaction (Figure 8) as a microsolvated cluster of solutes with n = 1, 2, 3, 4, 5, and 10 methanol molecules.
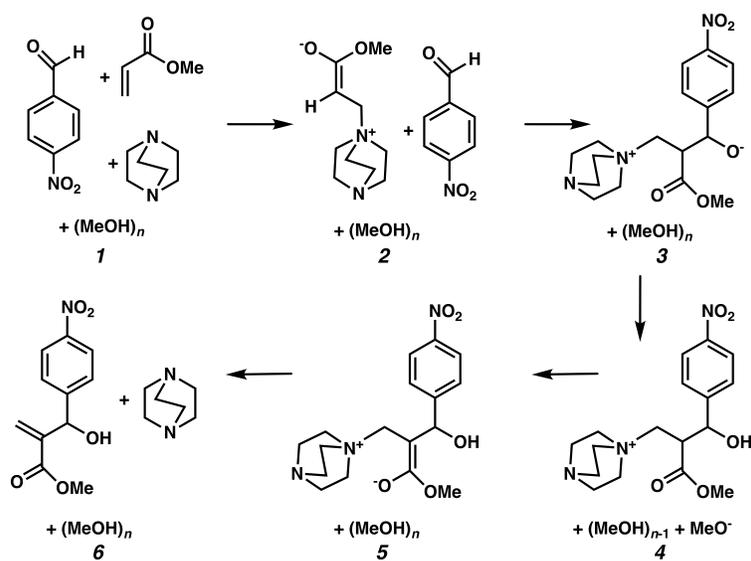
Figure 8: Mechanistic steps for the alcohol mediated MBH reaction, analogous to steps given in ref. [1].

2.2 Computational Methodology

In our study of the Morita Baylis-Hillman (MBH)[1] reaction mechanism, we used a filtering procedure where a global optimization code (ABCluster[181]) automatically generated 1,000 low energy candidates using CHARMM forcefield parameters from MacKerell's CGenFF website.[182] The 100 lowest energy structures from these cases were further optimized using semiempirical PM7[183] optimizations with MOPAC.[184] The five lowest energy structures from these cases were then optimized using Kohn-Sham density functional theory (DFT) at the BP86-D3BJ[185, 186]/Def2-SVP[187] level of theory with ORCA.[188] From now on, '-D3BJ' is shortened to '-D3'. We then compared the lowest energy QM-optimized structure using single point electronic energies at the same BP86-D3 level of theory, a hybrid functional (B3LYP[189]-D3[190]), and a high level ab initio method (DLPNO-CCSD(T)),[175, 176, 177, 178] each using the relatively large Def2-TZVP[187] basis set. Calculations made use of RI[191] and RIJCOSX[191] approximations when appropriate. We tested calculations accounting for extended solvation contributions using the SMD model (using B3LYP/Def2-TZVP calculations).

Our calculated reaction energies with no explicit solvent molecules followed an analogous procedure used by Plata and Singleton, where continuum solvation models were used and where thermal and entropic contributions for each solute was obtained using the full standard ideal gas, rigid rotor, and harmonic oscillator approximations. Low energy clusters with explicit solvent molecules were made for the six intermediates shown in Figure 8. Vibrational frequency calculations were carried out for the clusters with different number of methanol molecules to confirm there were no imaginary frequencies. Since each clustered intermediate had the same number and type of atoms, the vibrational, thermal, and entropic energy contributions from standard ideal gas, rigid rotor, and harmonic oscillator approximations can be expected to largely cancel out. In general, ΔE and ΔG values agree within about 6 kcal/mol, and these are differences that are much smaller than deviations shown from quantum chemistry modeling that does not account for explicit solvation.

All molecular dynamics (MD) simulations for the umbrella sampling were carried out using the TINKER[192] software. The simulations were carried out at 298 K in the NVT ensemble for 2 ns with 1 ps step size where the first 400 ps were used for equilibration and the remaining 1600 ps were used for data collection. In total 60 MD simulations were performed to scan the potential energy surface. In these simulations two constraints were in place: the distances between p-nitrobenzaldehyde and MA and p-nitrobenzaldehyde and DABCO. This was done by defining a harmonic potential between the center of masses of these molecules using a force constant of 100 kcal/mol. For each simulation, the distance between the molecules were varied between ∼4 to ∼15 Åwith a step size of 0.2 Å. At the end of the simulations, the distances were then calculated and the WHAM[193] analyses were carried out.

2.3   Results and Discussion

One might assume that intermediate **1** (shown in Figure 8) is not a good initial reference point for a reaction mechanism since reactant molecules are not infinitely separated and thus are interacting. To determine the free energy to form the cluster **1** from separated reactants in methanol solvent, we started with the lowest energy structure of **1** and then performed umbrella sampling simulations with classical forcefields using TINKER[192] simulation package to determine a quasi-static pathway that resulted in separated intermediates. Simulations used a cubic box starting with 500 solvent molecules. Dynamics simulations were run where the three intermediates were constrained at incremental intermolecular distances ranging from about 4 to 15 Å(see Figure 9). We then used the two-dimensional weighted histogram analysis method[193] to calculate the free energy profile along this pathway. We found a negligible free energy difference of about 1 kcal/mol to separate the three solvated reactants across this range of distances (see Figure 10). This confirms that this microsolvated cluster is in fact an appropriate reference point for the MBH reaction mechanism study. Future studies will help show if this is generally true for other microsolvated clusters of intermediate states in other reactions.
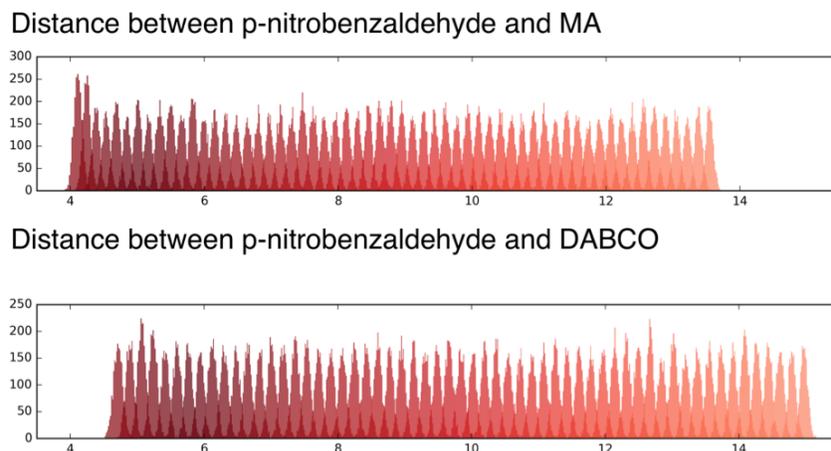
Figure 9: Overlap between US windows for aggregating different molecules together. Distances are defined as the distance between centers of mass for the three different molecules. The overlap of each window shows adequate sampling along the pathway.
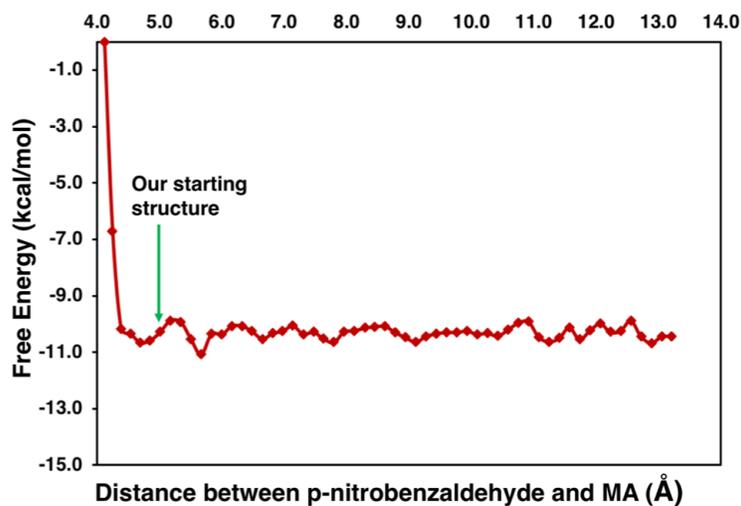
Figure 10: Free energy plot for aggregating three reactant species together. Distances are defined as the distance between centers of mass for p-nitrobenzaldehyde and MA, but simulations constrained distances between p-nitrobenzaldehyde and MA as well as p-nitorbenzaldehyde and DABCO simultaneously.

Figure 11 shows static quantum chemistry calculation data for each MBH reaction intermediate with different numbers of explicit methanol molecules. States labeled in red use a free energy calculation scheme with SMD continuum solvation and no explicit methanol molecules, analogous to the CSM-based model used in Plata and Singleton's study but now using DLPNO-CCSD(T)/Def2-TZVP electronic energies. We tested both SMD[84] and COSMO[85] solvation models and both provided effectively similar results. Figure 12 shows that differences between SMD and COSMO solvation models in these cases are small. Energies in Figure 11 are Gibbs free energies referenced to **1**. We note that once intermediates are clustered together, the relative free energies are quite similar to their respective relative electronic energies because the zero-point energies and other free energy contributions from IGRRHO approximations with the same number of atoms are similar (see Figure 13 for comparison of electronic and free energies).
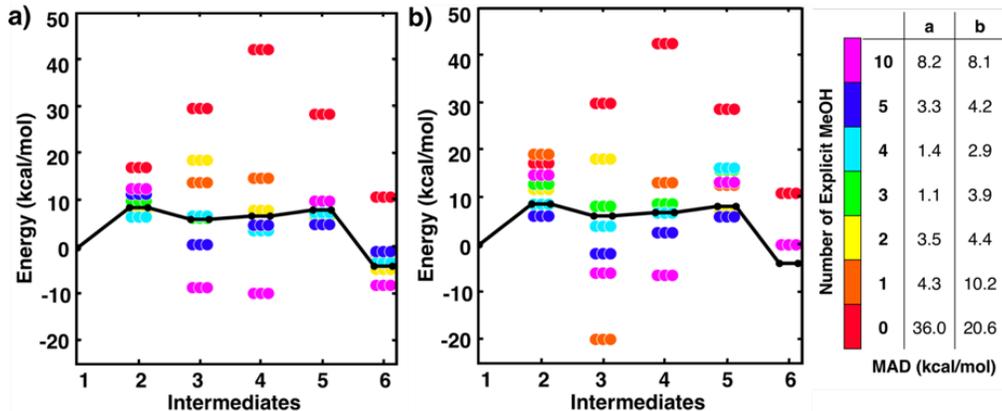
Figure 11: Free energies for MBH reaction intermediates (not including barriers) relative to intermediate **1**. Experimental data (black line) taken from ref. [1]. Data with '0' explicit solvent used a calculation scheme with SMD continuum solvation energies, analogous to ref. [1]. Relative free energies of clustered intermediates (a) without continuum solvation and (b) with continuum solvation. Mean absolute deviations (MAD, in kcal/mol) compared to experiment are reported in the table on the right. Energies are also tabulated in Tables 1 and 2.
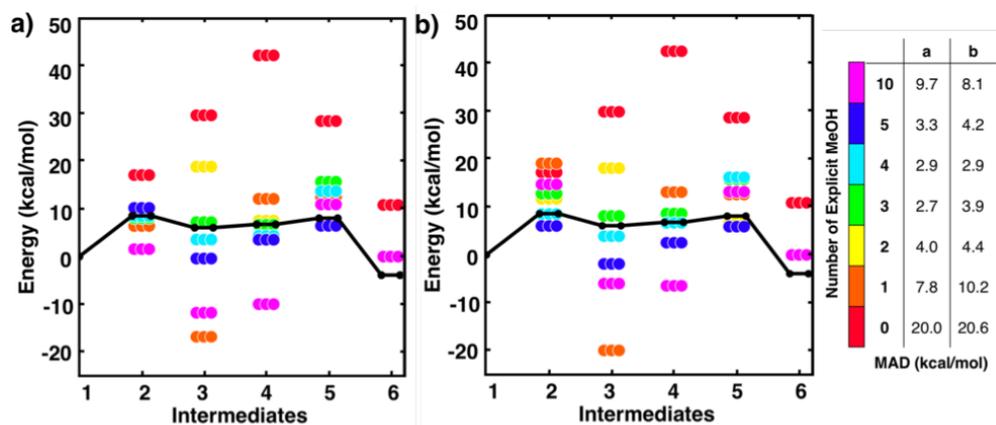
Figure 12: Energies for MBH reaction intermediates (not including barriers) relative to intermediate **1**. Experimental data (black line) taken from Ref. [1]. Relative (a) free energies with COSMO continuum solvation and (b) free energies with SMD continuum solvation. Mean absolute deviations (MAD, in kcal/mol) compared to experiment are reported in the right table.
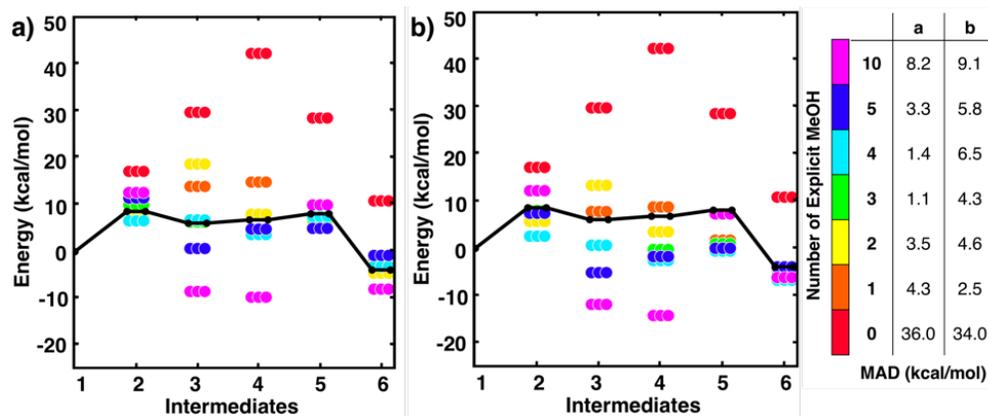
Figure 13: Energies for MBH reaction intermediates (not including barriers) relative to intermediate [1]. Data with '0' explicit solvent used a calculation scheme with SMD continuum solvation energies, analogous to Ref.[1]. All remaining calculations are in gas phase. Relative (a) free energies and (b) electronic energies of clustered intermediates. Mean absolute deviations (MAD, in kcal/mol) compared to experiment are reported in the right table.

Table 1: Free energies of clustered intermediates for microsolvated clusters relative to intermediate **1**.

# of explicit MeOH	Intermediate					
	1	2	3	4	5	6
0	0	12.1	26.5	152.7	18.0	3.5
1	0	8.12	14.0	14.9	5.2	-1.4
2	0	7.3	18.7	8.0	6.3	-4.6
3	0	10.0	6.3	4.9	7.0	-2.9
4	0	6.6	6.8	3.7	7.5	-3.2
5	0	11.5	0.7	4.8	4.5	-0.8
10	0	12.6	-8.5	-9.7	10.0	-8.0

Table 2: Free energies of clustered intermediates for clusters solvated using the SMD continuum solvation model relative to intermediate **1**.

# of explicit MeOH	Intermediate					
	1	2	3	4	5	6
0	0	17.2	29.8	42.4	28.5	10.9
1	0	19.1	-19.9	13.1	12.6	0.0
2	0	11.7	18.1	8.8	6.8	0.0
3	0	12.8	8.1	8.7	15.7	0.0
4	0	8.5	3.9	6.7	16.2	0.0
5	0	6.1	-1.9	2.5	5.9	0.0
10	0	14.7	-5.9	-6.4	13.2	0.0

Figure 11 thus showed that our initial hypothesis of gradually adding more solvent molecules into the system would improve agreement to experiment was categorically false. In fact, different numbers of explicit solvent molecules affect different states inconsistently. Furthermore, we unexpectedly found that including solvation energies via CSMs generally did not lower mean absolute deviation (MAD) to experimental data in any case compared to their respective gas phase calculations. Gas phase clusters with just three explicit methanol molecules had the smallest overall MAD, but we assumed this was due to error cancellation since adding more methanol molecules usually resulted in less agreement with experiment.

We then tested what might be causing errors that we attributed to solvation energy contributions. We hypothesized that different microsolvated clusters might have significantly different solute structures that then reflected different energies shown in Figure 11. To determine this, we analyzed the geometric similarities of solute structures using the Glosim [194] algorithm and the ReMatch-SOAP[194] kernel. Figure 14 shows the SOAP analysis for intermediates **2** and **3**. The SOAP analysis of different structures for **2** (Figure 14a) shows very high geometric similarities for all solute structures when gas phase optimizations were run with two or more explicit methanol molecules or when the structure with no explicit solvent was optimized using the SMD model.

In these cases, the C-N bond formed in the initial reaction step had a similar length ($R_{C-N} = 1.60$ Å). The gas phase optimized structures with and without one explicit methanol molecule had a significantly longer and unrealistic C-N bond length ($R_{C-N} = 2.72$ Å), showing those states had fallen downhill in energy into states best described as higher energy conformations of structure **1**. The SOAP analysis of structures for **3** (Figure 14b) showed the solute geometries within all the microsolvated clusters were highly similar regardless of solvent model. For the intermediate states **4** and **6**, the solute structures having two or more explicit methanol molecules were geometrically very similar to each other. The solute structures for **5**, however, were all found to be dissimilar across all cluster sizes (See Figure 15) and could not be interpreted. Apart from this exception, intermediates **2**, **3**, **4**, and **6** all had very similar respective solute structures (regardless of their respective microsolvated cluster sizes ranging from two to 10 methanol molecules). This shows that the 10 to 30 kcal/mol scatter in energies for each intermediate shown in Figure 11 is due to modeling

errors in solvation energy contributions that arise in static quantum chemistry calculations that all involve relatively small numbers of explicit solvent. Errors are present whether or not CSM modeling is used, but errors appear to be usually be larger when CSM models are used.
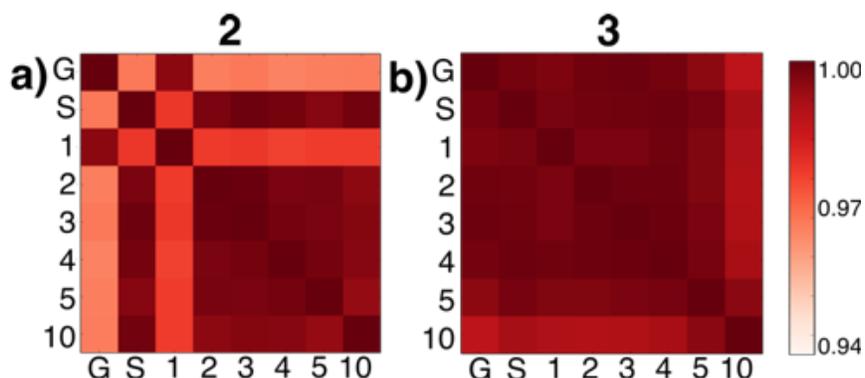
Figure 14: ReMatch-SOAP analysis on the solutes for clusters **2** and **3** with 0 methanol molecules ('G' represents a gas phase optimized structure and 'S' represents a structure optimized with SMD model), as well as 1, 2, 3, 4, 5, and 10 explicit methanol molecules. Colored boxes quantify similarities in different geometric structures (darker colors represent more similar structures).
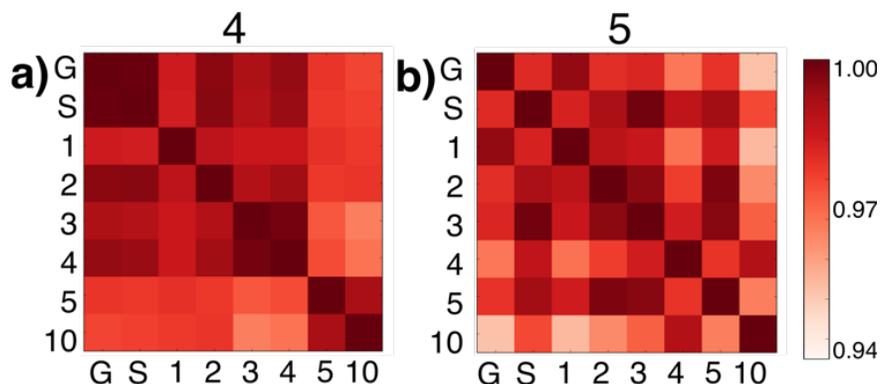
Figure 15: ReMatch-SOAP analysis on the solutes for clusters **4** and **5** with 0 methanol molecules ('G' represents a gas phase optimized structure and 'S' represents a structure optimized with SMD model), as well as 1, 2, 3, 4, 5, and 10 explicit methanol molecules. Colored boxes quantify similarities in different geometric structures.

To minimize errors in solvation energies arising from dissimilar local solvation structures, we then modeled the first bond-formation step of the MBH mechanism using Zimmerman's single-ended growing string method (GSM).[195, 196, 197, 198] We modeled pathways arising from the lowest energy configurations at each specific degree of solvation that was found using our filtering approach. Of course, one could also straightforwardly use this approach to model multiple pathways starting from different configurations of a single intermediate at the same specific degree of solvation (preferably with the smallest number of explicit solvent molecules). The only limitation to doing this is the higher computational cost of running multiple GSM calculations for each elementary step instead of just one.

All GSM pathway searches were performed with BP86-D3/Def2-SVP calculations with no CSM. We then calculated DLPNO-CCSD(T)/Def2-TZVP gas phase energies for the structures obtained from GSM calculations. Interestingly, transition states for systems with n = 1, 3, 4, and 10 methanol molecules each resulted in unreasonably large barriers ($\sim$40 kcal/mol), indicating an unphysical aspect with those microsolvated models for this step (more discussion below). Recall that the case with three methanol molecules also had the lowest MAD in reaction energies in Figure 11. In cases with n = 2 and 5 methanol molecules, more reasonable barriers of 23.6 kcal/mol and 17.6 kcal/mol were found.

Closer analysis revealed two points. First, the calculations with two and five methanol molecules both had explicit solvation interactions simultaneously at the two O atoms in **2** that undergo a tautomerization when forming **3**, while the other cases did not simultaneously solvate these two O atoms. (Note that these two atoms were also intuitively solvated by Harvey and Sunoj in their microsolvation models). Second, not only was the barrier with five methanol molecules is lower in energy, but it yielded the correct (S,R) isomer that was discussed in recent mechanism studies. The reaction with two methanol molecules had a higher barrier and resulted in an (S,S) isomer. Thus, the model with five explicit methanol molecules was the only case out of all solvation models considered that reasonably agreed with known experiment, and this model also resulted in a barrier height in reasonably close agreement with experiment (our calculation: 17.6 kcal/mol, experiment: 20.2 kcal/mol). For these reasons, this model system was the only microsolvated system used further.

We find that explicit solvation interactions are essential for an energetically feasible reaction pathway. However, modeling these interactions does not guarantee an experimentally observable pathway. Using this automatable modeling procedure allows one to see how and the degree that different explicit solvent configurations affect the same reaction step. It is also quite promising that the configuration resulting in the lowest energy pathway also yielded the same stereochemistry as found in previous studies. We see no reason why the configuration involving five methanol molecules is uniquely suited for this step, and one should at this point expect that alternate configurations with different numbers of explicit solvent molecules would also be in play and result in similar energy profiles.

Interestingly, this model for $2 \rightarrow 3$ resulted in a product state, $3'$, that was significantly higher in energy ($+7.5$ kcal/mol) relative to 3, the state that was found from our global optimization procedures. Not only was the barrier for $2 \rightarrow 3'$ ($\mathbf{TS}_{2\rightarrow 3'}$) is in reasonable agreement with experiment, but the overall energy of $3'$ was also in better agreement with experiment (calculated = 2.4 kcal/mol, experiment = 6.1 kcal/mol). Another single-ended GSM calculation starting from $3'$ was run to model the proton shuttling reaction that Plata and Singleton rationalized to be very fast, and this yielded a nearly barrier-less process leading to $4'$, which had reasonable energetics in agreement with experimental data (calculated = 3.5 kcal/mol, experiment = 6.8 kcal/mol). An additional single-ended GSM calculation found the pathway that tautomerized $4'$ into its enol form $5'$. The calculated barrier ($\mathbf{TS}_{4\rightarrow 5'}$, calculated = 17.3 kcal/mol) was also in reasonable agreement with experimental data (experiment = 21.2 kcal/mol), and the tautomerizing O atoms were again simultaneously interacting with explicit methanol molecules in this model. The relative energy of $5'$ was in reasonable agreement with experiment (calculated = 11.0 kcal/mol, experiment = 8.1 kcal/mol), while the energy of intermediate 6 was in very close agreement with experiment (calculated = 3.9 kcal/mol, experiment = 3.9 kcal/mol). Figure 16 summarizes our calculated pathway using five explicit methanol molecules and compares these data to the best calculated data from Harvey and Sunoj's study that used a combination of data from explicit and CSM solvation models. Hence, we have demonstrated a non-conventional, static, and automatable modeling scheme that identifies a complicated reaction mechanism with comparable accuracy as models using computationally demanding explicit solvation methods.
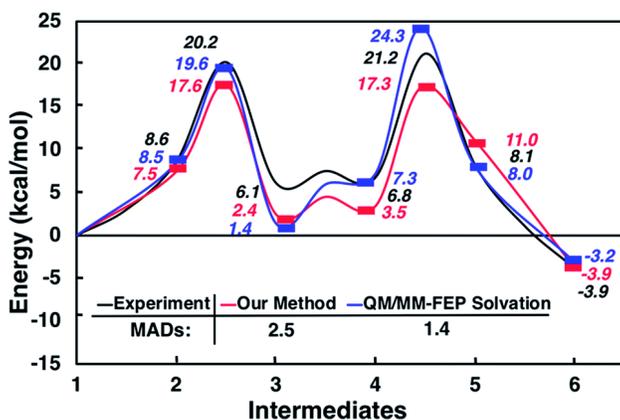
Figure 16: Reaction pathways obtained from GSM calculations (in red) compared to experimental data from ref. [1] (in black) and calculated energies from static and dynamics-based studies from ref. [2] (in blue).
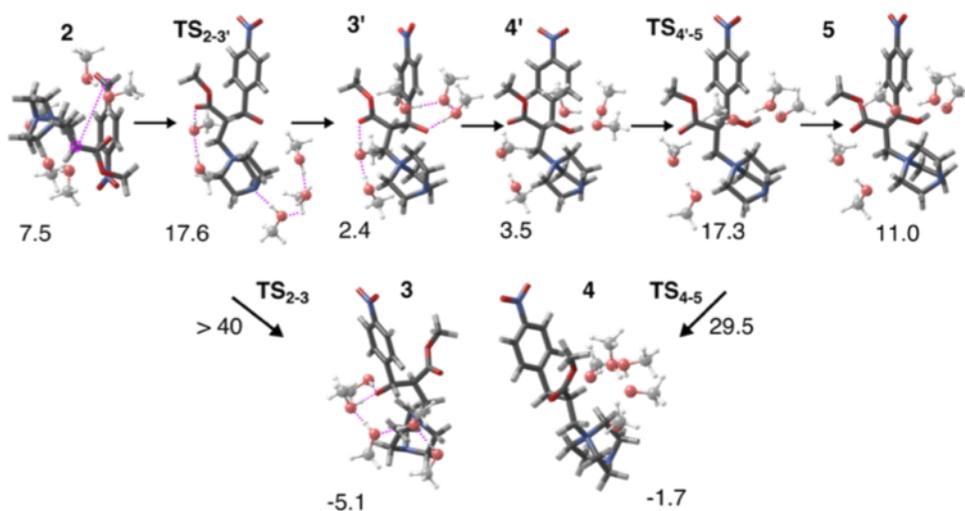
Figure 17: Structures for the MBH reaction pathway ($2 \rightarrow \textbf{TS(2-3')} \rightarrow 3' \rightarrow 4' \rightarrow \textbf{TS(3'-5)} \rightarrow 5$) highlighting the importance of local solvation stabilizing tautomerizing O groups. Though $3'$ is calculated to be higher in energy than 3, explicit hydrogen bonding opens a kinetically feasible pathway for C —C coupling. Reaction energies are reported relative to 1.

As another test for this MBH mechanism, we also ran double ended GSM calculations to identify barriers for $\mathbf{3'} \rightarrow \mathbf{3}$ and $\mathbf{4'} \rightarrow \mathbf{4}$ processes. Both barriers were greater than 28 kcal/mol and would be considered kinetically prohibited within this model. Figure 17 & 18 summarize these reaction intermediates and calculated data. Lastly, Figure 19 shows that reaction energies obtained using BP86-D3/Def2-TZVP (MAD = 3.7 kcal/mol) and B3LYP-D3/Def2-TZVP (MAD = 3.1 kcal/mol) single point energies are actually respectably similar to DLPNO-CCSD(T)/Def2-TVP calculations (MAD = 2.5 kcal/mol) as well as far less computationally demanding. We have thus shown calculation schemes using three very different levels of computational theory that are all consistent with each other and are significant improvements over results using a CSM with no explicit solvation. Consistent with our previous study on CO_2 reduction by borohydride in water[199], when modeling sufficiently microsolvated intermediates and transition states, there are only small differences between generalized gradient approximation (GGA) density functional theory, hybrid density functional theory, and high level wavefunction theory calculations. This suggests that even in solution phase reactions as complex as the MBH reaction mechanism, solvation energies are the most critical while electronic energy contributions from high level and computationally intensive methods may be less critical.
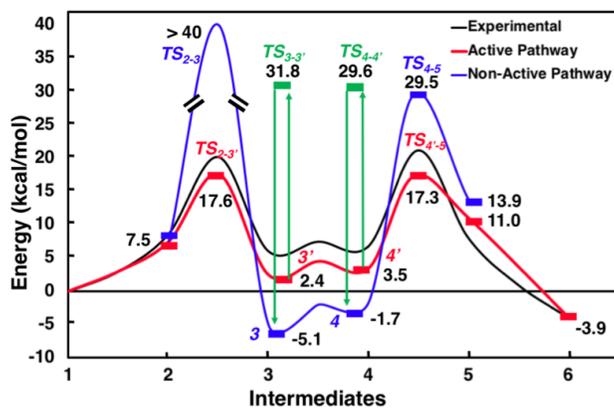
Figure 18: Reaction pathways relative to intermediate **1** with different intermediates and transition states obtained from GSM calculations compared to experimental data from Ref. [1]. The red line corresponds to the computationally predicted active pathway for the MBH reaction, and the blue line corresponds to a computationally predicted inactive pathway involving low energy (but kinetically inaccessible) intermediates.
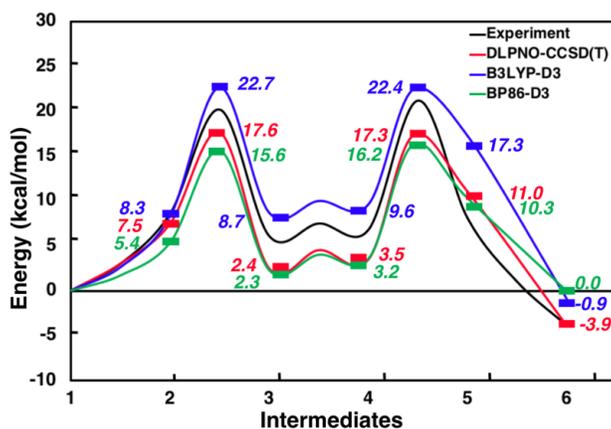
Figure 19: Red line is DLPNO-CCSD(T)/Def2-TZVP//BP86-D3/Def2-SVP, blue line is B3LYP-D3/Def2-TZVP//BP86-D3/Def2-SVP, the green line is BP86/Def2-TZVP//BP86-D3/Def2-SVP model chemistries with cluster modeling using five methanol molecules.

2.4 Conclusions

We have demonstrated a new, automatable, and paramedic, modeling scheme that reasonably models the MBH reaction mechanism and should be applicable for studying other challenging reaction mechanisms where CSM models can fail. No dynamics simulations are needed in this model, and transition states are automatically and efficiently found using GSM methods. Four points warrant consideration:

1. As has been stated before by others, CSMs can inadequately model significant local solvation effects in reaction mechanisms, and this affects not only proton shuttling mechanisms but also intramolecular charge transfers or tautomerizations. Remedying this requires some degree of explicit solvation.

2. The degree of explicit solvation required can be probed using this paramedic method that takes advantage of a globally optimized reactant state and error cancellation when modeling reaction pathways as a chronological sequence of GSM pathways.

3. The local solvent environment around a solute plays a critical role in stabilizing reaction intermediates, but any particular solvent environment should neither be assumed to be the same for all intermediates in a reaction mechanism nor easily transferable to different intermediates. We therefore encourage future efforts to report pathways that involve a globally optimized intermediate state followed by a sequence of reaction paths calculated using the growing string method.

4. Once a complete reaction pathway is found using the paramedic method, there appears to be only a marginal gain in accuracy when using high level methods, so a relatively efficient approach such as BP86-D3/Def2-TZVP//BP86-D3/Def2-SVP is likely adequate for qualitatively accurate mechanism predictions and comparison to experiment.

We now summarize steps taken for the paramedic method and recommend that others consider using it to model other reaction mechanism in solvents where CSM models appear to fail. Note that we do not use CSMs in any step of the procedure used here, but in other situations using a CSM might help.

- Step 1: identify globally optimized clustered states for hypothetical reactant states with different numbers of solvent molecules. Our umbrella sampling simulations using explicit solvent models suggest these are adequate representations of reactant states.

- Step 2: systematically explore reaction pathways using single-ended GSM calculations and eliminate models that give unrealistic barriers for processes known to occur and identify a microsolvation model that yields reasonable reaction profiles using energies from a trusted level of computational theory, e.g. BP86-D3/Def2-TZVP//BP86-D3/Def2-SVP or DLPNO-CCSD(T)/Def2-TZVP // BP86-D3/Def2-SVP when higher accuracy is needed.

- Step 3 (optional): use double-ended GSM calculations to identify any barrier heights between metastable intermediate states of interest for a complete mechanistic picture.

When successful, this paramedic treatment should be a robust and automatable way to model other challenging reaction mechanisms that involve explicit solvent molecules. Though the paramedic treatment is a multistep process that involves testing variable numbers of explicit solvent molecules, the static quantum chemistry calculations used here are significantly fewer and less computationally demanding than reaction dynamics simulation methods using quantum chemistry. In fact, the slowest step on our study were the single point DLPNO-CCSD(T) calculations. Furthermore, by clustering all atoms into a single microsolvated state, relative free energies (based on IGRRHO approximations) can usually be assumed to closely parallel electronic energies. Thus, hessian calculations might be considered unnecessary as long as GSM calculations (which do not involve hessian calculations) correctly identify stationary points. This paramedic approach appears to capture essential physical chemistry of chemical reactions involving solvent molecules, it appears relatively insensitive to levels of theory used, and it should be considered as a practical alternative to dynamics based computational studies in future studies.

3.0 Understanding Solvation Effects on Hydrogenation Barriers for CO_2 Reduction on Carbon Based Materials

3.1 Introduction

Society's continued consumption of fossil fuels results in increasing levels of CO_2 in the atmosphere, and the concentration of CO_2 in Antarctica recently passed 400 ppm, a level that has not been reached for four million years. Since CO_2 is correlated with extreme weather and global climate change, there have been efforts toward developing sources for renewable and sustainable energy that would supplant fossil fuels. In particular, there is great interest in converting CO_2 into value-added chemicals and fuels such as formic acid, CO, or methanol.[200, 201] Progress has been made in CO_2 reductions to useful products based on chemical,[202] thermal,[203] electrochemical,[204, 205, 206, 207], or photochemical means,[208, 209] or via combinations of different approaches.[210, 211, 212] Electrochemical and photochemical processes operating at room temperature show promise for scalability and favorable energetic efficiency, but it remains challenging to design electro- and photocatalysts with low overpotentials and high faradaic efficiencies for proton and electron transfers.[213, 214] Improved guidelines would be helpful for understanding how to effectively and selectively control proton and electron transfers within generalized proton coupled electron transfer (PCET) reactions.[215, 216, 217]

Computational quantum chemistry modeling can help interpret and guide experimental work in this area by providing insights into chemical reaction mechanisms. Advances in algorithms and hardware make it easier to computationally model larger scale systems with higher accuracy, but the central challenges of understanding *what* processes to model and *how* to physically model them in a reliable way still remain. Indeed, many chemical reactions have intermediate states that are stabilized by different degrees of solvating environments, and neglecting or incorrectly modeling these environments can significantly impact the quality of predictions from computational modeling. We have previously used thermodynamic Pourbaix diagram analyses to study a variety of homogeneous and heterogeneous catalysis

systems for CO_2 reduction. Interestingly, we have predicted that reaction conditions for several CO_2 electroreduction processes ranging from homogeneous pyridinium[39, 40] and homogeneous ruthenium[41]-complexes as well as heterogeneous N-doped nanocarbons[42] and partially reduced SnO_2 oxides[43]. Pourbaix diagram analyses give us ideas about the thermodynamics but they do not offer any information about the kinetic barriers of a given reaction. As illustrated by many different research groups, modeling reaction mechanisms without accounting for barriers provides an incomplete picture and can result in qualitatively different outcomes that might be wrong and/or misguide future research efforts.

In this project we model electrochemical hydrogenation barriers for CO_2 with various pyridine-derived catalysts. We model covalent hydrogen atom (H·) transfers, stepwise or coupled proton and electron transfers, and hydride (H$^-$) transfers that is coupled with a proton transfer while accounting for solvation and salt ions effects in our systems. We use both QM/MM simulations and potential of mean force calculations (PMF) as well as a calculation scheme that has previously developed in our group called "paramedic treatment"[218] which relies on mixed implicit/explicit modeling. We report barrier heights for different hydrogenation pathways and discuss the role of solvent molecules and salt ions.

3.2   Computational Methodology

All cluster calculations are performed with ORCA[188] simulation package. Cluster structures are optimized using Kohn-Sham density functional theory (DFT) at the BP86-D3BJ[185, 186]/Def2-SVP[187] level of theory. Starting from a fully optimized structure, we performed single-ended Growing String Method (GSM)[195, 196, 197, 219] calculations to model different reaction pathways. We then computed single point electronic energies using hybrid functionals (B3LYP[189]-D3[190], and ωB97X-D3[220]) and a high level ab initio method DLPNO-CCSD(T)[221, 222, 223, 224] each using the relatively large (Def2-TZVP[187]) basis set. Calculations made use of the RI[191] and RIJCOSX[191] approximations when appropriate. To treat outer-shell solvation effects we used CPCM implicit solvation with both geometry optimizations and single point energy calculations. Since each

clustered intermediate had the same number and type of atoms, the vibrational, thermal, and entropic energy contributions from standard ideal gas, rigid rotor, and harmonic oscillator approximations can be expected to largely cancel out.

The quantum mechanical/molecular mechanics–molecular dynamics (QM/MM-MD) simulations have been implemented to describe the free energy surfaces of the chemical reactions in solutions by applying the umbrella sampling technique. The simulations were performed on the spherical boundary condition surrounded by 285-290 water molecules to form a water sphere with 11 Å radius. To keep the volume of the sphere constant during simulations, we applied a harmonic restraint potential with a force constant of 2.0 kcal/mol/Å^2. The QM regions are described with ωB97X-D3[220]/def2-TZVP[187] level of theory and TIP5P water model has been applied to the MM regions. The NVT simulations were performed on each window at 300 K by the Nose-Hoover thermostat. Initially, 20 ps of NVT simulation is performed on the each window to equilibrate the systems with a time step of 1 fs. Next, 50 ps of NVT simulations are performed on the final structures from the equilibration step for the production runs. The PMF from the umbrella sampling simulations were obtained using the weighted histogram analysis method (WHAM).[193] GAMESS[225] simulation package is used to run the QM/MM-MD simulations.

3.3   Results and Discussion

First, we start by modelling single proton transfer from pyridinium cation to the CO_2 molecule. Reaction profile with GSM calculations and QM/MM calculations are shown in Figure 20. Different levels of theories showed similar energetics and the calculted reaction barrier with GSM calculations is ∼60 kcal/mol as shown in Figure 20a. Analogous PMF calculations based on QM/MM simulations are shown in Figure 20b. Calculated reaction barrier with this modeling scheme is ∼30 kcal/mol. Reaction energy calculated with both approaches agree well and is ∼ 20 kcal/mol. Even thought there is a significant difference with barrier heights, both calculation schemes suggest single proton transfer from pyridinium cation to CO_2 molecule is not favorable and not likely to happen under standard conditions.
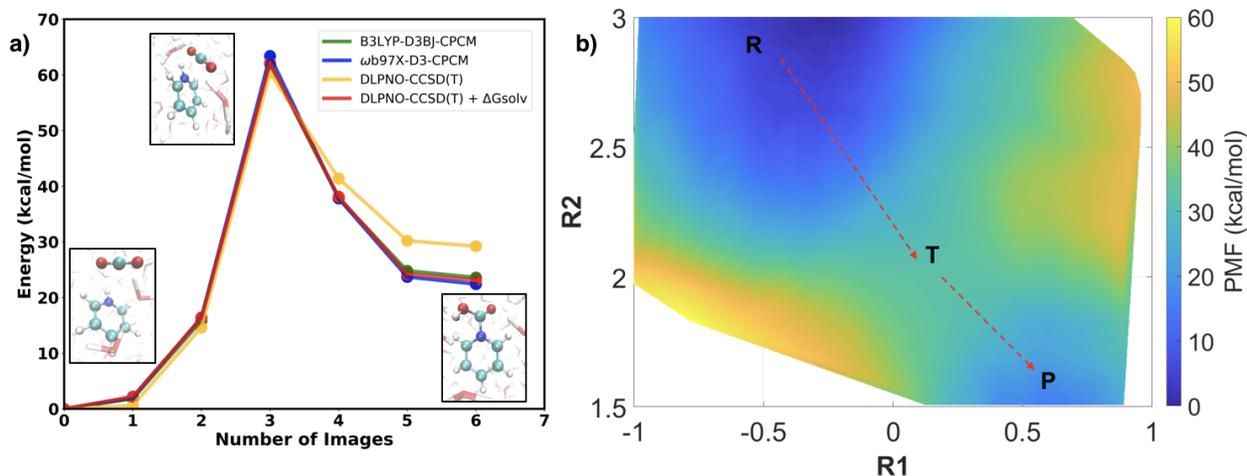
Figure 20: Single proton transfer from pyridinium cation to the CO_2. a) GSM calculations with zero explicit water molecule. Single point energy calculations on GSM geometries. To treat the outer-shell solvation effects CPCM implicit solvation is used with single point energy calculations. b) QM/MM calculations with zero explicit water molecule in the QM region.

Next, we have modeled single hydrogen atom transfer from pyridinyl radical to CO_2 molecule. Calculated reaction barrier for this mechanism is ∼30 kcal/mol with GSM calculations. Figure 21a shows the calculated reaction mechanism with different levels of theories on GSM pathway. B3LYP[189]-D3[190])/def2-TZVP[187] level of theory underestimates the reaction barrier compared to ωB97X-D3[220]/def2-TZVP[187] calculations. Calculated reaction barrier with ωB97X-D3[220]/def2-TZVP[187] agrees very well with the high level DLPNO-CCSD(T),[221, 222, 223, 224]/def2-TZVP[187] calculations. When we calculate the same reaction pathway with QM/MM simulations, we get a barrier height of ∼25 kcal/mol. GSM and PMF calculations show reasonable agreement with for both the barrier height and the reaction energy. Overall, both approaches show hydrogen atom transfer from pyridinyl radical to CO_2 molecule is again not likely to happen under standard conditions.
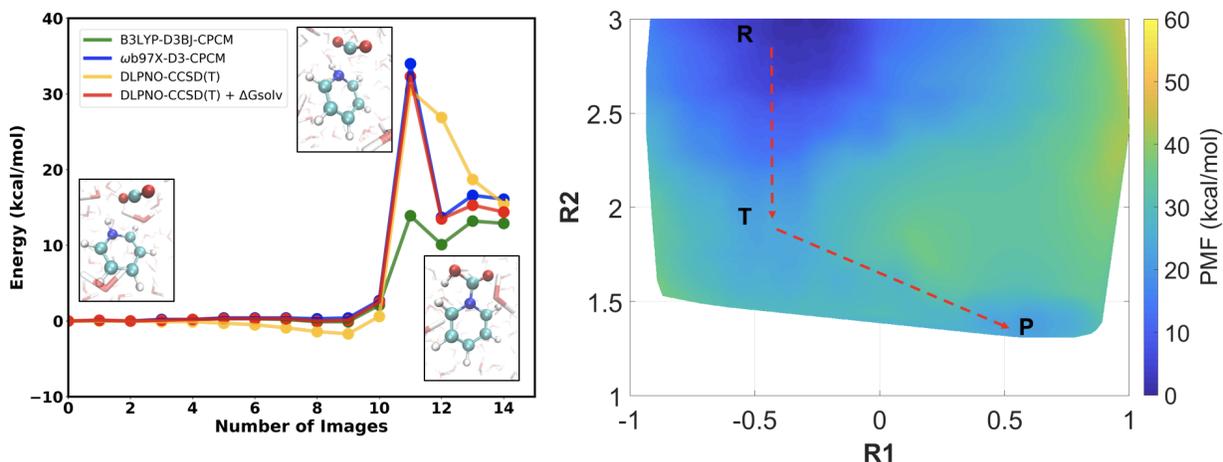
Figure 21: Hydrogen atom transfer from pyridinyl radical to the CO_2. a) GSM calculations with zero explicit water molecule. Single point energy calculations on GSM geometries. To treat the outer-shell solvation effects CPCM implicit solvation is used with single point energy calculations. b) QM/MM calculations with zero explicit water molecule in the QM region.

Then, we have considered hydride (H^-) transfer coupled with a proton transfer from dihydropyridine to CO_2 molecule. We have modeled hydride (H^-) coupled proton transfer using zero, one, and two explicit water molecules with our GSM calculations. Figure 22 shows different degrees of solvation treatments. When we use zero explicit water molecules in the GSM calculations, calculated reaction barrier with DLPNO-CCSD(T) calculations resulted with 25.1 kcal/mol barrier height. Figure 22a shows using B3LYP[189]-D3[190])/def2-TZVP[187] level of theory underestimates the reaction barrier compared to ωB97X-D3[220]/def2-TZVP[187] calculations and high level DLPNO-CCSD(T)),[221, 222, 223, 224]/def2-TZVP[187] calculations. Yellow line shown in Figure 22a is the gas phase calculations with DLPNO-CCSD(T) level of theory and it shows significant differences with the remianing data. Adding implicit solvation in this caluclation scheme plays a significant role by stabilizing the intermediate structures. When we add one explicit solvent molecule in our calculation scheme, calculated barrier heights decreased for all level of theories (as shown in Figure 22b). For all different level of theories, transition state structure is stabilized
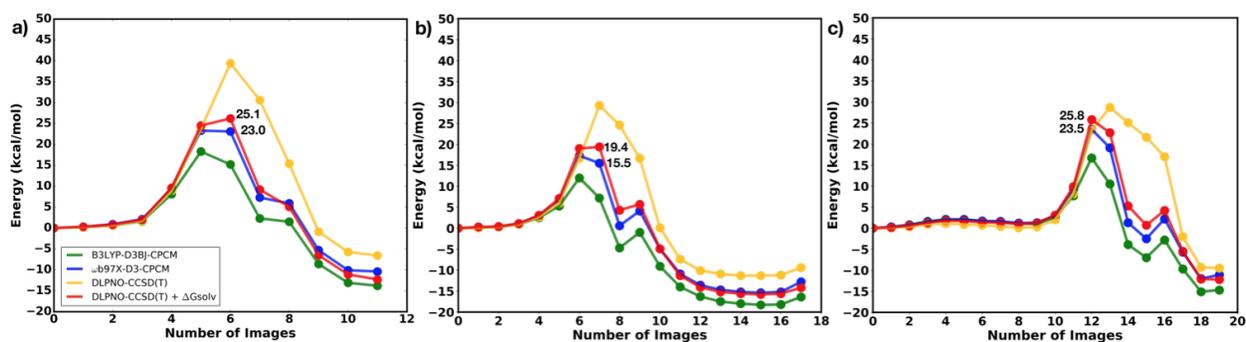
Figure 22: Hydride (H$^-$) transfer coupled with a proton transfer from dihydropyridine to the CO$_2$ molecule a) zero explicit water molecule, a) one explicit water molecule, c)two explicit water molecules in the cluster calculations.

by 5 kcal/mol resulting with 19.4 kcal/mol reaction barrier, as shown with the red data in Figure 22b. Finally when we increased the number of solvent molecules to two explicit water molecules, calculated barrier height increased to 25.kcal/mol. The second water molecule in the cluster calculation does not play an active role in the reaction but it forms hydrogen bonding network with first water molecule thus stabilizing the overall system. In all cases (Figure 22a, b, and c) B3LYP[189]-D3[190])/def2-TZVP[187] level of theory underestimates the reaction barriers proving higher level of theories are needed when modeling this reaction. Gas phase calculations demonstrate implicit solvation treatment is necessary when calculating hydrogenation barriers.

We have also calculated hydride (H$^-$) transfer coupled with a proton transfer with the PMF calculations and compared the results with the GSM calculations. To keep the calculations consistent we included one explicit water in the QM region and compared with the cluster calculations that included one explicit water molecule. Figure 23a shows the PMF pathway which resulted with ∼15 kcal/mol barrier height. When we use the same level of theory (ωB97X-D3[220]/def2-TZVP[187]) with our cluster approach, the calculated barrier height is 15.5 kcal/mol (Figure 23b). This shows we can calculate barrier heights as accurately as much more expensive PMF calculations using our clustering approach with mixed implicit/explicit modeling.
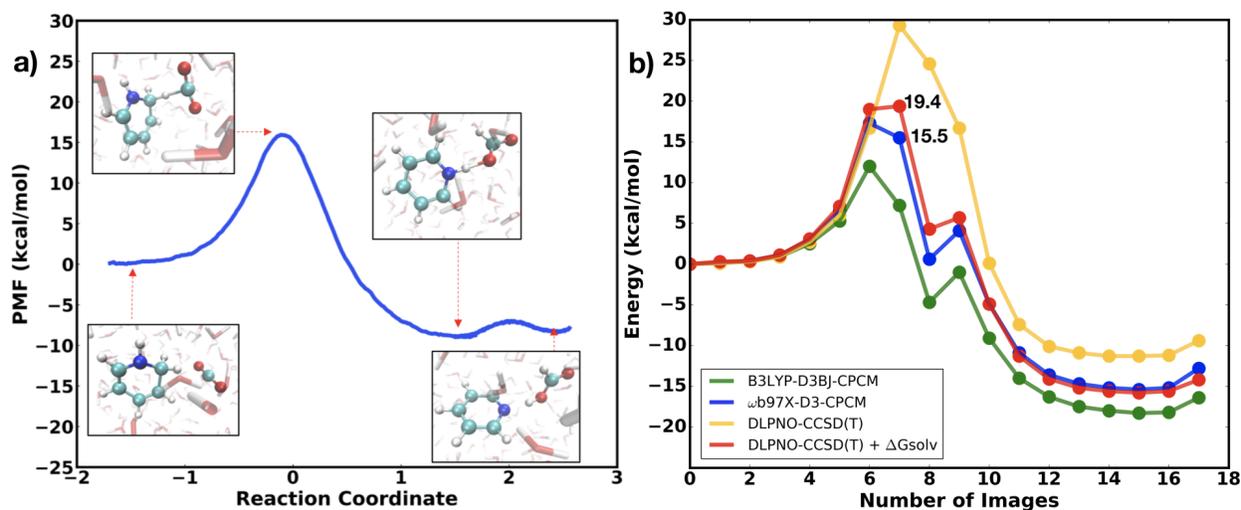
Figure 23: Hydride (H^-) coupled proton transfer from dihydropyridine to CO_2 molecule a) Reaction free energies with PMF calculations based on QM/MM simulations with one explicit water molecule in the QM region, b) GSM calculations with one explicit water molecule in the cluster. Single point energy calculations on GSM geometries. To treat the outer-shell solvation effects CPCM implicit solvation is used with single point energy calculations.

Finally, we have investigated how having a salt ion in the system effects the reaction energetics. We included a Na^+ ion in our cluster calculations with two explicit water molecules and performed another set of GSM calculations. Figure 24a shows the cluster calculations where we do not have the Na^+ ion in our system and Figure 24b shows the cluster calculations when we included the Na^+ ion near the dihydropyridine catalyst. Our original hypothesis of having the Na^+ ion in the cluster will stabilize the hydride transfer process with electrostatic interactions turned out to be false for this reaction mechanism. The calculated reaction barriers for these two cases are very similar (23.5 kcal/mol without Na^+ and 22.4 kcal/mol with Na^+ ion in the cluster using ωB97X-D3[220]/def2-TZVP[187] calculations.) When we compare the barriers heights for two cases using high level DLPNO-CCSD(T),[221, 222, 223, 224]/def2-TZVP[187] calculations, we see having the Na^+ ion the cluster actually results with a higher barrier.
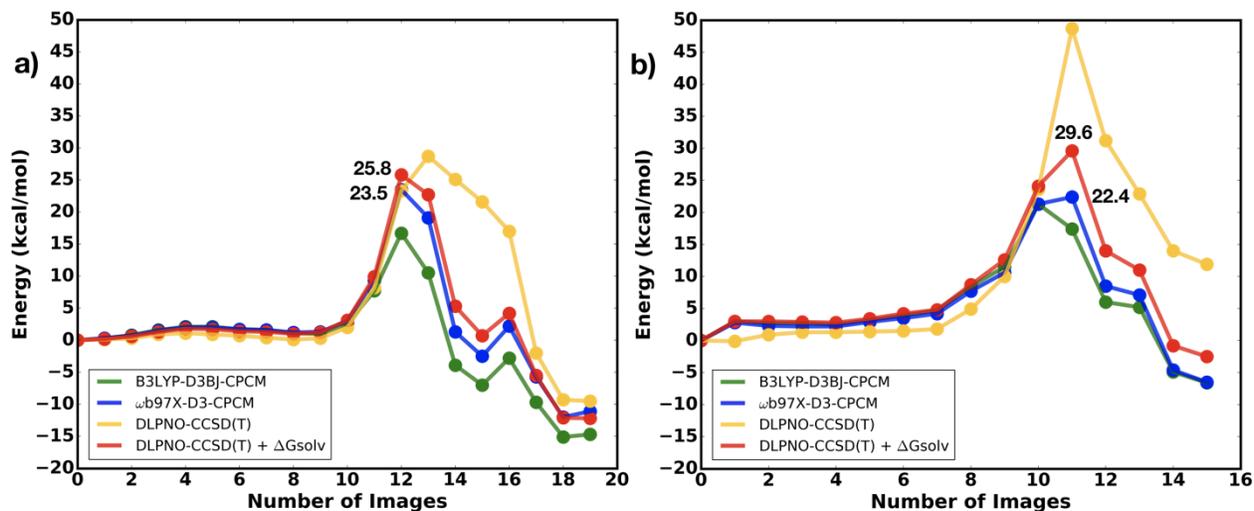
Figure 24: Hydride (H^-) coupled proton transfer from dihydropyridine to CO_2 molecule. GSM calculations with two explicit water molecules in the cluster. Single point energy calculations on GSM geometries. To treat the outer-shell solvation effects CPCM implicit solvation is used with single point energy calculations. a) Does not have a Na^+ ion in the cluster calculations whereas in b) the cluster has a Na^+ ion.

We have also performed PMF calculations to validate the effect of having an ion in the calculation scheme. Figure 25 shows the PMF results with and without the Na^+ ion. Calculated free energy barriers are almost exactly the same for two different cases. In the QM/MM simulations the Na^+ ion is completely solvated with MM waters and does not really interacts with the reactants. This shows having a salt ion in the system does not play an important role for this reaction pathway.
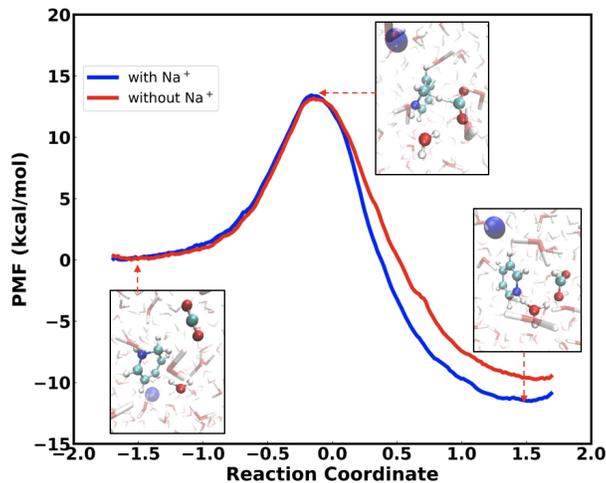
Figure 25: Hydride (H^-) coupled proton transfer from dihydropyridine to CO_2 molecule. Reaction free energies with PMF calculations based on QM/MM simulations with one explicit water molecule in the QM region. The red data does not have the Na^+ ion in the solvent box whereas the blue data has a Na^+ ion in the solvent box.

These analyses show CO_2 hydrogenation is most likely to occur as hydride (H^-) coupled proton transfer from the dihydropyridine catalyst. All other calculated barriers (proton transfer or hydrogen atom transfer) resulted with very high energies meaning they are not likely to occur under standard conditions. We also show having one explicit water molecule in the calculations play an important role in the reaction by stabilizing the intermediate structures. We point out using implicit solvation to treat the outer-shell solvation effects is also crucial for this reaction pathway. Finally, we look at the salt ion effect on the reaction mechanism and conclude having an ion in the system does not make a significant difference in the reaction energetics.

3.4   Conclusion

We have calculated barrier heights for different hydrogenation pathways for electrochemical CO_2 reduction. He have considered three main reaction mechanisms; proton transfer, hydrogen atom (H·) transfer and hydride (H^-) coupled proton transfer from pyridine-derived intermediates to the CO_2 molecule. We use two different calculation schemes; one based on mixed implicit/explicit modeling with different numbers of explicit solvent molecules other is more expensive PMF calculations based on QM/MM simulations. We show in most cases two different calculation schemes result with similar reaction pathways. In cases where the reaction is physically irrelevant two calculation schemes resulted with different barrier heights however both of them were able to identify that reaction will not likely to take place. We have considered number of different solvation treatments together with the effect of a salt ion in the system. We show one should be cautious when considering solvation effects and include explicit water molecules in the calculation scheme to model physically relevant states. Finally, we have investigated the salt ion effect and showed it does not play a significant role in the electrochemical CO_2 reduction with pyridine-derived catalyst materials. This approach is automatable and could be applied to any reaction mechanism where the solvent molecules affects the quantum level nature of reaction mechanisms.

4.0 Machine Learning Guided Approach for Studying Solvation Environments

This work has been published as Basdogan, Y., Groenenboom, M. C., Henderson, E., De, S., Rempe, S. B., & Keith, J. A. (2019). Machine Learning Guided Approach for Studying Solvation Environments. *Journal of Chemical Theory and Computation.*

4.1 Introduction

Solvation plays an essential role in chemical and biological processes ranging from homogeneous catalysis to ion channel transport to energy storage. In many cases, the explicit interactions between small ions with nearby solvent molecules are crucial for molecular-scale understanding. In such cases, single-ion solvation free energies can be several hundreds of kcal/mol (or greater than 10 eV), which can make accurate predictions quite challenging. Molecular dynamics (MD) or Monte Carlo (MC) simulations have been used, notably for systems that have anions and complex small molecules,[226, 227, 228] but the accuracy of these simulations depends on the availability of high-quality force field parameters. In the absence of reliable parameters, MD simulations involving quantum mechanics (QM) calculations can be accurate, but they are far more computationally laborious. Semi-empirical continuum solvation models (CSMs)[70, 78, 229, 85, 230, 84] have been developed to efficiently determine solvation free energies, but CSMs can sometimes result in large errors, especially with systems that have non-uniform charge distributions. Such errors can significantly impede predictions of thermodynamic properties and severely bias mechanistic predictions[1].

A standard approach to address these problems has been to include explicit solvent molecules with the QM calculation of the solute, using cluster-continuum or mixed implicit/explicit modeling, since this often provides better solvation free energies from thermodynamic cycles. Of these methods, the cluster formulation of quasi-chemical theory (QCT), developed by Pratt, Rempe, and colleagues, is a rigorous treatment that uses an electronic structure calculation on the ion with one or more solvation shells.[231, 232, 233]

This approach has produced accurate predictions of solvation free energies for hydration of alkali metal ions (Li^+, Na^+, K^+, Rb^+),[234, 235, 236, 237, 238, 239] alkaline earth metals (Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+}),[240, 241, 242, 243] transition metals,[244, 245] halide ions (F^-, Cl^-),[246, 247, 248] small molecules (Kr, H_2, CO_2),[249, 250, 251, 252] ion solvation from non-aqueous solvents,[253, 242] and binding sites of proteins and other macromolecules,[254, 255, 256, 257, 258, 259, 260] generally to within 5% error.[243] However, the correct use of QCT requires determining an appropriate solvation shell for the system, and this can be non-trivial.[261, 262]

Adding to the complexity of single-ion free-energy solvation predictions is that there are two different free energy scales that are frequently misunderstood or not acknowledged in the literature. One is often called the 'absolute' scale, while the other is called the 'real' scale. The real scale includes the phase potential[263] or surface potential,[264, 265, 266] which is the total reversible work to move an ion across the vacuum-liquid interface, whereas the absolute scale does not. The absolute scale is associated with data from Marcus, who studied and reported experimental solvation free energies for a large number of ions.[267] Those data rely on the 'classical' extra-thermodynamic assumption, referred to as the tetraphenyl-arsonium/tetraphenyl-borate (TATB) hypothesis, that two specific ions of opposite charges have similar absolute free energies. That hypothesis assumes the system is independent of any interfacial potential that arises from the anisotropic distribution of the solvent molecules near the interface.[268] In a real physical system, a solvation free energy will also include a phase potential contribution that depends on the interfacial potential at the air-water interface. The real scale can be associated with data from Tissandier et al., who have extrapolated conventional free energy measurements on small ionic hydrates to obtain real solvation free energies of ions in bulk phase.[269] This idea is often referred as the cluster pair-based (CPB) approximation.

The absolute solvation free energy scale can be converted into the real solvation free energy scale by incorporating the phase potential using the following equation:

$$\Delta G_{\text{solv}}^{\text{real}} = \Delta G_{\text{solv}}^{\text{abs}} + zF\phi \tag{4.1}$$

where F is the Faraday constant, z is the atomic charge and ϕ is the interfacial potential. Table 3 compares Marcus's data with data from Tissandier et al., and it highlights the phase potential contribution in real solvation free energy calculations, which is about -10 kcal/mol for alkali metals and about $+12$ kcal/mol for halides. With two sets of experimental data to compare to, there has often been a lack of consensus on which calculation schemes result in which solvation free energy scale and why. It is generally understood that free energy calculations using periodic boundary conditions (such as MD and MC simulations) do not include the phase potential contribution, and thus represent absolute solvation data because there is no physical vacuum-liquid interface.[270]

Table 3: Comparison of different solvation scales in kcal/mol.

Ion	TATB [a]	CPB [b]	Difference
Li^+	-117.3	-126.5	-9.2
Na^+	-91.0	-101.3	-10.3
K^+	-74.3	-84.1	-9.8
F^-	-114.9	-102.5	12.4
Cl^-	-85.0	-72.7	12.3
Br^-	-79.1	-66.3	12.8

[a] Data taken from Ref. [267]

[b] Data taken from Ref. [269]

For cluster-based calculations this is murkier. Specifically, QCT literature cites the absence of phase potentials in theoretical predictions and reports data in closest agreement with the absolute solvation data of Marcus,[243] while other computational studies using a similar thermodynamic cycle and cluster-continuum approach have reported closer agreement with the real solvation scale.[271, 272] Of course, solvation energies will depend on how many solvent molecules are used and where they are placed. Kemp and Gordon demonstrated the effective fragment potential (EFP) method coupled with Monte Carlo simulations can be used to study the solvation of F^- and Cl^- anions.[273] Their approach showed that 15 water molecules in this model were required to fully solvate a single F^- anion while 18

water molecules were required to completely solvate a Cl^- anion. Merz and coworkers also used molecular dynamics simulations to identify the first solvent shell that can be used in calculations using continuum solvation methods.[129] It would be beneficial to have a general and automatable way to model local solvation environments (of any solute in any general solvent) without using molecular dynamics simulations that can be computationally prohibitive.

This work elucidates the theory between two different thermodynamic cycles (Schemes 1 and 2) and how they result in two different solvation free energy scales. To automatically generate microsolvated clusters, we used a global optimization code called ABCluster [181, 274]. We then calculated the real solvation free energies with cluster-continuum modeling using the thermodynamic cycle outlined in Scheme 2. We initially hypothesized that solvation free energies will improve if we systematically add explicit water molecules around each ion while ensuring that each microsolvated state is a reasonable approximation of a thermodynamically low energy structure. A similar idea was previously studied by Bryantsev and co-workers by increasing water cluster sizes to 18 explicit solvent molecules around the Cu^{2+} ion, which significantly decreased the error compared to the CSM-computed solvation free energy.[128] Here, we introduce a way to leverage machine learning (ML) to study local solvation structures. Unlike other studies that use supervised ML to predict solvation energies with algorithms like random forest,[275] decision tree,[276] or artificial neural networks,[277] we are using unsupervised ML to study how similarities between microsolvated structures coincide with solvation energies so that one can learn the inherent arrangement of our data without using explicitly provided labels. We first assemble our data-set of microsolvated structures and then use dimensionality reduction algorithms to study microsolvation structures. One of the main challenges in applying ML techniques to chemistry problems is to find the best representation of the system so that it is complete and concise. In this study we use Smooth Overlap of Atomic Positions (SOAP) descriptors[278] to represent our microsolvated structures. Next we use sketch-maps to reduce the dimensions of our feature vectors and to study similar patterns.[194] By using the combination of SOAP and sketch-map ML algorithms we demonstrate that low energy molecular clusters produced by our procedure have structurally similar local solvation environments.

56

4.2 Theory

Cluster-continuum modeling has been used in different formulations to calculate solvation free energies of small ions.[127, 279, 180, 280, 129] These methods involve different approximations, ranging from including a single solvent molecule to using MD simulations to obtain physical solvent structures at room temperatures. This hybrid approach has received further developments in the theory (e.g. by using cluster expansion treatments).[281, 282, 283] QCT is the most robust approach of these because it is based on statistical mechanics,[284] and it has been proven to be reliable in different applications.[244, 254, 233, 253, 247, 248, 243, 234, 235, 236, 240]

The starting point for QCT is to partition the region around the solute into inner- and outer-shell solvent domains. Akin to cluster-continuum modeling schemes, the inner shell is typically treated quantum mechanically, while the outer shell is treated with a dielectric continuum model. When applied to the hydration of an ion X having a charge $m\pm$, the inner-shell reaction is given as a cluster association equilibrium:

$$\mathrm{X}^{m\pm} + n\mathrm{H_2O} \rightleftharpoons \mathrm{X(H_2O)}_n^{m\pm} \tag{4.2}$$

A clustering algorithm is applied to identify the populations of the clusters on the right side of Eq. 4.2, namely the $\mathrm{X(H_2O)}_n^{m\pm}$ species. One normally identifies inner-shell configurations for an ion by defining waters within a distance λ from the ion as an inner-shell partner. With n water ligands in the cluster, the excess chemical potential, or hydration free energy, consists of several terms,

$$\mu_{\mathrm{X}}^{\mathrm{(ex)}} = -kT \ln \left[K_n^{(0)} \rho_{\mathrm{H_2O}}{}^n \right] + kT \ln \left[p_{\mathrm{X}}(n) \right] + \left(\mu_{\mathrm{X(H_2O)}_n}^{\mathrm{(ex)}} - n\mu_{\mathrm{H_2O}}^{\mathrm{(ex)}} \right) \tag{4.3}$$

The terms in Eq. 4.3 describe contributions to the total ion hydration free energy from the inner and outer-shell solvation environments. The first term gives ion association reactions with water molecules in the inner shell taking place in an ideal gas phase. The association reactions are scaled by the water density, $\rho_{\mathrm{H_2O}}$, to account for the availability of water ligands to occupy the inner shell. The second term accounts for the thermal probability that a specific ion has n inner-shell partners in solution. The last terms describe the solvation

of the $X(H_2O)_n^{m\pm}$ cluster and the de-solvation of n individual water molecules from aqueous solution in the outer-shell environment.

A judicious selection of λ and n in Eq. 4.3 can simplify the free-energy analysis. Even though Eq. 4.3 does not depend explicitly on λ, by considering a specific λ, we can identify the the most probable n, denoted as $\bar{n}$. When $\bar{n}$ is used in Eq. 4.3, the $kT\ln[p_X(n)]$ term can be dropped to result in Eq. 4.4

$$\mu_X^{(ex)} \approx -kT\ln\left[K_{\bar{n}}^{(0)}\rho_{H_2O}{}^{\bar{n}}\right] + \left(\mu_{X(H_2O)_{\bar{n}}}^{(ex)} - \bar{n}\mu_{H_2O}^{(ex)}\right) \tag{4.4}$$

Alternatively, the magnitude of the contribution, $kT\ln[p_X(n)]$ from Eq. 4.3 can be estimated from molecular simulation results for any n. Second, CSMs can be used to determine outer-shell contributions. With most CSMs, the external boundary of the model cavity is defined by spheres centered on each of the atoms. Typically, CSM results are sensitive to the radii of the spheres that define the solute cavity, but when the ion is surrounded by a full shell of solvating ligands, the sensitivity is lessened (when the radii for the ligands are adequate), and this results in a fortuitous error cancellation in the last terms of Eq. 4.3 and 4.4. Third, selecting clusters with small n generally results in stronger solute-solvent interactions, which helps ensure that vibrational motions are characterized by small displacements from equilibrium, which is required when assuming a harmonic potential energy surface for the analysis of a free energy. Prior work suggests that anharmonic vibrational motions become prominent with clusters as small as $n=5$ for Na^+ and K^+ ions in clusters with water molecules,[257] and they can be even smaller sizes for anion-water clusters.[248]

As an aside, the solvation energy represented in Eq. 4.4 can also be equivalently represented using the thermodynamic cycle shown in Scheme 1 (Figure 26) which is mathematically expressed (using different notation) with Eq 4.5. This alternative notation is based on the observable macroscopic quantities coming from thermodynamics and is often used in chemistry and engineering communities (e.g. see Refs. [180] and [128]). We note that when Eq. 4.5 and Eq. 4.6 are written in per mole basis they are equivalent to chemical potentials.

$$\Delta G_{solv}^*(X^{m\pm}) = \Delta G_{g,bind}^\circ - \bar{n}\Delta G^{\circ\rightarrow*} + \Delta G_{solv}^*(X(H_2O)_{\bar{n}}^{m\pm}) - \bar{n}\Delta G_{solv}^*(H_2O) - \bar{n}RT\ln[H_2O] \tag{4.5}$$

58
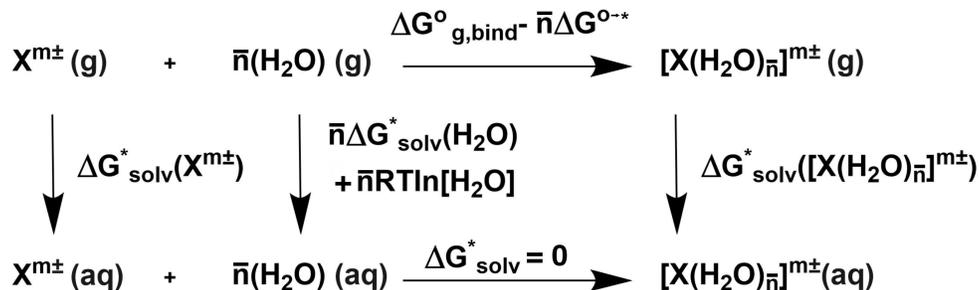
Figure 26: Monomer cycle (Scheme 1) for calculating an absolute solvation free energy.

Scheme 1 is sometimes called the 'monomer cycle' since it involves individual water monomers rather than a water cluster. The free energy of binding a gas-phase cluster is expressed as $\Delta G_{\text{g,bind}}^{\circ}$, where the circle denotes the free energy difference at a gas-phase standard state of 1 bar. The solvation free energies ($\Delta G_{\text{solv}}^{*}$) are labeled with asterisks to denote energies conventionally expressed at an aqueous standard state of 1 M, and they are calculated here using SMD implicit solvation. Additional corrections ($\Delta G^{\circ \to *}$, each having a magnitude of 1.9 kcal/mol or 0.08 eV) are needed to account for the change from a gas-phase standard state to an aqueous-phase standard state. Here, we use Scheme 1 to map the QCT theory on macroscopic variables. Solvation energies calculated using Scheme 1 are comparable to the Marcus scale and thus will be comparable to MD/free energy perturbation or QCT studies that do not have the phase potential contribution.[285]

Of the different cluster-continuum procedures that do not require dynamics, the procedure by Bryantsev et al. is promising since it appears to yield solvation free energies that agree well with the experimental data for both the proton and Cu^{2+}, and with results that appear to match the real solvation scale. Their cycle, outlined in Scheme 2, is similar to the monomer cycle in Scheme 1, but it involves pre-formed water clusters containing n interacting water molecules that have been optimized at 0 K and free energy contributions are obtained using standard ideal gas, rigid rotor, and harmonic oscillator approximations. The single-ion solvation free energy from the Scheme 2 cluster cycle is calculated with Eq. 4.6:

$$\Delta G_{\text{solv}}^{*}(X^{m\pm}) = \Delta G_{\text{g,bind}}^{\circ} - \Delta G^{\circ \to *} + \Delta G_{\text{solv}}^{*}(X(H_2O)_n^{m\pm}) - \Delta G_{\text{solv}}^{*}(H_2O)_n - \text{RTln}([H_2O]/n)$$

(4.6)

59

$$X^{m\pm} \text{(g)} + \text{(H}_2\text{O)}_n \text{(g)} \xrightarrow{\Delta G^o_{g,bind} - \Delta G^{o-*}} [X(H_2O)_n]^{m\pm} \text{(g)}$$

$$\Delta G^*_{solv}(X^{m\pm}) \qquad \Delta G^*_{solv}(H_2O)_n + RTln[[H_2O]/n] \qquad \Delta G^*_{solv}([X(H_2O)_n]^{m\pm})$$

$$X^{m\pm} \text{(aq)} + \text{(H}_2\text{O)}_n \text{(aq)} \xrightarrow{\Delta G^*_{solv} = 0} [X(H_2O)_n]^{m\pm}\text{(aq)}$$
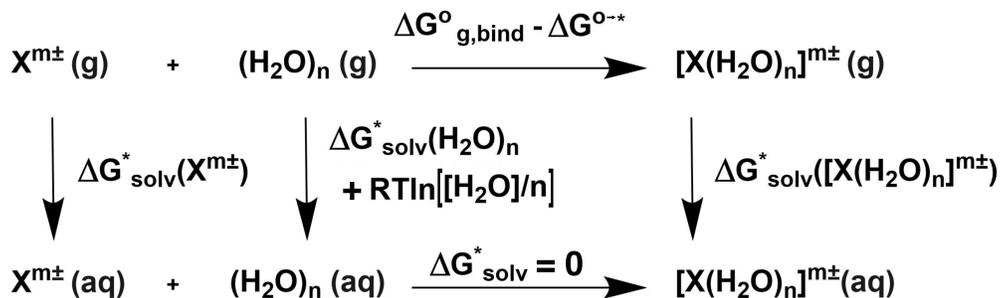
Figure 27: Cluster cycle (Scheme 2) for the calculation of real solvation free energy.

Scheme 2 also evaluates the same QCT theory shown in Eq. 4.3, but by applying QCT to both the water dehydration problem ($\mu^{(ex)}_{H_2O}$) as well as the ion hydration problem ($\mu^{(ex)}_X$). This dual QCT approach has advantages due to anticipated error cancellations.[252] Successful use of CSM is known to require properly chosen CSM parameters since results can vary greatly with surface type, cavity size, and continuum model used.[286] However, by using similar sizes of clusters for ion hydration and water dehydration, the boundary λ between inner and outer-shells is approximately balanced on both sides of the equation, leading to a cancellation of errors to the outer-shell solvation contribution from a CSM model.

The same balance in cluster sizes may also lead to an approximate cancellation of anharmonic contributions in the inner-shell contributions to the solvation free energy. Eq. 4.3 depends on using the most probable n to eliminate the $kT\ln[p_X(n)]$ term (as done in Eq. 4.4), or requires molecular simulations to explicitly evaluate that term. It also needs a filled inner-shell occupancy so that the CSM model is minimally dependent on specific radii used to compute the outer-shell contribution to hydration free energy. Scheme 2 approximately eliminates these constraints through error cancellations. With Scheme 2, large n values can be used; however, care must be applied when the cavity radius is around 6 Å. In such length scales and above, the surface or phase potential contributions to the solvation free energy, ϕ, should be included in the calculation.[287, 288, 289, 290, 291] In the analysis here, outer-shell contributions to the solvation free energy go to zero as cluster size increases,[240, 243] and then the phase potential enters into the calculation and then is accounted for naturally. This explains why results from Scheme 2 agree better with the real solvation scale.

To identify the degree of solvation necessary we employ unsupervised ML algorithms to assemble and then study the similarities between the microsolvated structures. The main challenge in using ML based algorithms, particularly with chemistry applications, is to come up with an appropriate representation that gives a complete description of the system. The SOAP kernel expresses the atomic neighborhood of a specific atom using a local expansion of a gaussian-smeared atomic densities. Previously, it has been used to study different geometries of fullerene, amorphous silicon, pentacene, and ice structures.[292, 194, 293, 294, 295] SOAP is also a good representation for microsolvated systems because it is invariant to rotations, translations, and permutations. Assembling techniques like SOAP identify an agglomeration of data in chemical space, but they do not offer a Euclidian relationship between different structures. Such relations can be determined and represented using maps that represent geometrically similar structures as data points that are adjacent to each other. Here, we use sketch-map non-linear dimensionality reduction technique to construct a two-dimensional representation of the free energy surface. Given a certain cutoff radius, this algorithm identifies structures that show similar local solvation motifs, e.g. if two points are close to each other on the sketch-map then the local solvation environments from the two data points are relatively similar.

In this study, we will apply Scheme 2 to systematically model microsolvated ions and water clusters using $n = 1 - 20$ water molecules. Below, we will show a modeling scheme that involves modern computational tools such as ABCluster, dispersion-corrected Kohn-Sham density functional theory, and the SOAP algorithm to analyze this thermodynamic cycle to quantify energy contributions, assess likely causes for errors, and understand the local structures of water molecules in these solvation environments. While more calculations are required for Scheme 2 than would be needed for Scheme 1, we find that the former scheme provides reasonably accurate single ion solvation free energies while also eschewing the need for *a priori* knowledge of the solvation environment. Thus, calculations from such cycles should be generalizable and easily automatable for any solute in any solvent environment.

4.3 Computational Methodology

We generated microsolvated structures using the rigid molecular optimizer module of the ABCluster program.[181, 274] We generated 1,000 low energy candidate structures using CHARMM force field parameters from MacKerell's CGenFF website together with TIP4P water parameters.[182] All force field parameters used in this study are reported in Table 4. The thermodynamic cycle reported in Scheme 2 also requires calculations on water clusters. To generate the water clusters, we followed the same procedure outlined above using TIP4P parameters for the water molecules.[296] We found that ABCluster can reliably identify Wales' low energy water clusters [297], but caution is still recommended in future work to ensure that meaningful microsolvated structures are obtained. Cluster geometries were then optimized at the same level of QM theory as the solute-solvent clusters, as discussed below.

The five lowest energy structures obtained from ABCluster then further optimized using computationally efficient BP86[185]-D3BJ[190]/def2-SVP[187] or B3LYP[189]-D3BJ[190]/def2-SVP[187] level of theory, as implemented in ORCA[188] using the RI-J or RIJCOSX approximations. To treat outer-shell solvation effects we used SMD implicit solvation with both geometry optimizations and single point energy calculations. Free energy contributions were calculated using the ideal gas, rigid rotor, and harmonic oscillator approximations at the same level of theory as the geometry optimizations. Finally, to assess the significance of higher levels of theory, we calculated single point energies on fully optimized geometries at the B3LYP[189]-D3BJ[190]/def2-TZVP[187] and ωB97X-D3[220]/def2-TZVP[187] levels of theory. Every energy reported in this manuscript is the Boltzmann-weighted average of the five lowest energy structures from our filtering procedure.

Finally, we compared our microsolvated structures with structures obtained from MD trajectories using the AMEOBA force field[298] as implemented in the TINKER[192] software package. First, we performed NPT simulations with a water box of 500 water molecules and equilibrated it for 200 picoseconds at 298.15 K and 1 atm. Next, we inserted an ion into the system (while removing one water molecule) and then performed NVT simulations for another 200 picoseconds. Finally, we extracted 100 structures from the NVT trajectory and compared them with the structures generated with our clustering approach.

Table 4: CGENFF parameters used with ABCluster [a]

Ions	q	epsilon (kJ/mol) [b]	sigma (Å)[c]
Li^+	+1.00	0.0097	2.31
Na^+	+1.00	0.1962	2.51
K^+	+1.00	0.3640	3.14
F^-	−1.00	0.3611	3.48
Cl^-	−1.00	0.6276	4.04
Br^-	−1.00	0.7606	4.76
Mg^{2+}	+2.00	0.0628	2.11
Ca^{2+}	+2.00	0.0502	2.44
Zn^{2+}	+2.00	0.0125	1.96
SO_4^{2-}	+2.40	1.0455	3.55
	−1.10	1.0455	3.15
TIP4P	+0.00	0.6485	3.15
	−1.04	0.0000	0.00
	+0.52	0.0000	0.00
	+0.52	0.0000	0.00
CO_3^{2-}	+0.161	0.0680	2.09
	−0.786	0.1921	1.76
	−0.786	0.1921	1.76
	−0.589	0.1650	1.69

[a] Parameters used in ABCluster algorithm. Following conversions needed to be performed to use CGENFF force field parameters in ABCluster:

[b] $\epsilon_{ABCluster} = \epsilon_{CHARMM}$ x (-4.184)

[c] $\sigma_{ABCluster} = \sigma_{CHARMM}$ x $(2^{5/6})$

4.4   Results and Discussion

We first benchmarked low energy water clusters generated from ABCluster to global minimum energy water clusters from the Cambridge Cluster Database that also used the TIP4P forcefield.[297] In all cases ($n = 1, 2, 4, 8, 12, 16, 20$), the energy differences between our structures and the reference structures from the database were all at most +1.2 kcal/mol (Table 5). This agreement demonstrates that ABCluster is an effective tool for identifying low energy structures and that our water clusters are comparable to well-established and globally optimized water cluster structures.

Table 5: Comparison of calculated electronic energies of water clusters in kcal/mol. [a]

n	Globally optimized clusters [b]	ABCluster clusters [c]	Difference
2	-95924.5	-95924.1	0.4
4	-191852.8	-191852.1	0.7
8	-383710.1	-383708.9	1.2
12	-575564.3	-575564.0	0.3
16	-767420.3	-767420.2	0.1
20	-959279.6	-959279.0	0.6

[a] Single point electronic energies calculated using ωB97X-D3/def2-TZVP.

[b] Geometries taken from Ref. [297] and then reoptimized using BP86-D3BJ/def2-SVP.

[c] Geometries obtained using the procedure stated in the main text.

Next, we calculated solvation free energies of $(H_2O)_n$ clusters using the thermodynamic cycle outlined in Figure 28 and compared it with solvation energy calculations using the SMD solvation model. Table 6 shows solvation free energies for water clusters derived using the thermodynamic cycle shown in Figure 28, solvation free energies calculated directly using a CSM, and the difference between the two calculation schemes. For cases where $n = 2, 4, 8$, the difference in the two sets of solvation free energies is less than 5 kcal/mol. However, for larger clusters ($n = 12, 16, 20$), the difference in free energies from these two calculation schemes significantly increases by as much as 30 kcal/mol. This trend shows that when

Table 6: Energy contributions (in kcal/mol) used in solvation free energy calculations of $(H_2O)_n$ clusters at $T = 298.15$ K in kcal/mol.

n	[a]	[b]	[c]	[d]	ΔG^*_{solv}(cycle) [e]	ΔG^*_{solv}(CSM) [f]	Difference
2	-12.6	8.5	3.4	3.8	-11.4	-14.7	-3.3
4	-25.2	17.0	5.6	3.4	-17.3	-16.7	0.6
8	-50.5	34.1	11.0	3.0	-30.5	-24.8	5.7
12	-75.7	51.1	19.6	2.8	-47.0	-29.1	17.9
16	-100.9	68.1	26.3	2.6	-61.8	-30.0	31.8
20	-126.2	85.2	27.9	2.5	-71.4	-38.9	32.2

[a] $n\Delta G^*_{\text{solv}}(H_2O)$ energies: $n\Delta G^*_{\text{solv}}(H_2O)$ was calculated by taking the experimentally derived $\Delta G^*_{\text{solv}}(H_2O)$ (see Ref. [299]) and then multiplying by n.

[b] Energy correction terms for the n water monomers: $n\Delta G^{\circ\rightarrow*} + n\text{RTln}([H_2O])$

[c] $G^\circ_{\text{g,bind}}$ energies calculated using DLPNO-CCSD(T)/def2-TZVP//B3LYP-D3BJ/def2-SVP.

[d] Energy correction terms for the water cluster: $\Delta G^{\circ\rightarrow*} + \text{RTln}([H_2O]/n)$

[e] ΔG^*_{solv}(cycle) values: Experimentally derived solvation energies using the thermodynamic cycle outlined in Figure 28. $\Delta G^*_{\text{solv}}(H_2O)_n = n\Delta G^*_{\text{solv}}(H_2O) - (\Delta G^\circ_{\text{g,bind}} - (n-1)\Delta G^{\circ\rightarrow*}) + \text{RTln}(n[H_2O]^{n-1})$.

[f] ΔG^*_{solv}(CSM) values: Solvation energies calculated using the SMD solvation model.

relatively large clusters of water are solvated with a CSM, the model appears to introduce significant errors that would then make them less reliable if used for calculations with Scheme 2. The observation also in part justifies the use of QCT methods that use Scheme 1 and relatively small cluster sizes. The lowest error arises with $n = 4$ because it is the most probable size for water clusters, making the ln $[p_X(n)]$ term in Eq. 4.3 approach to zero.

$n(H_2O)$ (g) $\xrightarrow{\Delta G^\circ_g(bind)}$ $(H_2O)_n$ (g)

$n\Delta G^*_{solv}(H_2O)$
+
$n\Delta G^{\circ \to *}$
+
$nRTln[H_2O]$

$\Delta G^*_{solv}(H_2O)_n$
+
$\Delta G^{\circ \to *}$
+
$RTln([H_2O]/n)$

$n(H_2O)$ (aq) $\xrightarrow{\Delta G^*_{solv} = 0}$ $(H_2O)_n$ (aq)
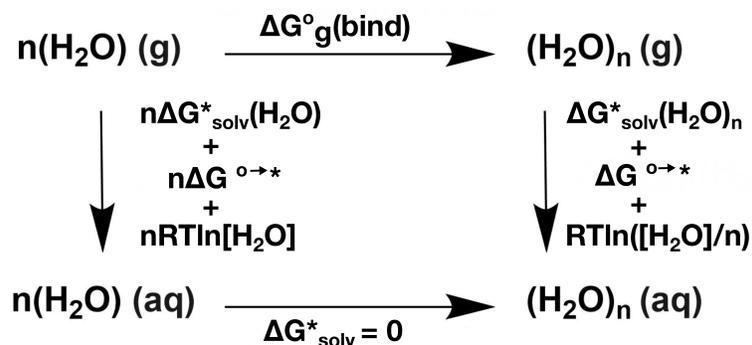
Figure 28: Thermodynamic cycle for the formation of water clusters.

We then benchmarked calculated gas-phase binding free energies of water molecules to different ions against experimental data.[300, 301] Table 7 shows gas-phase binding free energies for one water molecule and different ions. For all the ions (Li^+, Na^+, K^+, Cl^-, Br^-, F^-), calculations are in good agreement with the experimental data, and our errors are mostly under 5 kcal/mol. We further compared the Na^+ and K^+ binding free energies with one water molecule to other computational studies, and Table 7 shows that different calculations agree reasonably well with experimental data [239, 237]. We also calculated binding free energies involving four water molecules using Schemes 1 and 2, and compared them with both experimental and other computational studies (Table 8). The reference experimental data used for comparison add water molecules one by one to the system.[300, 301] Using Scheme 1, we get very good agreement with the experimental values and our errors are under 5 kcal/mol for all ions (Li^+, Na^+, K^+, F^-, Cl^-, Br^-). The data are also in relatively close agreement with gas-phase binding free energies calculated by Rempe and coworkers.[244] However, when we used Scheme 2 to calculate binding of ions to four water molecules, we obtained free energies that differ from the prior work by 10 kcal/mol. This difference is anticipated on the basis that Scheme 2 applies to solvation reactions, not gas-phase association reactions.

Table 7: Comparison of gas phase binding free energies for [X(H₂O)] in kcal/mol. [a]

Ions	Experiment [b]	Literature	$\Delta G^{\circ}_{\text{g,bind}}$ [e]
Li$^+$	-27.2	≈ -25[c]	-31.8
Na$^+$	-18.8	-18.6[d]	-21.7
K$^+$	-11.8	-12.0[d]	-15.6
F$^-$	-19.1	≈ -20[c]	-24.0
Cl$^-$	-8.6	-	-11.8
Br$^-$	-7.3	-	-9.0

[a] Free energies calculated using ωB97X-D3/def2-TZVP//B3LYP-D3BJ/def2-SVP.

[b] Ref [269]

[c] Ref [247] (calculated using B3LYP), [d] Ref [237] (calculated using B3LYP).

[e] This work.

Table 8: Gas phase binding free energies of [X(H₂O)₄] in kcal/mol. [a]

Ions	Experiment [b]	Literature	$\Delta G^{\circ}_{\text{g,bind}}$ [e]	$\Delta G^{\circ}_{\text{g,bind}}$ [f]
Li$^+$	-66.9	≈ -65[c]	-62.5	-77.1
Na$^+$	-46.9	-49.4[d]	-41.0	-53.3
F$^-$	-46.3	≈ -45[c]	-49.8	-52.0
Cl$^-$	-23.7	-	-24.2	-29.5
Br$^-$	-20.3	-	-18.2	-26.3

[a] Free energies calculated using ωB97X-D3/def2-TZVP//B3LYP-D3BJ/def2-SVP.

[b] Ref [269]

[c] Ref [247] (calculated using B3LYP), [d] Ref [237] (calculated using B3LYP).

[e] $\Delta G^{\circ}_{\text{g,bind}}$ values are calculated using Scheme 1.

[f] $\Delta G^{\circ}_{\text{g,bind}}$ values are calculated using Scheme 2.

After benchmarking our calculations, we calculated hydration free energies using Eq. 4.6, which relies on the cluster cycle in Scheme 2. We considered a data set of ions having different sizes and charges of 2+, 1+, 1−, and 2−. Figures 29 and 30 show hydration free energies for Na^+, Mg^{2+}, Cl^-, SO_4^{2-}. Table 9 shows hydration free energies for all the solutes, and the percent error calculated by taking absolute hydration free energies from Marcus's study and adding the phase potential contribution taken from Lamoureux and Roux,[270] using Eq. 4.1 or comparing with experimental values from Tissandier et. al.[269] We also calculated and compare hydration free energies for ion pairs. Table 10 shows hydration free energies for ion pairs and how they agree with both Marcus et al. [267] and Tissandier et al.[269] For all ions, we report solvation free energies with both BP86-D3BJ and B3LYP-D3BJ geometries to compare the importance of including exact exchange in these systems. In all cases, B3LYP-D3BJ geometries result with more consistent solvation energies compared to BP86-D3BJ geometries.

In Figure 29a and 29c, for Na^+ and Mg^{2+} cations, respectively, the hydration free energies appear to get closer to the experimental data when we gradually increase the number of water molecules in the system. For Na^+ starting with eight water molecules we get good agreement with the experimental data. All monovalent cations we have considered (Li^+, Na^+, K^+) have errors under 5% percent which corresponds to ∼3 kcal/mol error. Similarly, for Mg^{2+} starting with 12 water molecules we get good agreement with the experimental data. All divalent cations in our library (Mg^{2+}, Ca^{2+}, Zn^{2+}) have errors under 2% which corresponds to ∼7 kcal/mol error. In summary, using Scheme 2 with increasing cluster sizes consistently decreases the errors for cation systems.

In Figures 30a and 30c, for Cl^- and SO_4^{2-} anions, respectively, the data is not as straight forward. The anion hydration free energies are not particularly sensitive to water cluster size, and hydration free energies begin to deviate more from experiment when using 16 and 20 water molecules. When we looked at three 1− charged anions, we observed larger percent errors (∼10% percent for Cl^- and Br^-) relative to the cation data. Lastly, we considered two 2− charged anions and calculated errors are around ∼4%, which corresponds to ∼10 kcal/mol error in the hydration free energy. Thus, our initial hypothesis that adding more solvent molecules into the system should improve the accuracy was true only for the cations.

Table 9: Comparison of experimental and calculated solvation free energies in kcal/mol. [a]

Ions	Experiment [b]	Experiment [c]	Cluster Size	ΔG_{solv} [f]	Error [g]	% Error
Li^+	-117.3	-126.5 [d]	8, 12	-124.4	2.1	1.7
Na^+	-91.0	-101.3 [d]	8, 12	-103.3	-2.0	2.0
K^+	-74.3	-84.1 [d]	8, 12	-80.2	3.9	4.6
F^-	-114.9	-102.5 [d]	8, 12	-106.0	-3.5	3.4
Cl^-	-85.0	-72.7 [d]	8, 12	-77.7	-5.0	6.8
Br^-	-79.1	-66.3 [d]	8, 12	-72.0	-5.7	8.5
Mg^{2+}	-441.2	-466.1 [e]	12, 16	-460.0	6.1	1.3
Ca^{2+}	-363.4	-388.4 [e]	12, 16	-381.8	6.6	1.6
Zn^{2+}	-471.0	-495.9 [e]	12, 16	-487.3	8.6	1.8
SO_4^{2-}	-261.9	-237.0 [e]	12, 16	-245.3	-8.3	3.5
CO_3^{2-}	-318.1	-293.2 [e]	12, 16	-282.4	10.8	3.7

[a] Free energies calculated using ωB97X-D3/def2-TZVP//B3LYP-D3BJ-def2-SVP.

[b] Experimental free energies taken from Ref.[267]

[c] Experimental free energies taken from Ref.[267] and then corrected using the air/water interface potential.

[d] Experimental free energies taken from Ref.[269] (Free energies taken from Ref.[267] and corrected using the phase potential correction (i.e. difference) listed in Table 3.

[e] Corrected using the phase potential correction (-12.45 kcal/mol) from Ref.[270].

[f] ΔG_{solv}^* calculated in this work.

[g] Error is calculated with respect to column Experiment[c].

Table 10: Comparison of experimental and calculated solvation free energies for ion pairs in kcal/mol. [a]

Ions	Experiment [b]	Experiment [c]	ΔG_{solv} [d]
Li^+ - F^-	-232.2	-229.0	-230.4
Li^+ - Cl^-	-202.3	-199.2	-202.1
Li^+ - Br^-	-196.4	-192.8	-196.4
Na^+ - F^-	-205.9	-203.8	-209.3
Na^+ - Cl^-	-176.0	-174.0	-181.0
Na^+ - Br^-	-170.1	-167.6	-175.3
K^+ - F^-	-189.2	-186.6	-186.2
K^+ - Cl^-	-159.3	-156.8	-157.9
K^+ - Br^-	-153.4	-150.4	-152.2

[a] Free energies calculated using ωB97X-D3/def2-TZVP//B3LYP-D3BJ-def2-SVP.

[b] Experimental free energies taken from Ref.[267]

[c] Experimental free energies taken from Ref.[269]

[d] $\Delta G_{\mathrm{solv}}^*$ calculated in this work.

Next we considered that different microsolvated clusters might have significantly different solute structures, which result in different hydration free energies, as shown in Figures 29 and 30. To test this idea, we studied geometric similarities in cluster sizes. We used the SOAP kernel to quantify the similarity between solute environments.[278, 194] For the high-dimensional pair-similarity data, we used 'sketch-maps', a non-linear dimensionality reduction technique.[302, 303] Sketch-maps allow us to obtain a two-dimensional map that provides a visual representation of the similarity between solvent environments.

We define the local solvation environment as a sphere of radius $r+2$ Å, centered at the solute atom, where r is the atomic radius. For complex anions, the environment center was chosen to be the position of the central atom, and r was taken to be the average central atom-oxygen distance plus the atomic radius of the oxygen atom. This cutoff was chosen to capture the evolution of the local solvation environment around the ion as increasing numbers of water molecules were added, while disregarding solvent molecules further from the ion. The atoms within the cutoff distance (including those of the solute) contribute to a smooth representation of the atom density, which is used to define a similarity measure between structures invariant to permutation of atom indices as well as rigid translations and rotations.

In our previous study using SOAP, we applied the pairwise similarity between configurations to comment on structural analogies.[218] To provide an intuitive representation of the relationships between all pairs within a group of structures, we used a sketch-map based on the SOAP metric. Each point on the map represents a solute environment. Data points in close proximity indicate systems with high similarity in local solvent environments. The sketch-map algorithm follows a non-linear optimization procedure where the discrepancy between pairwise Euclidean distances in low dimension and the kernel-induced metric is minimized. A sigmoid function is applied to focus the optimization on the most relevant range of distances, e.g. disregarding thermal fluctuations. The parameters of this "filter" are in the format, sigma-a-b. In all cases, we used a=3 and b=8, while sigma values were adapted to different systems following the heuristics described in Ref [304].
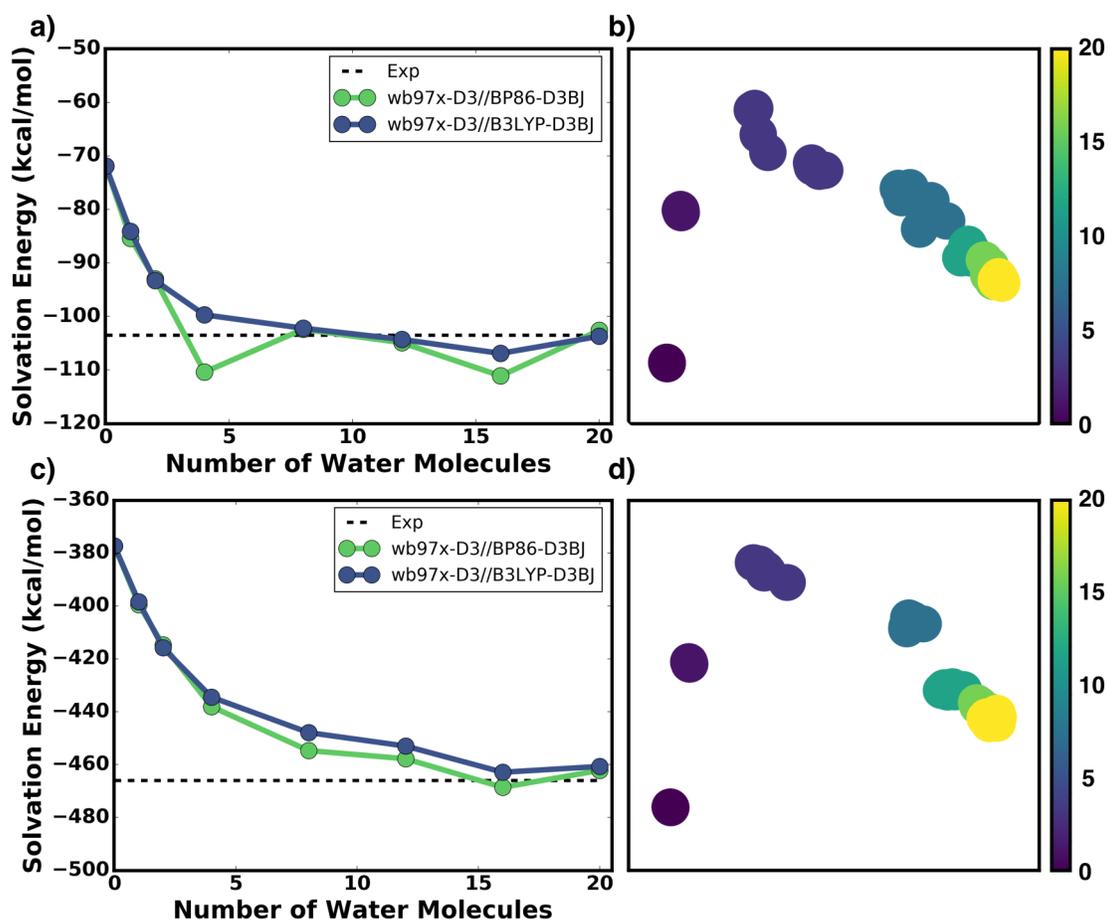
Figure 29: Hydration free energy plots and sketch-maps for Na$^+$ and Mg^{2+}. Plots show (a) hydration free energies calculated with Eq.4.6 for Na$^+$, (b) SOAP/sketch-map analysis for Na$^+$, (c) hydration free energies calculated with Eq.4.6 for Mg^{2+}, (d) SOAP/sketch-map analysis for Mg^{2+}. Data are from ωB97X-D3/def2-TZVP calculations on BP86-D3BJ/def2-SVP or B3LYP-D3BJ/def2-SVP geometries. Color bar shows the number of water molecules in the system.
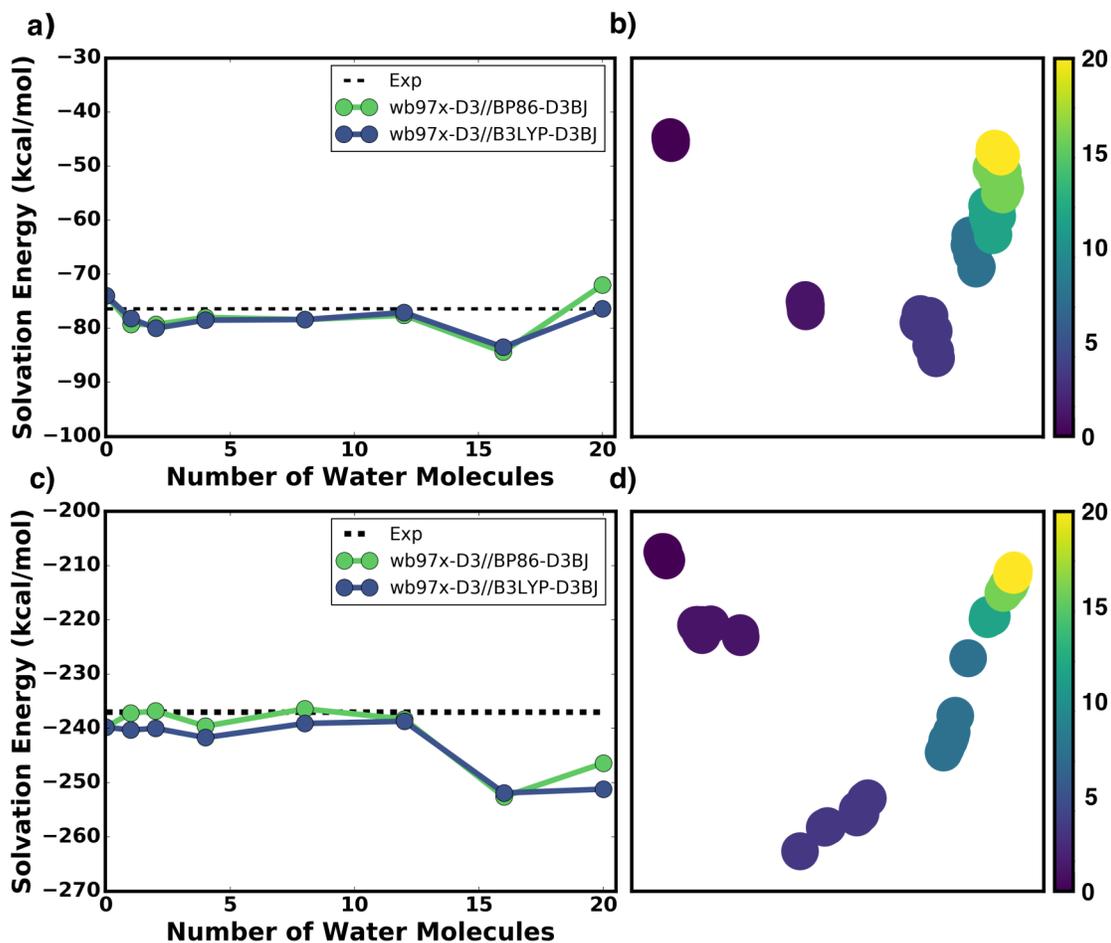
Figure 30: Hydration free energy plots and sketch-maps for Cl^- and SO_4^{2-}. Plots show (a) hydration free energies calculated Eq.4.6 for Cl^-, (b) SOAP/sketch-map analysis for Cl^-, (c) hydration free energies calculated with Eq.4.6 for SO_4^{2-}, (d) SOAP/sketch-map analysis for SO_4^{2-}. Data are from ωB97X-D3/def2-TZVP calculations on BP86-D3BJ/def2-SVP or B3LYP-D3BJ/def2-SVP geometries. Color bar shows the number of water molecules in the system.

Figure 29b demonstrates that models for Na^+ starting with eight or more water molecules give good agreement with the experimental values and all structures exhibit a similar solvent environment around the ion. Similarly for Mg^{2+}, Figure 29d shows a similar set of results except models should have 12 or more water molecules. Given the short-range cutoff for the SOAP descriptors, the results of this static calculation scheme with the cations indicate that hydration free energies gradually become more accurate while the local solvation environment around the cation also gradually becomes more similar with increasing cluster size.

However, this correspondence between structures and energetics is not observed for the anions (Cl^-, SO_4^{2-}). The sketch-map for Cl^- in Figure 30b shows that local solvation environments start becoming progressively similar with $n = 8$, but solvation free energies for larger clusters deviate by about 10 kcal/mol for clusters with 16 or more water molecules. A similar trend is seen on the sketch-map for SO_4^{2-} in Figure 30d. We see the local solvation structures start to become more similar with about $n = 12$, but the solvation free energies for the 16 and 20 molecule water clusters are inaccurate by 15 kcal/mol compared to experimental values. It is unlikely that the errors with the anions arise from not having a globally optimized $X(H_2O)_n^{m\pm}$ cluster since Eq.4.6 tells us that a more stable cluster would generally result in a more negative solvation energy. This can explain for calculated underpredictions of solvation energies (Table 9), and even more sampling of microsolvated structures might further improve these data. However, the calculated overpredictions seen with anions suggests that the errors shown in Figure 30 with the 16 and 20 molecule water clusters may arise from an imbalance in anharmonic effects in Scheme 2 because the ion-water clusters have more anharmonicity than the water-water clusters. Nevertheless, we can still use sketch-maps to identify the number of solvent molecules needed to see a analogous solvent arrangements. For example with Cl^-, the local solvation environment starts becoming similar with $n = 8$, and when we calculate the hydration free energy with eight water molecules we get a relatively good agreement with the experimental data without introducing errors we are currently attributing to anharmonicity. Similarly with SO_4^{2-}, we identify 12, 16, and 20 water clusters have similar solvent arrangements, but the smallest of these structures will have the least anharmonicity and thus lowest error. Table 9 summarizes all of our calculations and compares calculated hydration free energies with experimental

74

data. Thus, the sketch-map analysis appears to be useful for identifying how many solvent molecules are needed to calculate an accurate hydration free energy. Similar analysis is done for every ion in our library and shown in Figures 31-37.
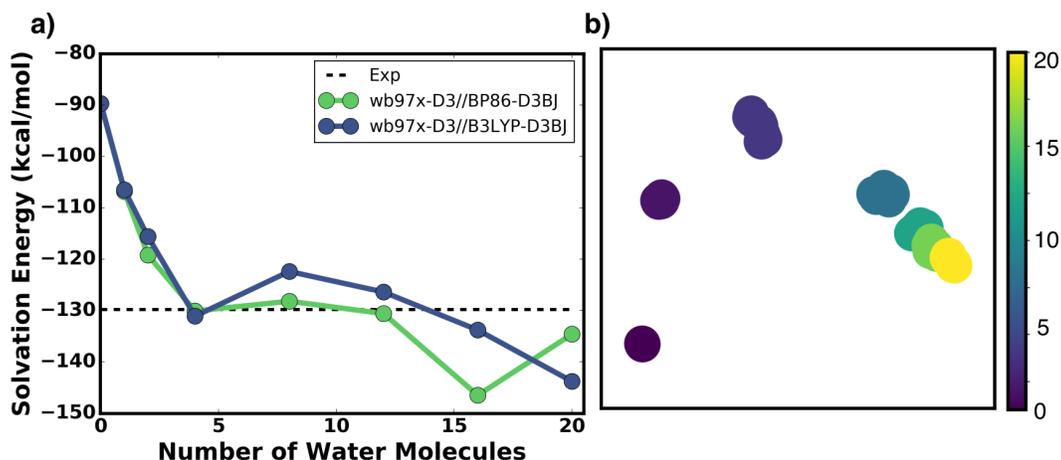
Figure 31: Hydration free energy plot and sketch-map for Li$^+$. Figure 31a shows hydration free energies calculated with Eq. 4.6 for Li$^+$. Figure 31b is the SOAP/sketch-map analysis for Li$^+$.
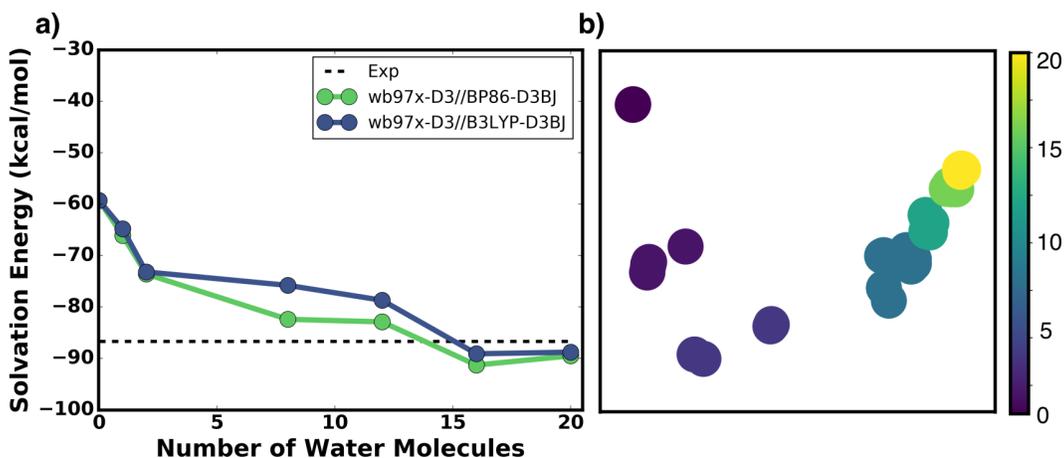
Figure 32: Hydration free energy plot and sketch-map for K$^+$. Figure 32a shows hydration free energies calculated with Eq. 4.6 for K$^+$. Figure 32b is the SOAP/sketch-map analysis for K$^+$.
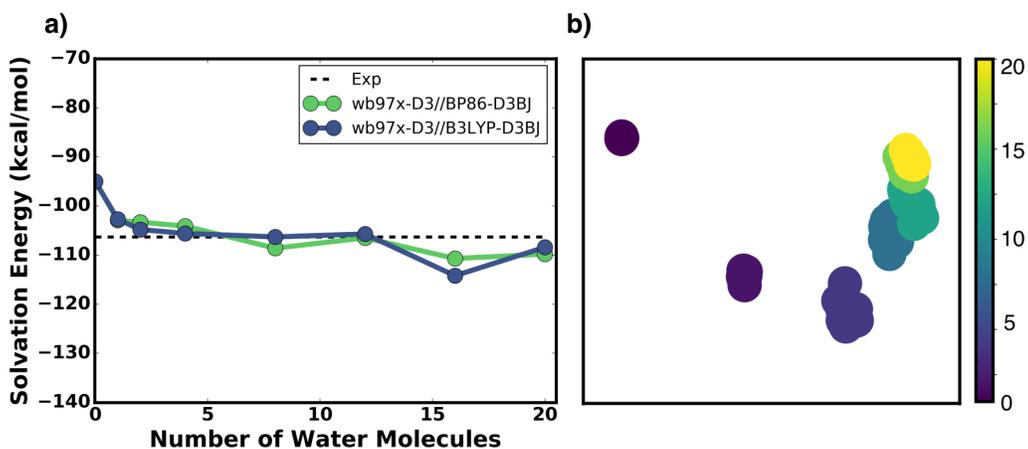
Figure 33: Hydration free energy plot and sketch-map for F$^-$. Figure 33a shows hydration free energies calculated with Eq. 4.6 for F$^-$. Figure 33b is the SOAP/sketch-map analysis for F$^-$.
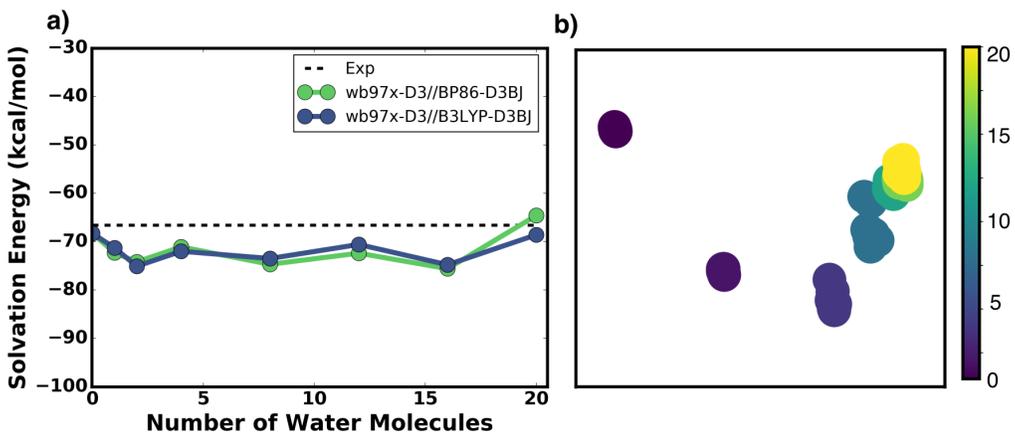
Figure 34: Hydration free energy plot and sketch-map for Br$^-$. Figure 34a shows hydration free energies calculated with Eq. 4.6 for Br$^-$. Figure 34b is the SOAP/sketch-map analysis for Br$^-$.
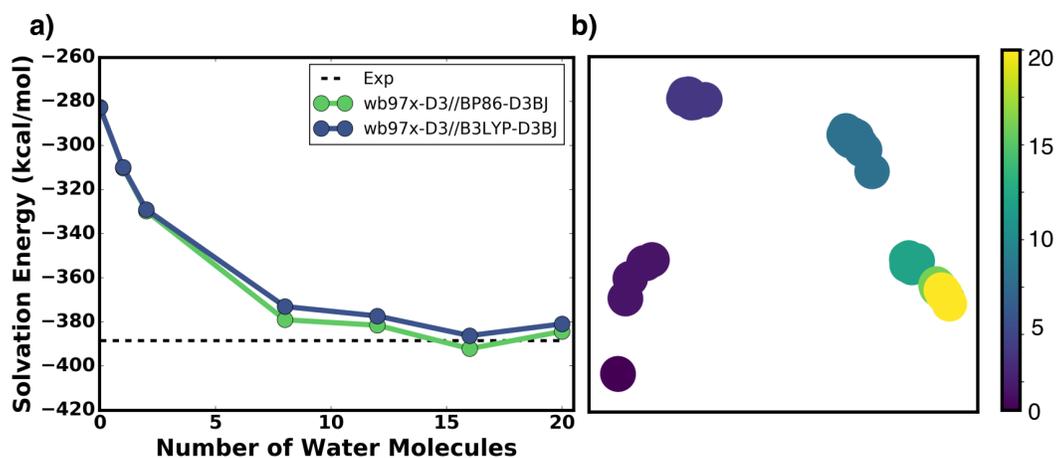
Figure 35: Hydration free energy plot and sketch-map for Ca^{2+}. Figure 35a shows hydration free energies calculated with Eq. 4.6 for Ca^{2+}. Figure 35b is the SOAP/sketch-map analysis for Ca^{2+}.
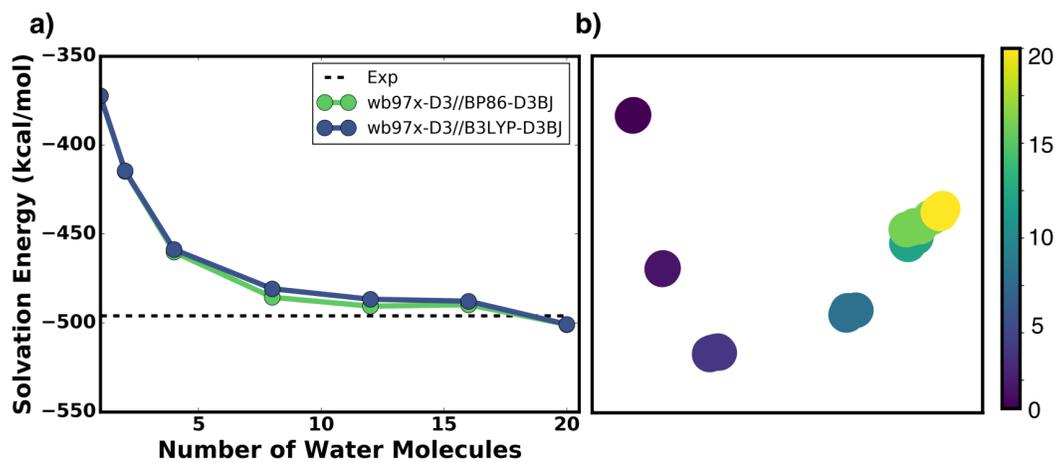
Figure 36: Hydration free energy plot and sketch-map for Zn^{2+}. Figure 36a shows hydration free energies calculated with Eq. 4.6 for Zn^{2+}. Figure 36b is the SOAP/sketch-map analysis for Zn^{2+}.
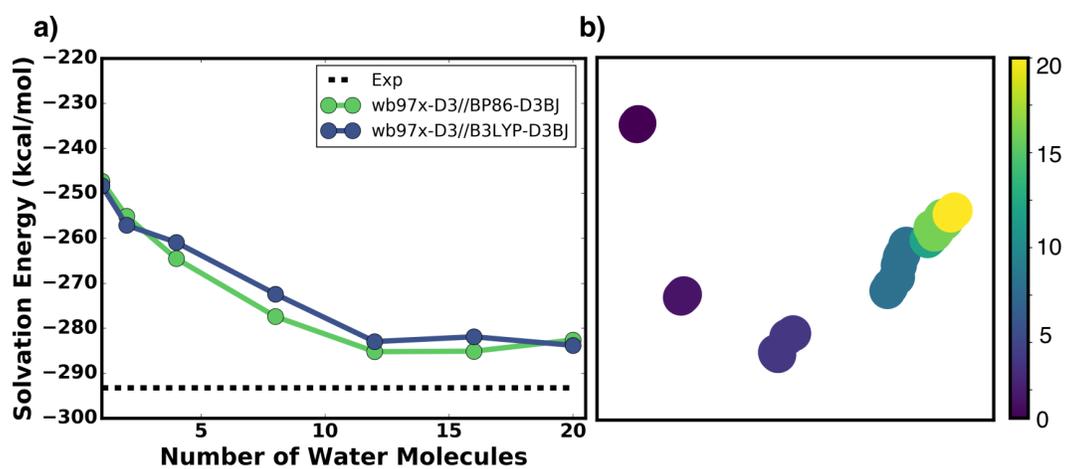
Figure 37: Hydration free energy plot and sketch-map for $CO_3{}^{2-}$. Figure 37a shows hydration free energies calculated with Eq. 4.6 for $CO_3{}^{2-}$. Figure 37b is the SOAP/sketch-map analysis for $CO_3{}^{2-}$.

To obtain a more detailed understanding of where the errors come from, we performed MD simulations using the AMEOBA force field[298] with the TINKER[192] software package to account for both chemical and thermal energy scales. We performed simulations for Na^+ and Cl^- and used a cubic box starting with 500 solvent molecules. We picked 100 frames from the trajectory and carved out clusters containing eight water molecules. Figure 38 shows SOAP/sketch-map analysis for these 100 structures and our DFT optimized structures. For Na^+, DFT optimization resulted in similar structures as those found from the MD simulation regardless of the functional used in the optimization. However for Cl^-, DFT optimization resulted in quite different structures than those found from the MD simulation. Again, we attribute this result to enhanced anharmonic effects in ion-water clusters with anions. The lack of error cancellation in Scheme 2 with anions suggests that one should perform dynamics simulations within Scheme 1 to capture the correct geometry for outer-shell contributions and correct for anharmonicity contributions in the inner-shell contributions to hydration free energy, as demonstrated recently.[248, 247, 250]
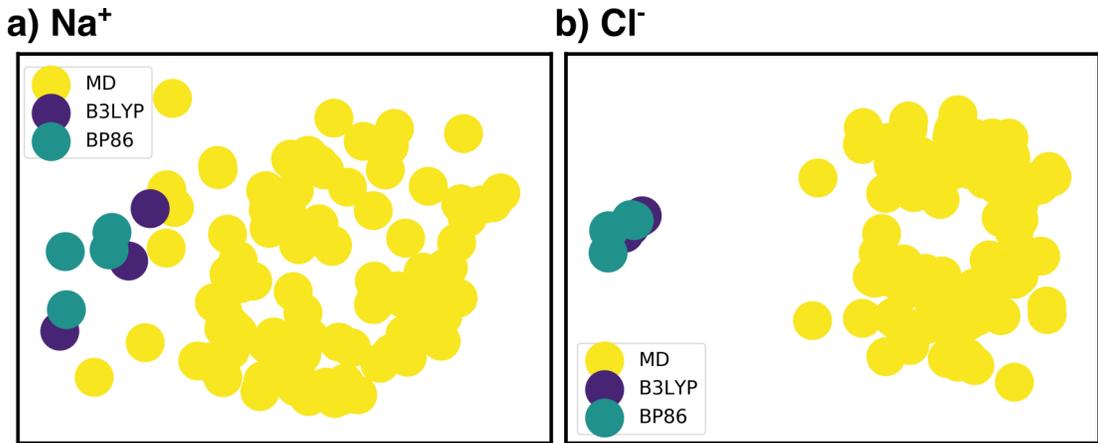
Figure 38: SOAP/sketch-map analysis for Na^+ and Cl^- with eight water clusters. Yellow data points are obtained from an MD trajectory, purple and green data points are obtained from a full QM optimization in this work.

Therefore, in the absence of well-parameterized force fields or computationally intensive BOMD simulations, we propose the following practical treatment (that exploits error cancellations) for automated calculations of real solvation free energies: 1) Calculate solvation

free energies using the thermodynamic cycle in Scheme 2 for various values of n. 2) Use a SOAP/sketch-map analysis to identify a relatively similar local solvent environment. 3) Use the smallest possible cluster to calculate solvation free energy contributions. In most cases, particularly with cations, using larger cluster sizes improves the calculated solvation free energies. However, for anions (as demonstrated above) there may be unbalanced errors that are known to arise from a CSM analysis of outer-shell contributions combined with harmonic analysis of inner-shell contributions, and using cluster sizes with more than 15 water molecules may cause errors as high as 10 kcal/mole. By using unsupervised machine learning (i.e. the SOAP/sketch-map analysis) we identify salient solvation environments needed for solvation free energy calculations using static microsolvated clusters.

4.5   Conclusions

We have demonstrated an automatable cluster-continuum (i.e., mixed implicit/explicit) modeling approach that leverages unsupervised machine learning to calculate solvation free energies of ions and small molecules. We showed that Scheme 1 has practical applications for small values of n when CSM models are correctly accounting for outer-shell contributions, and these results correspond more to absolute solvation free energies that do not contain surface potential contributions. In contrast, Scheme 2 has practical applications for larger values of n, and error cancellation from this scheme gives results that appear to correspond to real solvation energies. We also showed how adding explicit solvent molecules improves calculated solvation free energies by creating a more physical local solvation environment around cations, but adding too many solvent molecules can lead to significant errors due to an imbalance in anharmonic and harmonic energy contributions that are more predominant for solvated anions. Overall, we show a generalizable approach to systematically investigate atomic scale microsolvation environments and predict solvation free energies of ions in water and presumably other general solvents.

5.0 Publications and Future Work

5.1 Publications

5.1.1 Current Publications

1. **Basdogan, Y.**, Maldonado, A and Keith, J. (2020). Advances and challenges in modeling solvated reaction mechanisms for renewable fuels and chemicals. *Wiley Interdiscip. Rev. Comput. Mol. Sci. DOI: 10.1002/wcms.1446*

2. **Basdogan, Y.**, Groenenboom, M. C., Rempe, S., and Keith, J. (2019). Machine Learning Guided Approach for Studying Solvation Environments. *J. Chem. Theory. Comp.*

3. **Basdogan, Y.**, and Keith, J. A. (2018). A paramedic treatment for modeling explicitly solvated chemical reaction mechanisms. *Chem. Sci, 9(24), 5341-5346.*

4. Vo, M. N., **Basdogan, Y.**, Derksen, B. S., Kowall, C., ... and Johnson, J. K. (2018). Mechanism of Isobutylene Polymerization: Quantum Chemical Insight into AlCl3/H2O-Catalyzed Reactions. *ACS Catal., 8(9), 8006-8013.*

5. Ilic, S., **Basdogan, Y.**, Keith, J. A., and Glusac, K. D. (2018). Thermodynamic hydricities of biomimetic organic hydride donors. *J. Am. Chem. Soc., 140(13), 4569-4579.*

6. Groenenboom, M. C., Anderson, R. M., Horton, D. J., **Basdogan, Y.**, Roeper, D. F., Policastro, S. A., and Keith, J. A. (2017). Doped amorphous Ti oxides to deoptimize oxygen reduction reaction catalysis. *J. Phys. Chem. C., 121(31), 16825-16830.*

7. Saravanan, K., **Basdogan, Y.**, Dean, J., and Keith, J. A. (2017). Computational investigation of CO_2 electroreduction on tin oxide and predictions of Ti, V, Nb and Zr dopants for improved catalysis. *J. Mater. Chem. A, 5(23), 11756-11763.*

5.1.2 Expected Publications

8. Wang Y*., **Basdogan, Y.***, Keith, J. A. and Gilbertson L. M. (2020). Identifying the role of different oxygen functional groups in the graphene mediated glutathione oxidation reaction. ***Submitted (co-first authorship)***

9. Maldonado, A., **Basdogan, Y.**, Rempe, S., and Keith, J. (2020). Modeling reactions in mixed solvents: Where to go from here? ***Submitted to an invited issue of J. Chem Phys.***

10. **Basdogan, Y.**, Choi, T. H., and Keith, J. A. (2020). Understanding solvation effects on hydrogenation barriers for CO_2 reduction on carbon-based materials. ***In Preparation.***

11. Maldonado, A., **Basdogan, Y.**, and Keith, J. (2021). Using machine learning to model reactions in mixed solvents. ***In Preparation.***

12. Choi, T. H., **Basdogan, Y.**, and Keith, J. A. (2021). Modelling CO_2 Reduction using aromatic N-heterocycles. ***In Preparation.***

5.2 Future Work

5.2.1 Using Machine Learning and First Principles Modeling of Chemistry in Mixed Solvents

Mixed solvents (i.e. binary or higher order mixtures of ionic or non-ionic liquids) play crucial roles in chemical syntheses, separations, and electrochemical devices because they can be tuned for specific reactions and applications. Apart from fully explicit solvation treatments that can be difficult to parameterize or computationally expensive, there is currently no well-established first-principles regimen for reliably modeling atomic-scale chemistry in mixed solvent environments. The approach demonstrated in this thesis should be applicable to study mixed solvent systems and calculate solvation free energies of any ion in any complicated solvation environment. The theoretical foundation and robustness of QCT effectively splits the problem of mixed solvents into inner and outer contributions as with pure solvents. The dual QCT approach in the cluster cycle should make solvation free energies tractable in mixed solvents without explicit simulations. Standard implicit solvation models treat mixed solvents as a homogeneous medium which should be appropriate for the outer shell contribution where there is no preferential solvation. Then, the SOAP and sketch-map analysis will guide selection of microsolvated clusters. We have started studying mixed solvents by using ABCluster as a global cluster optimization tool, imposing a local solvent environment selection criteria through SOAP/sketch-map analysis, and using a dual QCT approach with the cluster thermodynamic cycle. We believe that this approach is promising because it combines the theoretical rigour of QCT with a state-of-the-art automatable solvation analysis that should be extendable to mixed solvents.

Bibliography

[1] R Erik Plata and Daniel A Singleton. A case study of the mechanism of alcohol-mediated morita baylis–hillman reactions. the importance of experimental observations. *Journal of the American Chemical Society*, 137(11):3811–3826, 2015.

[2] Zhen Liu, Chandan Patel, Jeremy N. Harvey, and Raghavan B. Sunoj. Mechanism and reactivity in the morita–baylis–hillman reaction: the challenge of accurate computations. *Phys. Chem. Chem. Phys.*, 19:30647–30657, 2017.

[3] Ibrahim Dincer. Renewable energy and sustainable development: a crucial review. *Renewable and sustainable energy reviews*, 4(2):157–175, 2000.

[4] Phebe Asantewaa Owusu and Samuel Asumadu-Sarkodie. A review of renewable energy sources, sustainability issues and climate change mitigation. *Cogent Engineering*, 3(1):1167990, 2016.

[5] Zhi Wei Seh, Jakob Kibsgaard, Colin F Dickens, Ib Chorkendorff, Jens K Nørskov, and Thomas F Jaramillo. Combining theory and experiment in electrocatalysis: Insights into materials design. *Science*, 355(6321):eaad4998, 2017.

[6] Morris A Bender, Thomas R Knutson, Robert E Tuleya, Joseph J Sirutis, Gabriel A Vecchi, Stephen T Garner, and Isaac M Held. Modeled impact of anthropogenic warming on the frequency of intense atlantic hurricanes. *Science*, 327(5964):454–458, 2010.

[7] Hervé Le Treut. Historical overview of climate change. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, 2007.

[8] U.S. Energy Information Administration. International energy outlook 2018. July 30, 2018.

[9] Mary Rakowski Dubois and Daniel L Dubois. Development of molecular electrocatalysts for CO_2 reduction and H_2 production/oxidation. *Acc. Chem. Res.*, 42(12):1974–1982, 2009.

[10] Xiaorong Zhu and Yafei Li. Review of two-dimensional materials for electrochemical CO_2 reduction from a theoretical perspective. *WIRES: Comput. Mol. Sci.*, page e1416.

[11] Shenzhen Xu and Emily A Carter. Theoretical insights into heterogeneous (photo)electrochemical CO_2 reduction. *Chem. Rev.*, 2018.

[12] Stephanie Nitopi, Erlend Bertheussen, Soren B Scott, Xinyan Liu, Albert K Engstfeld, Sebastian Horch, Brian Seger, Ifan EL Stephens, Karen Chan, Christopher Hahn, Jens K. Nørskov, Thomas F. Jaramillo, and Ib Chorkendorff. Progress and perspectives of electrochemical CO_2 reduction on copper in aqueous electrolyte. *Chem. Rev.*, 2019.

[13] Wan-Hui Wang, Yuichiro Himeda, James T Muckerman, Gerald F Manbeck, and Etsuko Fujita. CO_2 hydrogenation to formate and methanol as an alternative to photo- and electrochemical CO_2 reduction. *Chem. Rev.*, 115(23):12936–12973, 2015.

[14] Huimin Liu, Li Wei, Fei Liu, Zengxia Pei, Jeffrey Shi, Zhou-jun Wang, Dehua He, and Yuan Chen. Homogeneous, heterogeneous, and biological catalysts for electrochemical N_2 reduction towards NH_3 under ambient conditions. *ACS Catal.*, 9(6):5245–5267, 2019.

[15] Brian M Lindley, Aaron M Appel, Karsten Krogh-Jespersen, James M Mayer, and Alexander JM Miller. Evaluating the thermodynamics of electrocatalytic N_2 reduction in acetonitrile. *ACS Energy Letters*, 1(4):698–704, 2016.

[16] Wangchuk Rabten, Markus D Kärkäs, Torbjörn Åkermark, Hong Chen, Rong-Zhen Liao, Fredrik Tinnis, Junliang Sun, Per EM Siegbahn, Pher G Andersson, and Bjoörn Åkermark. Catalytic water oxidation by a molecular ruthenium complex: unexpected generation of a single-site water oxidation catalyst. *Inorg. Chem.*, 54(10):4611–4620, 2015.

[17] Shunsuke Sato, Takeo Arai, Takeshi Morikawa, Keiko Uemura, Tomiko M Suzuki, Hiromitsu Tanaka, and Tsutomu Kajino. Selective CO_2 conversion to formate conjugated with H_2O oxidation utilizing semiconductor/complex hybrid photocatalysts. *J. Am. Chem. Soc.*, 133(39):15240–15243, 2011.

[18] Matthew Jouny, Wesley Luc, and Feng Jiao. General techno-economic analysis of CO_2 electrolysis systems. *Ind. Eng. Chem. Res.*, 57(6):2165–2177, 2018.

[19] Jingguang G Chen, Richard M Crooks, Lance C Seefeldt, Kara L Bren, R Morris Bullock, Marcetta Y Darensbourg, Patrick L Holland, Brian Hoffman, Michael J Janik, Anne K Jones, Mercouri G. Kanatzidis, Paul King, Kyle M. Lancaster, Sergei V. Lymar, Peter Pfromm, William F. Schneider, and Richard R. Schrock. Beyond fossil fuel–driven nitrogen transformations. *Science*, 360(6391):eaar6611, 2018.

[20] J Li, R Güttinger, R Moré, F Song, W Wan, and GR Patzke. Frontiers of water oxidation: the quest for true catalysts. *Chem. Soc. Rev.*, 46(20):6124–6147, 2017.

[21] Wenhui Li, Haozhi Wang, Xiao Jiang, Jie Zhu, Zhongmin Liu, Xinwen Guo, and Chunshan Song. A short review of recent advances in CO_2 hydrogenation to hydrocarbons over heterogeneous catalysts. *RSC Adv.*, 8(14):7651–7669, 2018.

[22] Qi Lu and Feng Jiao. Electrochemical CO_2 reduction: Electrocatalyst, reaction mechanism, and process engineering. *Nano Energy*, 29:439–456, 2016.

[23] James T Muckerman, Patrick Achord, Carol Creutz, Dmitry E Polyansky, and Etsuko Fujita. Calculation of thermodynamic hydricities and the design of hydride donors for CO_2 reduction. *Proc. Nat. Acad. Sci.*, 109(39):15657–15662, 2012.

[24] Sharon Hammes-Schiffer. Theory of proton-coupled electron transfer in energy conversion processes. *Acc. Chem. Res.*, 42(12):1881–1889, 2009.

[25] Sharon Hammes-Schiffer and Alexei A Stuchebrukhov. Theory of coupled electron and proton transfer reactions. *Chem. Rev.*, 110(12):6939–6960, 2010.

[26] Sharon Hammes-Schiffer and Alexander V Soudackov. Proton-coupled electron transfer in solution, proteins, and electrochemistry. *J. Phys. Chem. B*, 112(45):14108–14123, 2008.

[27] Jeffrey J Warren, Tristan A Tronic, and James M Mayer. Thermochemistry of proton-coupled electron transfer reagents and its implications. *Chem. Rev.*, 110(12):6961–7001, 2010.

[28] James M Mayer. Proton-coupled electron transfer: a reaction chemist's view. *Annual Review of Physical Chemistry*, 55:363–390, 2004.

[29] Cyrille Costentin, Marc Robert, and Jean-Michel Savéant. Catalysis of the electrochemical reduction of carbon dioxide. *Chemical Society Reviews*, 42(6):2423–2436, 2013.

[30] Cyrille Costentin, Marc Robert, and Jean-Michel Savéant. Concerted proton-electron transfers: Electrochemical and related approaches. *Acc. Chem. Res.*, 43(7):1019–1029, 2010.

[31] Cyrille Costentin, Marc Robert, and Jean-Michel Savéant. Electrochemical and homogeneous proton-coupled electron transfers: Concerted pathways in the one-electron oxidation of a phenol coupled with an intramolecular amine-driven proton transfer. *J. Am. Chem. Soc.*, 128(14):4552–4553, 2006.

[32] Luis GC Rego and Victor S Batista. Quantum dynamics simulations of interfacial electron transfer in sensitized TiO_2 semiconductors. *J. Am. Chem. Soc.*, 125(26):7989–7997, 2003.

[33] Sharon Hammes-Schiffer and Stephen J Benkovic. Relating protein motion to catalysis. *Annual Review of Biochemistry*, 75:519–541, 2006.

[34] J Jacq. Schema carre: Etablissement et discussion de l'equation generale de la courbe intensite-potentiel en regime stationnaire et diffusion convective. *J. Electroanal. Chem. and Interfacial Electrochemistry*, 29(1):149–180, 1971.

[35] John A Keith, Kyle A Grice, Clifford P Kubiak, and Emily A Carter. Elucidation of the selectivity of proton-dependent electrocatalytic CO_2 reduction by $fac−Re(bpy)(CO)_3Cl$. *J. Am. Chem. Soc.*, 135(42):15823–15829, 2013.

[36] Karsten Reuter and Matthias Scheffler. First-principles atomistic thermodynamics for oxidation catalysis: Surface phase diagrams and catalytically interesting regions. *Phys. Rev. Lett.*, 90:046103, 2003.

[37] Aleksandr V Marenich, Junming Ho, Michelle L Coote, Christopher J Cramer, and Donald G Truhlar. Computational electrochemistry: prediction of liquid-phase reduction potentials. *Phys. Chem. Chem. Phys.*, 16(29):15068–15106, 2014.

[38] Marcel Pourbaix. Atlas of electrochemical equilibria in aqueous solution. *NACE*, 307, 1974.

[39] Aude Marjolin and John A Keith. Thermodynamic descriptors for molecules that catalyze efficient CO_2 electroreductions. *ACS Catal.*, 5(2):1123–1130, 2015.

[40] Mitchell C Groenenboom, Karthikeyan Saravanan, Yaqun Zhu, Jeffrey M Carr, Aude Marjolin, Gabriel G Faura, Eric C Yu, Raymond N Dominey, and John A Keith.

Structural and substituent group effects on multielectron standard reduction potentials of aromatic n-heterocycles. *J. Phys. Chem. A*, 120(34):6888–6894, 2016.

[41] K Saravanan and JA Keith. Standard redox potentials, pK$_a$s, and hydricities of inorganic complexes under electrochemical conditions and implications for CO_2 reduction. *Dalton Trans.*, 45(39):15336–15341, 2016.

[42] Karthikeyan Saravanan, Eric Gottlieb, and John A Keith. Nitrogen-doped nanocarbon materials under electroreduction operating conditions and implications for electrocatalysis of CO_2. *Carbon*, 111:859–866, 2017.

[43] Karthikeyan Saravanan, Yasemin Basdogan, James Dean, and John A Keith. Computational investigation of CO_2 electroreduction on tin oxide and predictions of ti, v, nb and zr dopants for improved catalysis. *Journal of Materials Chemistry A*, 5(23):11756–11763, 2017.

[44] Jean-Michel Savéant and Cédric Tard. Attempts to catalyze the electrochemical CO_2-to-methanol conversion by biomimetic $2\,e^- + 2\,H^+$ transferring molecules. *J. Am. Chem. Soc.*, 138(3):1017–1021, 2016.

[45] Cyrille Costentin, Jean-Michel Savéant, and Cédric Tard. Catalysis of CO_2 electrochemical reduction by protonated pyridine and similar molecules. useful lessons from a methodological misadventure. *ACS Energy Letters*, 3(3):695–703, 2018.

[46] Shunsuke Sato and Takeshi Morikawa. [Ir(tpy)(bpy)Cl] as a photocatalyst for CO_2 reduction under visible-light irradiation. *ChemPhotoChem*, 2(3):207–212, 2018.

[47] Patrick K Giesbrecht, Dion B Nemez, and David E Herbert. Electrochemical hydrogenation of a benzannulated pyridine to a dihydropyridine in acidic solution. *Chemical Communications*, 54(4):338–341, 2018.

[48] Jens Kehlet Nörskov, Jan Rossmeisl, Ashildur Logadottir, LRKJ Lindqvist, John R Kitchin, Thomas Bligaard, and Hannes Jonsson. Origin of the overpotential for oxygen reduction at a fuel-cell cathode. *J. Phys. Chem. B*, 108(46):17886–17892, 2004.

[49] Kai S Exner and Herbert Over. Kinetics of electrocatalytic reactions from first-principles: a critical comparison with the ab initio thermodynamics approach. *Accounts of chemical research*, 50(5):1240–1247, 2017.

[50] Xiaowa Nie, Monica R Esopi, Michael J Janik, and Aravind Asthagiri. Selectivity of CO_2 reduction on copper electrodes: the role of the kinetics of elementary steps. *Angew. Chem. Int. Ed.*, 52(9):2459–2462, 2013.

[51] Emily A Carter. Challenges in modeling materials properties without experimental input. *Science*, 321(5890):800–803, 2008.

[52] Renat R Nazmutdinov, Michael D Bronshtein, Tamara T Zinkicheva, and Dmitrii V Glukhov. Modeling of electron transfer across electrochemical interfaces: State-of-the art and challenges for quantum and computational chemistry. *Int. J. Quantum Chem.*, 116(3):189–201, 2016.

[53] Federico Calle-Vallejo and Marc TM Koper. First-principles computational electrochemistry: Achievements and challenges. *Electrochim. Acta*, 84:3–11, 2012.

[54] Paul M Zimmerman. Automated discovery of chemically reasonable elementary reaction steps. *J. Comput. Chem.*, 34(16):1385–1392, 2013.

[55] Andrew Behn, Paul M Zimmerman, Alexis T Bell, and Martin Head-Gordon. Efficient exploration of reaction paths via a freezing string method. *J. Chem. Phys.*, 135(22):224108, 2011.

[56] Christopher D Fu, Luiz FL Oliveira, and Jim Pfaendtner. Determining energy barriers and selectivities of a multi-pathway system with infrequent metadynamics. *J. Chem. Phys.*, 146(1):014108, 2017.

[57] Georg Kastlunger, Per Lindgren, and Andrew A Peterson. Controlled-potential simulation of elementary electrochemical reactions: Proton discharge on metal surfaces. *J. Phys. Chem. C*, 122(24):12771–12781, 2018.

[58] Sneha A Akhade, Nicole J Bernstein, Monica R Esopi, Michael J Regula, and Michael J Janik. A simple method to approximate electrode potential-dependent activation energies using density functional theory. *Catal. Today*, 288:63–73, 2017.

[59] Tao Cheng, Hai Xiao, and William A Goddard III. Free-energy barriers and reaction mechanisms for the electrochemical reduction of co on the cu(100) surface, including multiple layers of explicit solvent at ph 0. *J. Phys. Chem. Lett.*, 6(23):4767–4773, 2015.

[60] Meenesh R Singh, Jason D Goodpaster, Adam Z Weber, Martin Head-Gordon, and Alexis T Bell. Mechanistic insights into electrochemical reduction of CO_2 over ag using density functional theory and transport models. *Proc. Nat. Acad. Sci.*, 114(42):E8812–E8821, 2017.

[61] Karen Chan and Jens K Nørskov. Electrochemical barriers made simple. *J. Phys. Chem. Lett.*, 6(14):2663–2668, 2015.

[62] Christopher D Taylor, Sally A Wasileski, Jean-Sebastien Filhol, and Matthew Neurock. First principles reaction modeling of the electrochemical interface: Consideration and calculation of a tunable surface potential from atomic and electronic structure. *Phys. Rev. B*, 73(16):165402, 2006.

[63] Sneha A Akhade, Robert M Nidzyn, Gholamreza Rostamikia, and Michael J Janik. Using brønsted-evans-polanyi relations to predict electrode potential-dependent activation energies. *Catal. Today*, 312:82–91, 2018.

[64] Ling Liu and Chungen Liu. Origin of the overpotentials for hcoo- and co formation in the electroreduction of co 2 on cu (211): the reductive desorption processes decide. *Phys. Chem. Chem. Phys.*, 20(8):5756–5765, 2018.

[65] Federico Calle-Vallejo, Rodrigo Ferreira de Morais, Francesc Illas, David Loffreda, and Philippe Sautet. Affordable estimation of solvation contributions to the adsorption energies of oxygenates at metal nanoparticles. *J. Phys. Chem. C*, 123(9):5578–5582, 2019.

[66] Stephan N Steinmann, Philippe Sautet, and Carine Michel. Solvation free energies for periodic surfaces: comparison of implicit and explicit solvation models. *Phys. Chem. Chem. Phys.*, 18(46):31850–31861, 2016.

[67] Sina Behtash, Jianmin Lu, Osman Mamun, Christopher T Williams, John R Monnier, and Andreas Heyden. Solvation effects in the hydrodeoxygenation of propanoic acid over a model pd(211) catalyst. *J. Phys. Chem. C*, 120(5):2724–2736, 2016.

[68] Hankyul Lee, Hyung-Kyu Lim, and Hyungjun Kim. Hydration thermodynamics of non-polar aromatic hydrocarbons: Comparison of implicit and explicit solvation models. *Molecules*, 23(11):2927, 2018.

[69] Xiaohong Zhang, Torrie E Sewell, Brittany Glatz, Sapna Sarupria, and Rachel B Getman. On the water structure at hydrophobic interfaces and the roles of water on transition-metal catalyzed reactions: A short review. *Catal. Today*, 285:57–64, 2017.

[70] Jacopo Tomasi and Maurizio Persico. Molecular interactions in solution: an overview of methods based on continuous distributions of the solvent. *Chemical Reviews*, 94(7):2027–2094, 1994.

[71] Jacopo Tomasi. Selected features of the polarizable continuum model for the representation of solvation. *WIRES: Comput. Mol. Sci.*, 1(5):855–867, 2011.

[72] Andreas Klamt. The cosmo and cosmo-rs solvation models. *WIRES: Comput. Mol. Sci.*, 1(5):699–709, 2011.

[73] Matthias Heyden. Disassembling solvation free energies into local contributions——toward a microscopic understanding of solvation processes. *WIRES: Comput. Mol. Sci.*, page e1390.

[74] Christopher J Cramer and Donald G Truhlar. Implicit solvation models: equilibria, structure, spectra, and dynamics. *Chem. Rev.*, 99(8):2161–2200, 1999.

[75] Jacopo Tomasi, Benedetta Mennucci, and Roberto Cammi. Quantum mechanical continuum solvation models. *Chemical reviews*, 105(8):2999–3094, 2005.

[76] Christopher J Cramer and Donald G Truhlar. A universal approach to solvation modeling. *Acc. Chem. Res.*, 41(6):760–768, 2008.

[77] Maurizio Cossi, Nadia Rega, Giovanni Scalmani, and Vincenzo Barone. Energies, structures, and electronic properties of molecules in solution with the c-pcm solvation model. *J. Comput. Chem.*, 24(6):669–681, 2003.

[78] S Miertuš, E Scrocco, and J Tomasi. Electrostatic interaction of a solute with a continuum. a direct utilizaion of ab initio molecular potentials for the prevision of solvent effects. *Chemical Physics*, 55(1):117–129, 1981.

[79] Maurizio Cossi, Vincenzo Barone, Roberto Cammi, and Jacopo Tomasi. Ab initio study of solvated molecules: a new implementation of the polarizable continuum model. *Chem. Phys. Lett.*, 255(4–6):327–335, 1996.

[80] Claudio Amovilli, Vincenzo Barone, Roberto Cammi, Eric Cancès, Maurizio Cossi, Benedetta Mennucci, Christian S Pomelli, and Jacopo Tomasi. Recent advances in the description of solvent effects with the polarizable continuum model. In *Adv. Quantum Chem.*, volume 32, pages 227–261. Elsevier, 1998.

[81] E Cancès, Benedetta Mennucci, and J Tomasi. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. *J. Chem. Phys.*, 107(8):3032–3041, 1997.

[82] Benedetta Mennucci, E Cancès, and J Tomasi. Evaluation of solvent effects in isotropic and anisotropic dielectrics and in ionic solutions with a unified integral equation method: theoretical bases, computational implementation, and numerical applications. *J. Phys. Chem. B*, 101(49):10506–10517, 1997.

[83] Maurizio Cossi and Vincenzo Barone. Time-dependent density functional theory for molecules in liquid solutions. *J. Chem. Phys.*, 115(10):4708–4717, 2001.

[84] Aleksandr V Marenich, Christopher J Cramer, and Donald G Truhlar. 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. *The Journal of Physical Chemistry B*, 113(18):6378–6396, 2009.

[85] Andreas Klamt and GJGJ Schüürmann. Cosmo: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society, Perkin Transactions 2*, (5):799–805, 1993.

[86] Andreas Klamt. *COSMO-RS: from quantum chemistry to fluid phase thermodynamics and drug design.* Elsevier, 2005.

[87] Paweł Grochowski and Joanna Trylska. Continuum molecular electrostatics, salt effects, and counterion binding–a review of the poisson–boltzmann theory and its modifications. *Biopolymers*, 89(2):93–113, 2008.

[88] Jean-Luc Fattebert and François Gygi. First-principles molecular dynamics simulations in a continuum solvent. *Int. J. Quantum Chem.*, 93(2):139–147, 2003.

[89] SA Petrosyan, AA Rigos, and TA Arias. Joint density-functional theory: Ab initio study of cr2o3 surface chemistry in solution. *J. Phys. Chem. B*, 109(32):15436–15444, 2005.

[90] Deniz Gunceler, Kendra Letchworth-Weaver, Ravishankar Sundararaman, Kathleen A Schwarz, and TA Arias. The importance of nonlinear fluid response in joint density-functional theory studies of battery systems. *Model. Simul. Mater. Sci. Eng.*, 21(7):074005, 2013.

[91] Oliviero Andreussi, Ismaila Dabo, and Nicola Marzari. Revised self-consistent continuum solvation in electronic-structure calculations. *The Journal of chemical physics*, 136(6):064102, 2012.

[92] Giuseppe Fisicaro, Luigi Genovese, Oliviero Andreussi, Nicola Marzari, and Stefan Goedecker. A generalized poisson and poisson-boltzmann solver for electrostatic environments. *J. Chem. Phys.*, 144(1):014103, 2016.

[93] C Dupont, Oliviero Andreussi, and N Marzari. Self-consistent continuum solvation (sccs): The case of charged systems. *J. Chem. Phys.*, 139(21):214110, 2013.

[94] Ravishankar Sundararaman and William A Goddard III. The charge-asymmetric nonlocally determined local-electric (candle) solvation model. *The Journal of chemical physics*, 142(6):064107, 2015.

[95] Matthew Truscott and Oliviero Andreussi. Field-aware interfaces in continuum solvation. *J. Phys. Chem. B*, 123(16):3513–3524, 2019.

[96] Joseph A Gauthier, Stefan Ringe, Colin F Dickens, Alejandro J Garza, Alexis T Bell, Martin Head-Gordon, Jens K Nørskov, and Karen Chan. Challenges in modeling electrochemical reaction energetics with polarizable continuum models. *ACS Catal.*, 9:920–931, 2018.

[97] Giuseppe Fisicaro, Luigi Genovese, Oliviero Andreussi, Sagarmoy Mandal, Nisanth N Nair, Nicola Marzari, and Stefan Goedecker. Soft-sphere continuum solvation in electronic-structure calculations. *J. Chem. Theory Comput.*, 13(8):3829–3845, 2017.

[98] Oliviero Andreussi, Nicholas Georg Hörmann, Francesco Nattino, Giuseppe Fisicaro, Stefan Goedecker, and Nicola Marzari. Solvent-aware interfaces in continuum solvation. *J. Chem. Theory Comput.*, 15(3):1996–2009, 2019.

[99] Jinshuai Song, Eric L Klein, Frank Neese, and Shengfa Ye. The mechanism of homogeneous CO_2 reduction by ni(cyclam): Product selectivity, concerted proton–electron transfer and c–o bond cleavage. *Inorg. Chem.*, 53(14):7500–7507, 2014.

[100] Christoph Riplinger and Emily A Carter. Influence of weak brønsted acids on electro-catalytic CO_2 reduction by manganese and rhenium bipyridine catalysts. *ACS Catal.*, 5(2):900–908, 2015.

[101] Charles J Stanton III, Charles W Machan, Jonathon E Vandezande, Tong Jin, George F Majetich, Henry F Schaefer III, Clifford P Kubiak, Gonghu Li, and Jay Agarwal. Re(i) nhc complexes for electrocatalytic conversion of CO_2. *Inorg. Chem.*, 55(6):3136–3144, 2016.

[102] Chern-Hooi Lim, Aaron M Holder, James T Hynes, and Charles B Musgrave. Roles of the lewis acid and base in the chemical reduction of CO_2 catalyzed by frustrated lewis pairs. *Inorg. Chem.*, 52(17):10062–10066, 2013.

[103] David W Shaffer, Samantha I Johnson, Arnold L Rheingold, Joseph W Ziller, William A Goddard III, Robert J Nielsen, and Jenny Y Yang. Reactivity of a se-ries of isostructural cobalt pincer complexes with CO_2, co, and H^+. *Inorg. Chem.*, 53(24):13031–13041, 2014.

[104] Martina Lessio and Emily A Carter. What is the role of pyridinium in pyridine-catalyzed CO_2 reduction on p-gap photocathodes? *J. Am. Chem. Soc.*, 137(41):13248–13251, 2015.

[105] Jason D Goodpaster, Alexis T Bell, and Martin Head-Gordon. Identification of pos-sible pathways for c–c bond formation during electrochemical reduction of CO_2: New theoretical insights from an improved electrochemical model. *J. Phys. Chem. Lett.*, 7(8):1471–1477, 2016.

[106] Alejandro J Garza, Alexis T Bell, and Martin Head-Gordon. Mechanism of CO_2 reduction at copper surfaces: pathways to C_2 products. *ACS Catal.*, 8(2):1490–1499, 2018.

[107] Mu-Jeng Cheng, Ezra L Clark, Hieu H Pham, Alexis T Bell, and Martin Head-Gordon. Quantum mechanical screening of single-atom bimetallic alloys for the selective reduc-tion of CO_2 to C_1 hydrocarbons. *ACS Catal.*, 6(11):7769–7777, 2016.

[108] Mu-Jeng Cheng, Youngkook Kwon, Martin Head-Gordon, and Alexis T Bell. Tailoring metal-porphyrin-like active sites on graphene to improve the efficiency and selectivity of electrochemical CO_2 reduction. *J. Phys. Chem. C*, 119(37):21345–21352, 2015.

[109] Timo Jacob. The mechanism of forming H_2O from H_2 and O_2 over a pt catalyst via direct oxygen reduction. *Fuel Cells*, 6(3-4):159–181, 2006.

[110] Corinne M Gray, Karthikeyan Saravanan, Guofeng Wang, and John A Keith. Quantifying solvation energies at solid/liquid interfaces using continuum solvation methods. *Mol. Sim.*, 43(5-6):420–427, 2017.

[111] Florian Libisch, Chen Huang, and Emily A Carter. Embedded correlated wavefunction schemes: theory and applications. *Acc. Chem. Res.*, 47(9):2768–2775, 2014.

[112] Frederick R Manby, Martina Stella, Jason D Goodpaster, and Thomas F Miller III. A simple, exact density-functional-theory embedding scheme. *J. Chem. Theory Comput.*, 8(8):2564–2568, 2012.

[113] Xinzheng Yang. Hydrogenation of carbon dioxide catalyzed by pnp pincer iridium, iron, and cobalt complexes: a computational design of base metal catalysts. *ACS Catal.*, 1(8):849–854, 2011.

[114] A Bondi. van der waals volumes and radii. *J. Phys. Chem.*, 68(3):441–451, 1964.

[115] Junming Ho and Michelle L Coote. A universal approach for continuum solvent pK_a calculations: are we there yet? *Theor. Chem. Acc.*, 125(1-2):3, 2010.

[116] Klaas Jan P Schouten, Zisheng Qin, Elena Pérez Gallent, and Marc TM Koper. Two pathways for the formation of ethylene in co reduction on single-crystal copper electrodes. *J. Am. Chem. Soc.*, 134(24):9864–9867, 2012.

[117] KJP Schouten, Y Kwon, CJM Van der Ham, Z Qin, and MTM Koper. A new mechanism for the selectivity to C_1 and C_2 species in the electrochemical reduction of carbon dioxide on copper electrodes. *Chem. Sci.*, 2(10):1902–1909, 2011.

[118] Victor Rosca, Matteo Duca, Matheus T de Groot, and Marc TM Koper. Nitrogen cycle electrocatalysis. *Chem. Rev.*, 109(6):2209–2244, 2009.

[119] Marc TM Koper. Thermodynamic theory of multi-electron transfer reactions: Implications for electrocatalysis. *J. Electroanal. Chem.*, 660(2):254–260, 2011.

[120] Ruud Kortlever, Jing Shen, Klaas Jan P Schouten, Federico Calle-Vallejo, and Marc TM Koper. Catalysts and reaction pathways for the electrochemical reduction of carbon dioxide. *J. Phys. Chem. Lett.*, 6(20):4073–4082, 2015.

[121] Jing Shen, Manuel J Kolb, Adrien J Göttle, and Marc TM Koper. Dft study on the mechanism of the electrochemical reduction of CO_2 catalyzed by cobalt porphyrins. *J. Phys. Chem. C*, 120(29):15714–15721, 2016.

[122] Alejandro J Garza, Srimanta Pakhira, Alexis T Bell, Jose L Mendoza-Cortes, and Martin Head-Gordon. Reaction mechanism of the selective reduction of CO_2 to CO by a tetraaza $[Co^{II}N_4H]^{2+}$ complex in the presence of protons. *Phys. Chem. Chem. Phys.*, 20(37):24058–24064, 2018.

[123] Andreas Klamt, Frank Eckert, and Wolfgang Arlt. Cosmo-rs: an alternative to simulation for calculating thermodynamic properties of liquid mixtures. *Annu. Rev. Chem. Biomol. Eng.*, 1:101–122, 2010.

[124] C Kalidas, Glenn Hefter, and Yizhak Marcus. Gibbs energies of transfer of cations from water to mixed aqueous organic solvents. *Chem. Rev.*, 100(3):819–852, 2000.

[125] Yizhak Marcus. Gibbs energies of transfer of anions from water to mixed aqueous organic solvents. *Chem. Rev.*, 107(9):3880–3897, 2007.

[126] Linlin Cao, Chuanzhi Sun, Nan Sun, Lin Meng, and Dezhan Chen. Theoretical mechanism studies on the electrocatalytic reduction of CO_2 to formate by water-stable iridium dihydride pincer complex. *Dalton Trans.*, 42(16):5755–5763, 2013.

[127] Josefredo R Pliego and José M Riveros. The cluster- continuum model for the calculation of the solvation free energy of ionic species. *The Journal of Physical Chemistry A*, 105(30):7241–7247, 2001.

[128] Vyacheslav S Bryantsev, Mamadou S Diallo, and William A Goddard Iii. Calculation of solvation free energies of charged solutes using mixed cluster/continuum models. *The Journal of Physical Chemistry B*, 112(32):9709–9719, 2008.

[129] Eirik F da Silva, Hallvard F Svendsen, and Kenneth M Merz. Explicitly representing the solvation shell in continuum solvent calculations. *The Journal of Physical Chemistry A*, 113(22):6404–6409, 2009.

[130] Mårten SG Ahlquist. Iridium catalyzed hydrogenation of CO_2 under basic conditions—mechanistic insight from theory. *J. Mol. Catal. A*, 324(1-2):3–8, 2010.

[131] Chern-Hooi Lim, Aaron M Holder, James T Hynes, and Charles B Musgrave. Reduction of CO_2 to methanol catalyzed by a biomimetic organo-hydride produced from pyridine. *J. Am. Chem. Soc.*, 136(45):16081–16095, 2014.

[132] Chern-Hooi Lim, Aaron M Holder, and Charles B Musgrave. Mechanism of homogeneous reduction of CO_2 by pyridine: proton relay in aqueous solvent and aromatic stabilization. *J. Am. Chem. Soc.*, 135(1):142–154, 2012.

[133] Chern-Hooi Lim, Aaron M Holder, James T Hynes, and Charles B Musgrave. Dihydropteridine/pteridine as a $2\,H^+/2\,e^-$ redox mediator for the reduction of CO_2 to methanol: A computational study. *J. Phys. Chem. B*, 121(16):4158–4167, 2017.

[134] Dongmei Xiang, Donny Magana, and R Brian Dyer. Co2 reduction catalyzed by mercaptopteridine on glassy carbon. *J. Am. Chem. Soc.*, 136(40):14007–14010, 2014.

[135] Kai Rohmann, Jens Kothe, Matthias W Haenel, Ulli Englert, Markus Hölscher, and Walter Leitner. Hydrogenation of CO_2 to formic acid with a highly active ruthenium acriphos complex in dmso and dmso/water. *Angew. Chem. Int. Ed.*, 55(31):8966–8969, 2016.

[136] Shenzhen Xu and Emily A Carter. 2-pyridinide as an active catalytic intermediate for CO_2 reduction on p-gap photoelectrodes: Lifetime and selectivity. *J. Am. Chem. Soc.*, 140(28):8732–8738, 2018.

[137] Ping Li, Graeme Henkelman, John A Keith, and J Karl Johnson. Elucidation of aqueous solvent-mediated hydrogen-transfer reactions by ab initio molecular dynamics and nudged elastic-band studies of nabh4 hydrolysis. *J. Phys. Chem. C*, 118(37):21385–21399, 2014.

[138] Mitchell C Groenenboom and John A Keith. Explicitly unraveling the roles of counterions, solvent molecules, and electron correlation in solution phase reaction pathways. *J. Phys. Chem. B*, 120(41):10797–10807, 2016.

[139] Richard Car and Mark Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, 55(22):2471, 1985.

[140] Adam W Duster, Chun-Hung Wang, Christina M Garza, Danielle E Miller, and Hai Lin. Adaptive quantum/molecular mechanics: What have we learned, where are we, and where do we go from here? *WIRES: Comput. Mol. Sci.*, 7(5):e1310, 2017.

[141] Christopher N Rowley and Benoît Roux. The solvation structure of na+ and k+ in liquid water determined from high level ab initio molecular dynamics simulations. *J. Chem. Theory Comput.*, 8(10):3526–3535, 2012.

[142] Motoyuki Shiga and Marco Masia. Boundary based on exchange symmetry theory for multilevel simulations. i. basic theory. *J. Chem. Phys.*, 139(4):044120, 2013.

[143] Hideaki Takahashi, Hiroyuki Kambe, and Akihiro Morita. A simple and effective solution to the constrained qm/mm simulations. *J. Chem. Phys.*, 148(13):134119, 2018.

[144] Min Zheng and Mark P Waller. Adaptive quantum mechanics/molecular mechanics methods. *WIRES: Comput. Mol. Sci.*, 6(4):369–385, 2016.

[145] Rosa E Bulo, Bernd Ensing, Jetze Sikkema, and Lucas Visscher. Toward a practical method for adaptive qm/mm simulations. *J. Chem. Theory Comput.*, 5(9):2212–2221, 2009.

[146] Hiroshi C Watanabe and Qiang Cui. Quantitative analysis of qm/mm boundary artifacts and correction in adaptive qm/mm simulations. *J. Chem. Theory Comput.*, 2019.

[147] Maliheh Shaban Tameh, Albert K Dearden, and Chen Huang. Accuracy of density functional theory for predicting kinetics of methanol synthesis from co and co2 hydrogenation on copper. *J. Phys. Chem. C*, 122(31):17942–17953, 2018.

[148] Hai Lin and Donald G Truhlar. Qm/mm: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.*, 117(2):185, 2007.

[149] Pratyush Tiwary and Michele Parrinello. From metadynamics to dynamics. *Phys. Rev. Lett.*, 111(23):230602, 2013.

[150] Alessandro Barducci, Massimiliano Bonomi, and Michele Parrinello. Metadynamics. *WIRES: Comput. Mol. Sci.*, 1(5):826–843, 2011.

[151] Ludovico Sutto, Simone Marsili, and Francesco Luigi Gervasio. New advances in metadynamics. *WIRES: Comput. Mol. Sci.*, 2(5):771–779, 2012.

[152] Shalini Awasthi and Nisanth N Nair. Exploring high-dimensional free energy landscapes of chemical reactions. *WIRES: Comput. Mol. Sci.*, 9(3):e1398, 2019.

[153] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proc. Nat. Acad. Sci.*, 99(20):12562–12566, 2002.

[154] Alessandro Laio, Antonio Rodriguez-Fortea, Francesco Luigi Gervasio, Matteo Ceccarelli, and Michele Parrinello. Assessing the accuracy of metadynamics. *J. Phys. Chem. B*, 109(14):6714–6721, 2005.

[155] Atsushi Urakawa, Marcella Iannuzzi, Jürg Hutter, and Alfons Baiker. Towards a rational design of ruthenium CO_2 hydrogenation catalysts by ab initio metadynamics. *Chem. Eur. J.*, 13(24):6828–6840, 2007.

[156] Mireille Ghoussoub, Shwetank Yadav, Kulbir Kaur Ghuman, Geoffrey A Ozin, and Chandra Veer Singh. Metadynamics-biased ab initio molecular dynamics study of heterogeneous CO_2 reduction via surface frustrated lewis pairs. *ACS Catal.*, 6(10):7109–7117, 2016.

[157] Grégoire A Gallet, Fabio Pietrucci, and Wanda Andreoni. Bridging static and dynamical descriptions of chemical reactions: An ab initio study of CO_2 interacting with water molecules. *J. Chem. Theory Comput.*, 8(11):4029–4039, 2012.

[158] András Stirling. HCO_3 formation from CO_2 at high ph: Ab initio molecular dynamics study. *J. Phys. Chem. B*, 115(49):14683–14687, 2011.

[159] Mirza Galib and Gabriel Hanna. The role of hydrogen bonding in the decomposition of h2co3 in water: mechanistic insights from ab initio metadynamics studies of aqueous clusters. *J. Phys. Chem. B*, 118(22):5983–5993, 2014.

[160] Tao Cheng, Hai Xiao, and William A Goddard III. Nature of the active sites for co reduction on copper nanoparticles; suggestions for optimizing performance. *J. Am. Chem. Soc.*, 139(34):11642–11645, 2017.

[161] Yanwei Lum, Tao Cheng, William A Goddard III, and Joel W Ager. Electrochemical co reduction builds solvent water into oxygenate products. *J. Am. Chem. Soc.*, 140(30):9337–9340, 2018.

[162] Tao Cheng, Hai Xiao, and William A Goddard III. Reaction mechanisms for the electrochemical reduction of CO_2 to co and formate on the cu (100) surface at 298 k from quantum mechanics free energy calculations with explicit water. *J. Am. Chem. Soc.*, 138(42):13802–13805, 2016.

[163] Tao Cheng, Hai Xiao, and William A Goddard III. Full atomistic reaction mechanism with kinetics for co reduction on cu(100) from ab initio molecular dynamics free-energy calculations at 298 k. *Proc. Nat. Acad. Sci.*, 114(8):1795–1800, 2017.

[164] Glenn M Torrie and John P Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, 23(2):187–199, 1977.

[165] Johannes Kästner. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(6):932–942, 2011.

[166] Peter Virnau and Marcus Müller. Calculation of free energy through successive umbrella sampling. *J. Chem. Phys.*, 120(23):10925–10930, 2004.

[167] Shankar Kumar, John M Rosenberg, Djamal Bouzida, Robert H Swendsen, and Peter A Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *J. Comput. Chem.*, 13(8):1011–1021, 1992.

[168] Marc Souaille and Benoît Roux. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.*, 135(1):40–57, 2001.

[169] Johannes Kästner and Walter Thiel. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "umbrella integration". *J. Chem. Phys.*, 123(14):144104, 2005.

[170] Eugenia Iskrenova-Tchoukova, Andrey G Kalinichev, and R James Kirkpatrick. Metal cation complexation with natural organic matter in aqueous solutions: molecular dynamics simulations and potentials of mean force. *Langmuir*, 26(20):15909–15919, 2010.

[171] Kevin Leung, Ida MB Nielsen, Na Sai, Craig Medforth, and John A Shelnutt. Cobalt-porphyrin catalyzed electrochemical reduction of carbon dioxide in water. 2. mechanism from first principles. *J. Phys. Chem. A*, 114(37):10174–10184, 2010.

[172] M in het Panhuis, CH Patterson, and RM Lynden-Bell. A molecular dynamics study of carbon dioxide in water: diffusion, structure and thermodynamics. *Mol. Phys.*, 94(6):963–972, 1998.

[173] Kevin Leung, Ida MB Nielsen, and Ira Kurtz. Ab initio molecular dynamics study of carbon dioxide and bicarbonate hydration and the nucleophilic attack of hydroxide on CO_2. *J. Phys. Chem. B*, 111(17):4453–4459, 2007.

[174] Mitchell C Groenenboom and John A Keith. Quantum chemical analyses of BH_4^- and BH_3OH^- hydride transfers to CO_2 in aqueous solution with potentials of mean force. *ChemPhysChem*, 18(22):3148–3152, 2017.

[175] Christoph Riplinger, Barbara Sandhoefer, Andreas Hansen, and Frank Neese. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *The Journal of Chemical Physics*, 139(13):134101, 2013.

[176] Christoph Riplinger and Frank Neese. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *The Journal of Chemical Physics*, 138(3):034106, 2013.

[177] Peter Pinski, Christoph Riplinger, Edward F. Valeev, and Frank Neese. Sparse maps—a systematic infrastructure for reduced-scaling electronic structure methods. i. an efficient and simple linear scaling local mp2 method that uses an intermediate basis of pair natural orbitals. *The Journal of Chemical Physics*, 143(3):034108, 2015.

[178] Christoph Riplinger, Peter Pinski, Ute Becker, Edward F. Valeev, and Frank Neese. Sparse maps—a systematic infrastructure for reduced-scaling electronic structure methods. ii. linear scaling domain based pair natural orbital coupled cluster theory. *The Journal of Chemical Physics*, 144(2):024109, 2016.

[179] Junming Ho, Andreas Klamt, and Michelle L Coote. Comment on the correct use of continuum solvent models. *The Journal of Physical Chemistry A*, 114(51):13442–13444, 2010.

[180] Casey P Kelly, Christopher J Cramer, and Donald G Truhlar. Aqueous solvation free energies of ions and ion- water clusters based on an accurate value for the absolute aqueous solvation free energy of the proton. *The Journal of Physical Chemistry B*, 110(32):16066–16081, 2006.

[181] Jun Zhang and Michael Dolg. Abcluster: the artificial bee colony algorithm for cluster global optimization. *Physical Chemistry Chemical Physics*, 17(37):24173–24181, 2015.

[182] Kenno Vanommeslaeghe, Elizabeth Hatcher, Chayan Acharya, Sibsankar Kundu, Shijun Zhong, Jihyun Shim, Eva Darian, Olgun Guvench, P Lopes, Igor Vorobyov, et al. Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *Journal of computational chemistry*, 31(4):671–690, 2010.

[183] James JP Stewart. Optimization of parameters for semiempirical methods vi: more modifications to the nddo approximations and re-optimization of parameters. *Journal of molecular modeling*, 19(1):1–32, 2013.

[184] James JP Stewart. Mopac: a semiempirical molecular orbital program. *Journal of computer-aided molecular design*, 4(1):1–103, 1990.

[185] John P Perdew. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Physical Review B*, 33(12):8822, 1986.

[186] Axel D Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical review A*, 38(6):3098, 1988.

[187] Florian Weigend and Reinhart Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics*, 7(18):3297–3305, 2005.

[188] Frank Neese. The orca program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(1):73–78, 2012.

[189] Axel D Becke. Density-functional thermochemistry. iii. the role of exact exchange. *The Journal of chemical physics*, 98(7):5648–5652, 1993.

[190] Stefan Grimme, Stephan Ehrlich, and Lars Goerigk. Effect of the damping function in dispersion corrected density functional theory. *Journal of computational chemistry*, 32(7):1456–1465, 2011.

[191] John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical review letters*, 77(18):3865, 1996.

[192] Jay W Ponder. Tinker: Software tools for molecular design. *Washington University School of Medicine, Saint Louis, MO*, 3, 2004.

[193] Alan Grossfield. Wham: the weighted histogram analysis method. *version*, 2(9):06, 2012.

[194] Sandip De, Albert P Bartók, Gábor Csányi, and Michele Ceriotti. Comparing molecules and solids across structural and alchemical space. *Physical Chemistry Chemical Physics*, 18(20):13754–13769, 2016.

[195] Paul Zimmerman. Reliable transition state searches integrated with the growing string method. *Journal of chemical theory and computation*, 9(7):3043–3050, 2013.

[196] Paul M Zimmerman. Single-ended transition state finding with the growing string method. *Journal of computational chemistry*, 36(9):601–611, 2015.

[197] Mina Jafari and Paul M Zimmerman. Reliable and efficient reaction path and transition state finding for surface reactions with the growing string method. *Journal of computational chemistry*, 38(10):645–658, 2017.

[198] Paul M. Zimmerman. Growing string method with interpolation and optimization in internal coordinates: Method and examples. *The Journal of Chemical Physics*, 138(18):184102, 2013.

[199] Mitchell C. Groenenboom and John A. Keith. Explicitly unraveling the roles of counterions, solvent molecules, and electron correlation in solution phase reaction pathways. *The Journal of Physical Chemistry B*, 120(41):10797–10807, 2016.

[200] Rui Zhang, Weixin Lv, and Lixu Lei. Role of the oxide layer on sn electrode in electrochemical reduction of {CO2} to formate. *Applied Surface Science*, 356:24 – 29, 2015.

[201] Aaron M. Appel, John E. Bercaw, Andrew B. Bocarsly, Holger Dobbek, Daniel L. DuBois, Michel Dupuis, James G. Ferry, Etsuko Fujita, Russ Hille, Paul J. A. Kenis, Cheryl A. Kerfeld, Robert H. Morris, Charles H. F. Peden, Archie R. Portis, Stephen W. Ragsdale, Thomas B. Rauchfuss, Joost N. H. Reek, Lance C. Seefeldt, Rudolf K. Thauer, and Grover L. Waldrop. Frontiers, opportunities, and challenges in biochemical and chemical catalysis of CO2fixation. *Chemical Reviews*, 113(8):6621–6658, aug 2013.

[202] N. M. Rezayee, C. A. Huff, and M. S. Sanford. Tandem Amine and Ruthenium-Catalyzed Hydrogenation of CO_2 to Methanol. *Journal of the American Chemical Society*, 137(3):1028–1031, January 2015.

[203] J. Kim, T. A. Johnson, J. E. Miller, E. B. Stechel, and C. T. Maravelias. Fuel production from CO_2 using solar-thermal energy: system level analysis. *Energy & Environmental Science*, 5(9):8417–8429, September 2012.

[204] Jinli Qiao, Yuyu Liu, Feng Hong, and Jiujun Zhang. A review of catalysts for the electroreduction of carbon dioxide to produce low-carbon fuels. *Chemical Society Reviews*, 43(2):631–675, December 2013.

[205] D. T. Whipple and P. J. A. Kenis. Prospects of CO_2 Utilization via Direct Heterogeneous Electrochemical Reduction. *Journal of Physical Chemistry Letters*, 1(24):3451–3458, December 2010.

[206] J. Rosenthal. Progress Toward the Electrocatalytic Production of Liquid Fuels from Carbon Dioxide. *Progress in Inorganic Chemistry, Vol 59*, 59:299–338, 2014.

[207] S. C. Ma, M. Sadakiyo, R. Luo, M. Heima, M. Yamauchi, and P. J. A. Kenis. One-step electrosynthesis of ethylene and ethanol from CO_2 in an alkaline electrolyzer. *Journal of Power Sources*, 301:219–228, January 2016.

[208] B. Kumar, M. Llorente, J. Froehlich, T. Dang, A. Sathrum, and C. P. Kubiak. Photochemical and Photoelectrochemical Reduction of CO_2. *Annual Review of Physical Chemistry, Vol 63*, 63:541–+, 2012.

[209] W. B. Hou, W. H. Hung, P. Pavaskar, A. Goeppert, M. Aykol, and S. B. Cronin. Photocatalytic Conversion of CO_2 to Hydrocarbon Fuels via Plasmon-Enhanced Absorption and Metallic Interband Transitions. *Acs Catalysis*, 1(8):929–936, August 2011.

[210] S. Licht. STEP (Solar Thermal Electrochemical Photo) Generation of Energetic Molecules: A Solar Chemical Process to End Anthropogenic Global Warming. *Journal of Physical Chemistry C*, 113(36):16283–16292, September 2009.

[211] S. Licht, O. Chitayat, H. Bergmann, A. Dick, H. Ayub, and S. Ghosh. Efficient STEP (solar thermal electrochemical photo) production of hydrogen - an economic assessment. *International Journal of Hydrogen Energy*, 35(20):10867–10882, October 2010.

[212] Y. J. Zhu, B. H. Wang, X. L. Liu, H. Y. Wang, H. J. Wu, and S. Licht. STEP organic synthesis: an efficient solar, electrochemical process for the synthesis of benzoic acid. *Green Chemistry*, 16(11):4758–4766, 2014.

[213] Seunghwa Lee and Jaeyoung Lee. Electrode build-up of reducible metal composites toward achievable electrochemical conversion of carbon dioxide. *ChemSusChem*, 9(4):333–344, nov 2015.

[214] Masashi Azuma. Electrochemical reduction of carbon dioxide on various metal electrodes in low-temperature aqueous KHCO[sub 3] media. *Journal of The Electrochemical Society*, 137(6):1772, 1990.

[215] David R. Weinberg, Christopher J. Gagliardi, Jonathan F. Hull, Christine Fecenko Murphy, Caleb A. Kent, Brittany C. Westlake, Amit Paul, Daniel H. Ess, Dewey Granville McCafferty, and Thomas J. Meyer. Proton-Coupled Electron Transfer. *Chemical Reviews*, 112(7):4016–4093, July 2012.

[216] J. L. Dempsey and T. T. Eisenhart. New approaches to probe proton-coupled electron transfer reaction pathways. *Abstracts of Papers of the American Chemical Society*, 247, March 2014.

[217] S. Hammes-Schiffer. Theory of Proton-Coupled Electron Transfer in Energy Conversion Processes. *Accounts of Chemical Research*, 42(12):1881–1889, December 2009.

[218] Yasemin Basdogan and John A Keith. A paramedic treatment for modeling explicitly solvated chemical reaction mechanisms. *Chemical science*, 9(24):5341–5346, 2018.

[219] Paul M Zimmerman. Growing string method with interpolation and optimization in internal coordinates: Method and examples. *The Journal of chemical physics*, 138(18):184102, 2013.

[220] Jeng-Da Chai and Martin Head-Gordon. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Physical Chemistry Chemical Physics*, 10(44):6615–6620, 2008.

[221] Peter Pinski, Christoph Riplinger, Edward F Valeev, and Frank Neese. Sparse maps—a systematic infrastructure for reduced-scaling electronic structure methods. i. an efficient and simple linear scaling local mp2 method that uses an intermediate basis of pair natural orbitals. *The Journal of chemical physics*, 143(3):034108, 2015.

[222] Christoph Riplinger, Peter Pinski, Ute Becker, Edward F Valeev, and Frank Neese. Sparse maps—a systematic infrastructure for reduced-scaling electronic structure methods. ii. linear scaling domain based pair natural orbital coupled cluster theory. *The Journal of chemical physics*, 144(2):024109, 2016.

[223] Christoph Riplinger and Frank Neese. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *The Journal of chemical physics*, 138(3):034106, 2013.

[224] Frank Neese, Frank Wennmohs, Andreas Hansen, and Ute Becker. Efficient, approximate and parallel hartree–fock and hybrid dft calculations. a 'chain-of-spheres' algorithm for the hartree–fock exchange. *Chemical Physics*, 356(1-3):98–109, 2009.

[225] Michael W Schmidt, Kim K Baldridge, Jerry A Boatz, Steven T Elbert, Mark S Gordon, Jan H Jensen, Shiro Koseki, Nikita Matsunaga, Kiet A Nguyen, Shujun Su, et al. General atomic and molecular electronic structure system. *Journal of computational chemistry*, 14(11):1347–1363, 1993.

[226] Perla B Balbuena, Keith P Johnston, and Peter J Rossky. Molecular dynamics simulation of electrolyte solutions in ambient and supercritical water. 1. ion solvation. *The Journal of Physical Chemistry*, 100(7):2706–2715, 1996.

[227] Frank C Pickard IV, Gerhard König, Andrew C Simmonett, Yihan Shao, and Bernard R Brooks. An efficient protocol for obtaining accurate hydration free energies using quantum chemistry and reweighting from molecular dynamics simulations. *Bioorganic & medicinal chemistry*, 24(20):4988–4997, 2016.

[228] Becky L Eggimann and J Ilja Siepmann. Size effects on the solvation of anions at the aqueous liquid- vapor interface. *The Journal of Physical Chemistry C*, 112(1):210–218, 2008.

[229] Roberto Cammi and Jacopo Tomasi. Remarks on the use of the apparent surface charges (asc) methods in solvation problems: Iterative versus matrix-inversion procedures and the renormalization of the apparent charges. *Journal of Computational Chemistry*, 16(12):1449–1458, 1995.

[230] Vincenzo Barone and Maurizio Cossi. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *The Journal of Physical Chemistry A*, 102(11):1995–2001, 1998.

[231] Lawrence R. Pratt and Susan B. Rempe. Quasi-chemical theory and implicit solvent models for simulations. *AIP Conf. Proc.*, 492(1):172–201, 1999.

[232] D. Asthagiri, P.D. Dixit, S. Merchant, M.E. Paulaitis, L.R. Pratt, S.B. Rempe, and S. Varma. Ion selectivity from local configurations of ligands in solutions and ion channels. *Chem. Phys. Lett.*, 485(1):1 – 7, 2010.

[233] David M Rogers, Dian Jiao, Lawrence R Pratt, and Susan B Rempe. Structural models and molecular thermodynamics of hydration of ions and small molecules. In *Annual Reports in Computational Chemistry*, volume 8, pages 71–127. Elsevier, 2012.

[234] Susan B Rempe, Lawrence R Pratt, Gerhard Hummer, Joel D Kress, Richard L Martin, and Antonio Redondo. The hydration number of li+ in liquid water. *Journal of the American Chemical Society*, 122(5):966–967, 2000.

[235] Susan B Rempe and Lawrence R Pratt. The hydration number of na+ in liquid water. *Fluid Phase Equilibria*, 183:121–132, 2001.

[236] Susan B Rempe, D Asthagiri, and Lawrence R Pratt. Inner shell definition and absolute hydration free energy of k+(aq) on the basis of quasi-chemical theory and ab initio molecular dynamics. *Physical Chemistry Chemical Physics*, 6(8):1966–1969, 2004.

[237] Sameer Varma and Susan B Rempe. Structural transitions in ion coordination driven by changes in competition for ligand binding. *Journal of the American Chemical Society*, 130(46):15405–15419, 2008.

[238] D Sabo, D Jiao, S Varma, LR Pratt, and SB Rempe. Case study of rb$^+$(aq), quasi-chemical theory of ion hydration, and the no split occupancies rule. *Annual Reports Section C (Physical Chemistry)*, 109:266–278, 2013.

[239] Marielle Soniat, David M Rogers, and Susan B Rempe. Dispersion-and exchange-corrected density functional theory for sodium ion hydration. *Journal of chemical theory and computation*, 11(7):2958–2967, 2015.

[240] Mangesh I Chaudhari, Marielle Soniat, and Susan B Rempe. Octa-coordination and the aqueous ba^{2+} ion. *J. Phys. Chem. B*, 119(28):8746–8753, 2015.

[241] Mangesh I Chaudhari and Susan B Rempe. Strontium and barium in aqueous solution and a potassium channel binding site. *J. Chem. Phys.*, 148(22):222831, June 2018.

[242] Mangesh I Chaudhari, Lawrence R Pratt, and Susan B Rempe. Utility of chemical computations in predicting solution free energies of metal ions. *Molecular Simulation*, 44(2):110–116, 2018.

[243] Biomolecular hydration mimicry in ion permeation through membrane channels. *Acc. Chem. Res.*, 2019.

[244] D Asthagiri, Lawrence R Pratt, Michael E Paulaitis, and Susan B Rempe. Hydration structure and free energy of biomolecularly specific aqueous dications, including zn2+ and first transition row metals. *Journal of the American Chemical Society*, 126(4):1285–1289, 2004.

[245] Dian Jiao, Kevin Leung, Susan B Rempe, and Tina M Nenoff. First principles calculations of atomic nickel redox potentials and dimerization free energies: A study of metal nanoparticle growth. *J. Chem. Theory Comput.*, 7(2):485–495, 2010.

[246] Mangesh I Chaudhari, Susan B Rempe, and Lawrence R Pratt. Quasi-chemical theory of f$^-$(aq): The "no split occupancies rule" revisited. *J. Chem. Phys.*, 147(16):161728, 2017.

[247] Ajay Muralidharan, Lawrence R Pratt, Mangesh I Chaudhari, and Susan B Rempe. Quasi-chemical theory with cluster sampling from ab initio molecular dynamics: Fluoride (f–) anion hydration. *The Journal of Physical Chemistry A*, 122(51):9806–9812, 2018.

[248] A. Muralidharan, L. R. Pratt, M. I. Chaudhari, and S. B. Rempe. Quasi-Chemical Theory for Anion Hydration and Specific Ion Effects: Cl$^-$(aq) *vs.* F$^-$(aq). *Chem. Phys. Lett.*, 2019.

[249] Henry S Ashbaugh, D Asthagiri, Lawrence R Pratt, and Susan B Rempe. Hydration of krypton and consideration of clathrate models of hydrophobic effects from the perspective of quasi-chemical theory. *Biophysical chemistry*, 105(2-3):323–338, 2003.

[250] Dubravko Sabo, Sameer Varma, Marcus G Martin, and Susan B Rempe. Studies of the thermodynamic properties of hydrogen gas in bulk water. *J. Phys. Chem. B*, 112(3):867–876, 2008.

[251] Jacalyn S Clawson, Randall T Cygan, Todd M Alam, Kevin Leung, and Susan B Rempe. Ab initio study of hydrogen storage in water clathrates. *J. Comput. Theor. Nanosci.*, 7(12):2602–2606, 2010.

[252] Dian Jiao and Susan B Rempe. Co$_2$ solvation free energy using quasi-chemical theory. *J. Chem. Phys.*, 134(22):224506, 2011.

[253] Mangesh I Chaudhari, Jijeesh R Nair, Lawrence R Pratt, Fernando A Soto, Perla B Balbuena, and Susan B Rempe. Scaling atomic partial charges of carbonate solvents for lithium ion solvation and diffusion. *J. Chem. Theory Comput.*, 12(12):5709–5718, 2016.

[254] Sameer Varma and Susan B. Rempe. Tuning ion coordination architectures to enable selective partitioning. *Biophys. J.*, 93(4):1093 – 1099, 2007.

[255] Sameer Varma, Dubravko Sabo, and Susan B. Rempe. K$^+$/na$^+$ selectivity in k channels and valinomycin: Over-coordination versus cavity-size constraints. *J. Mol. Biol.*, 376(1):13 – 22, 2008.

[256] Sameer Varma, David M. Rogers, Lawrence R. Pratt, and Susan B. Rempe. Design principles for k$^+$ selectivity in membrane transport. *J. Gen. Physiol.*, 137(6):479–488, 2011.

[257] David M Rogers and Susan B Rempe. Probing the thermodynamics of competitive ion binding using minimum energy structures. *J. Phys. Chem. B*, 115(29):9116–9129, 2011.

[258] Dian Jiao and Susan B Rempe. Combined density functional theory (dft) and continuum calculations of pk_a in carbonic anhydrase. *Biochem.*, 51(30):5979–5989, 2012.

[259] Mariana Rossi, Alexandre Tkatchenko, Susan B Rempe, and Sameer Varma. Role of methyl-induced polarization in ion binding. *Proc. Natl. Acad. Sci. U.S.A.*, 110(32):12978–12983, 2013.

[260] Mark J Stevens and Susan LB Rempe. Ion-specific effects in carboxylate binding sites. *J. Phys. Chem. B*, 120(49):12519–12530, 2016.

[261] Benedetta Mennucci and José M Martínez. How to model solvation of peptides? insights from a quantum-mechanical and molecular dynamics study of n-methylacetamide. 1. geometries, infrared, and ultraviolet spectra in water. *The Journal of Physical Chemistry B*, 109(19):9818–9829, 2005.

[262] Christopher J Fennell and Ken A Dill. Physical modeling of aqueous solvation. *Journal of statistical physics*, 145(2):209–226, 2011.

[263] Lawrence R Pratt. Contact potentials of solution interfaces: phase equilibrium and interfacial electric fields. *J. Phys. Chem.*, 96(1):25–33, 1992.

[264] Kevin Leung, Susan B. Rempe, and O. Anatole von Lilienfeld. Ab initio molecular dynamics calculations of ion hydration free energies. *J. Chem. Phys.*, 130(20):204507, 2009.

[265] Johannes Lyklema. Interfacial Potentials: Measuring the Immeasurable? *Substantia*, 1(2), 2017.

[266] Carrie Conor Doyle, Yu Shi, and Thomas L Beck. The importance of the water molecular quadrupole for estimating interfacial potential shifts acting on ions near the liquid-vapor interface. *J. Phys. Chem. B*, 2019.

[267] Yizhak Marcus. Thermodynamics of solvation of ions. part 5.—gibbs free energy of hydration at 298.15 k. *Journal of the Chemical Society, Faraday Transactions*, 87(18):2995–2999, 1991.

[268] Yizhak Marcus. The thermodynamics of solvation of ions. part 4.—application of the tetraphenylarsonium tetraphenylborate (tatb) extrathermodynamic assumption to the hydration of ions and to properties of hydrated ions. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, 83(9):2985–2992, 1987.

[269] Michael D Tissandier, Kenneth A Cowen, Wan Yong Feng, Ellen Gundlach, Michael H Cohen, Alan D Earhart, James V Coe, and Thomas R Tuttle. The proton's absolute aqueous enthalpy and gibbs free energy of solvation from cluster-ion solvation data. *The Journal of Physical Chemistry A*, 102(40):7787–7794, 1998.

[270] Guillaume Lamoureux and Benoît Roux. Absolute hydration free energy scale for alkali and halide ions established from simulations with a polarizable force field. *The journal of physical chemistry B*, 110(7):3308–3322, 2006.

[271] Vyacheslav S Bryantsev, Mamadou S Diallo, and William A Goddard III. Computational study of copper (ii) complexation and hydrolysis in aqueous solutions using mixed cluster/continuum models. *The Journal of Physical Chemistry A*, 113(34):9559–9567, 2009.

[272] Wenkun Wu and John Kieffer. New hybrid method for the calculation of the solvation free energy of small molecules in aqueous solutions. *Journal of chemical theory and computation*, 15(1):371–381, 2018.

[273] Daniel D Kemp and Mark S Gordon. Theoretical study of the solvation of fluorine and chlorine anions by water. *J. Phys. Chem. A*, 109(34):7688–7699, 2005.

[274] Jun Zhang and Michael Dolg. Global optimization of clusters of rigid molecules using the artificial bee colony algorithm. *Physical Chemistry Chemical Physics*, 18(4):3003–3010, 2016.

[275] Brandon J Jaquis, Ailin Li, Nolan D Monnier, Robert G Sisk, William E Acree, and Andrew SID Lang. Using machine learning to predict enthalpy of solvation. *J. Solution Chem.*, 48(4):564–573, 2019.

[276] NS Hari Narayana Moorthy, Silvia A Martins, Sergio F Sousa, Maria J Ramos, and Pedro A Fernandes. Classification study of solvation free energies of organic molecules using machine learning techniques. *RSC Adv.*, 4(106):61624–61630, 2014.

[277] Hyuntae Lim and YounJoon Jung. Delfos: Deep learning model for prediction of solvation free energies in generic organic solvents. *Chem. Sci.*, 2019.

[278] Albert P Bartók, Risi Kondor, and Gábor Csányi. On representing chemical environments. *Physical Review B*, 87(18):184115, 2013.

[279] Chang-Guo Zhan and David A Dixon. Absolute hydration free energy of the proton from first-principles electronic structure calculations. *The Journal of Physical Chemistry A*, 105(51):11534–11540, 2001.

[280] Demian Riccardi, Hao-Bo Guo, Jerry M Parks, Baohua Gu, Liyuan Liang, and Jeremy C Smith. Cluster-continuum calculations of hydration free energies of anions and group 12 divalent cations. *Journal of chemical theory and computation*, 9(1):555–569, 2012.

[281] Josefredo R Pliego. Shells theory of solvation and the long-range born correction. *Theor. Chem. Acc.*, 128(3):275–283, 2011.

[282] Josefredo R Pliego Jr. Cluster expansion of the solvation free energy difference: Systematic improvements in the solvation of single ions. *J. Chem. Phys.*, 147(3):034104, 2017.

[283] Guilherme Ferreira de Lima, Hélio Anderson Duarte, and Josefredo R Pliego Jr. Dynamical discrete/continuum linear response shells theory of solvation: convergence test for nh4+ and oh- ions in water solution using dft and dftb methods. *J. Phys. Chem. B*, 114(48):15941–15947, 2010.

[284] T. L. Beck, M. E Paulaitis, and L. R. Pratt. *The Potential Distribution Theorem and Models of Molecular Solutions.* Cambridge University Press, New York, 2006.

[285] Benoît Roux and Haibo Yu. Assessing the accuracy of approximate treatments of ion hydration based on primitive quasichemical theory. *J. Chem. Phys.*, 132(23):06B606, 2010.

[286] Keith E Gutowski and David A Dixon. Predicting the energy of the water exchange reaction and free energy of solvation for the uranyl ion in aqueous solution. *J. Phys. Chem. A*, 110(28):8840–8856, 2006.

[287] Travis Pollard and Thomas L Beck. Quasichemical analysis of the cluster-pair approximation for the thermodynamics of proton hydration. *The Journal of chemical physics*, 140(22):224507, 2014.

[288] Travis P Pollard and Thomas L Beck. The thermodynamics of proton hydration and the electrochemical surface potential of water. *The Journal of chemical physics*, 141(18):18C512, 2014.

[289] Travis P Pollard and Thomas L Beck. Re-examining the tetraphenyl-arsonium/tetraphenyl-borate (tatb) hypothesis for single-ion solvation free energies. *The Journal of chemical physics*, 148(22):222830, 2018.

[290] Yu Shi and Thomas L Beck. Length scales and interfacial potentials in ion hydration. *The Journal of chemical physics*, 139(4):044504, 2013.

[291] Thomas L Beck. The influence of water interfacial potentials on ion hydration in bulk water and near interfaces. *Chemical Physics Letters*, 561:1–13, 2013.

[292] Edgar A Engel, Andrea Anelli, Michele Ceriotti, Chris J Pickard, and Richard J Needs. Mapping uncharted territory in ice from zeolite networks to ice structures. *Nat. Commun.*, 9(1):2173, 2018.

[293] Albert P Bartók, Sandip De, Carl Poelking, Noam Bernstein, James R Kermode, Gábor Csányi, and Michele Ceriotti. Machine learning unifies the modeling of materials and molecules. *Sci. Adv.*, 3(12):e1701816, 2017.

[294] Félix Musil, Sandip De, Jack Yang, Joshua E Campbell, Graeme M Day, and Michele Ceriotti. Machine learning for the structure–energy–property landscapes of molecular crystals. *Chem. Sci*, 9(5):1289–1300, 2018.

[295] Michele Ceriotti. Unsupervised machine learning in atomistic simulations, between predictions and understanding. *J. Chem. Phys.*, 150(15):150901, 2019.

[296] Jose LF Abascal and Carlos Vega. A general purpose model for the condensed phases of water: Tip4p/2005. *The Journal of chemical physics*, 123(23):234505, 2005.

[297] David J Wales and Matthew P Hodges. Global minima of water clusters (h2o) n, n 21, described by an empirical potential. *Chemical physics letters*, 286(1-2):65–72, 1998.

[298] Jay W Ponder, Chuanjie Wu, Pengyu Ren, Vijay S Pande, John D Chodera, Michael J Schnieders, Imran Haque, David L Mobley, Daniel S Lambrecht, Robert A DiStasio Jr, and Martin Head-Gordon. Current status of the amoeba polarizable force field. *The journal of physical chemistry B*, 114(8):2549–2564, 2010.

[299] John A Keith and Emily A Carter. Quantum chemical benchmarking, validation, and prediction of acidity constants for substituted pyridinium ions and pyridinyl radicals. *Journal of chemical theory and computation*, 8(9):3187–3206, 2012.

[300] M Arshadi, R Yamdagni, and Paul Kebarle. Hydration of the halide negative ions in the gas phase. ii. comparison of hydration energies for the alkali positive and halide negative ions. *The Journal of Physical Chemistry*, 74(7):1475–1482, 1970.

[301] I Dzidic and Paul Kebarle. Hydration of the alkali ions in the gas phase. enthalpies and entropies of reactions m+ (h2o) n-1+ h2o= m+ (h2o) n. *The Journal of Physical Chemistry*, 74(7):1466–1474, 1970.

[302] Michele Ceriotti, Gareth A Tribello, and Michele Parrinello. Simplifying the representation of complex free-energy landscapes using sketch-map. *Proceedings of the National Academy of Sciences*, 108(32):13023–13028, 2011.

[303] Michele Ceriotti, Gareth A Tribello, and Michele Parrinello. Demonstrating the transferability and the descriptive power of sketch-map. *Journal of chemical theory and computation*, 9(3):1521–1532, 2013.

[304] Sandip De, Felix Musil, Teresa Ingram, Carsten Baldauf, and Michele Ceriotti. Mapping and classifying molecules from a high-throughput structural database. *Journal of cheminformatics*, 9(1):6, 2017.