A MULTIFACETED COMPUTATIONAL INVESTIGATION OF A PROSPECTIVE UNIVERSAL CHEMICAL ROLE OF WATER IN ENZYME CATALYSIS

by

Caleb Michael-Thomas Sindic

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Chemistry

MONTANA STATE UNIVERSITY

Bozeman, Montana

July 2024

DEDICATION

While I would not be anywhere worthwhile if not for the persistent and superfluous support of my family, I would not be writing a dissertation most tersely summarized as “computational biophysical chemistry” if not specifically for the mind and manners of my grandfather, Dr. William T. Wagner. You listened patiently when I presented my nonsensical theory of everything at age 12. At age 28, I hope you will take my theory of enzymes with equal tolerance.

ACKNOWLEDGEMENTS

I would like to acknowledge research assistant support from Montana State Research Expansion Funds Grant Program for assistance during the fall of 2021. I would also like to acknowledge the invaluable resources offered by the San Diego Supercomputer Center, and the MSU computer clusters Hyalite, Mfruits, and Tempest. Several past group members laid the foundations for this work: James Vivian, PhD; Max Yates; and Pedro L. Muiño, PhD, who collaborated directly. Of course, I would like to acknowledge the guidance and effort of my advisor, Patrik Callis, PhD.

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

TABLE OF CONTENTS CONTINUED

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

LIST OF FIGURES CONTINUED

Figure Page

LIST OF FIGURES CONTINUED

Figure Page

LIST OF FIGURES CONTINUED

Figure Page

LIST OF FIGURES CONTINUED

Figure Page

ABSTRACT

The functionality of enzymes is intricately linked to their solvation environment, yet the role of water within enzyme active sites remains insufficiently understood, particularly regarding its direct chemical involvement in catalysis. This dissertation investigates the presence and significance of water molecules within enzyme active sites, combining structural and quantum mechanical analyses to elucidate their potential catalytic roles.

First, an extensive survey of 1013 enzyme crystals, resolved to high resolution (< 1.5 Å) using X-ray crystallography, revealed that 98% of these structures possess water-filled tunnels connecting the active sites to bulk water. This structural insight suggests a ubiquitous presence of water molecules intricately positioned within the catalytic cores of enzymes. Building upon these structural findings, analysis of 1179 enzyme crystal structures demonstrated that the average electric field experienced by water molecules in enzyme active sites is greater in magnitude than that of water molecules elsewhere in or near enzymes, and substantially larger than those in bulk solvent. Moreover, a classical molecular dynamics simulation of triose phosphate isomerase and its substrate in bulk water showed that the electric fields incurred by water hydrogen near the enzyme-bound substrate are significantly greater than those near the unbound substrate.

Lastly, measuring these fields in eight enzyme active sites during quantum molecular dynamics calculations showed greater variance in the magnitude of the electric field incurred along water hydroxyl bonds relative to those in bulk solvent, implicating heterogeneity of the electrostatic profiles of enzyme active sites. During these 100 fs atom centered density matric propagation simulations, water ionizing events were observed in two of the eight enzyme active sites, which is remarkable considering the slow ionization rate of pure liquid water.

The combined structural and electrostatic evidence presented in this study supports the hypothesis that water molecules within enzyme active sites contribute critically to the catalytic efficiency and specificity of enzymatic reactions. By highlighting the chemically active nature of these water molecules, this research advances our understanding of enzymatic catalysis and opens new avenues for exploring enzyme function and design.

CHAPTER ONE

INTRODUCTION

An Overview of Enzymes

Enzymes are remarkable biological molecules that serve as catalysts for a vast array of biochemical reactions essential for life. These proteins accelerate the rates of chemical reactions by lowering the activation energy required for the transformation of reactants into products. Enzymes are found in all living organisms, and they play crucial roles in numerous physiological processes.

At their core, enzymes are highly specialized macromolecules composed of amino acids arranged in precise three-dimensional structures. This structural organization allows enzymes to interact selectively with specific substrates, facilitating their conversion into products. Enzymes exhibit remarkable catalytic efficiency and specificity, often accelerating reactions by factors of millions to billions compared to uncatalyzed reactions.

Among the most critical features of enzymes is their ability to stabilize the transition states of chemical reactions, thereby reducing the energy barrier required for the conversion of reactants into products. This process occurs through interactions between the enzyme's active site and the substrate(s), leading to the formation of an enzyme-substrate (ES) complex. Within the active site, amino acid residues and cofactors play critical roles in catalysis, facilitating chemical transformations through various mechanisms such as acid-base catalysis, covalent catalysis, and metal ion catalysis.

Enzymes exhibit remarkable versatility and diversity, catalyzing a wide range of reactions including bond cleavage, bond formation, isomerization, and more. They are classified into several classes based on the types of reactions they catalyze. At the highest level of categorization, these classes are oxidoreductases, translocases, transferases, hydrolases, lyases, isomerases, and ligases.

The reactions which are catalyzed by enzymes in biological systems serve as steps in countless biological processes. Subsequently, enzymes are involved in metabolism, signaling pathways, DNA replication and repair, protein synthesis, and many other cellular functions. Enzymes also play essential roles in biotechnological applications, including industrial processes, pharmaceutical drug development, and diagnostics.

Understanding the structure, function, and regulation of enzymes is a central focus of biochemistry and molecular biology research. Advances in techniques such as X-ray crystallography (XRC), nuclear magnetic resonance (NMR) spectroscopy, and computational modeling have provided unprecedented insights into enzyme mechanisms and dynamics, paving the way for the design of novel enzymes with tailored functions and applications. Recently, many computational methods of enzyme analysis have grown in popularity, allowing immense leaps in scientific progress. Despite this, much of their function remains misunderstood.

Enzymes are remarkably effective catalysts, capable of expediting reactions by up to 17 orders of magnitude.[1] Over time, numerous theories and principles have been constructed from experimental and computational knowledge to explain this capability. Upon edification, one will find that these ideas often corroborate or compete with one another, as is frequently the case amidst efforts to explain phenomena just beyond the realm of contemporary understanding.

Energetics of Catalysis

One common scheme used to depict the process of enzyme catalysis is a free energy diagram (fig. 1).[2] This tool allows one compare and contrast the thermodynamic journeys of a reaction whether or not it is catalyzed by an enzyme. In the catalyzed reaction, the enzyme binds the substrate to form a more thermodynamically favorable ES complex. The ES complex then undergoes shift in reaction coordinate to the bound transition state complex, $ES^{\ddagger}$. $ES^{\ddagger}$ has a lower free energy than the transition state of the uncatalyzed reaction, $E+S^{\ddagger}$. The lifetime of an enzyme transition state complex can be as short as a few femtoseconds (fs).[3] This difference in transition state energy between catalyzed and uncatalyzed pathways is known as transition state stabilization (TSS) and is one of two ideas involving the reaction coordinates of an enzyme-catalyzed reaction to justify reaction rate enhancement; the other being reactant/ground state destabilization (RSD/GSD) where the free energy of the ES complex is higher than that of the unbound substrate and apoenzyme.[4]

In accordance with TSS, the energy required to reach the transition state of the reaction, the activation energy—which is formally the Gibbs free energy of activation ($\Delta G^{\ddagger}$)— is lower in the enzyme-catalyzed reaction ($\Delta G^{\ddagger}_{cat}$) than the uncatalyzed pathway. Regardless of whether TSS or RSD is used to justify catalysis, it is this discrepancy in activation energy that underlies the power of enzymes. After formation of the enzyme-bound product, EP, energy is required to release the enzyme-bound product (E+P). Notably, both pathways begin with E+S, and end with E+P the free energies of each being unchanged by the reaction pathway Thus, in a traditional sense, the thermodynamics of the reaction are unaffected by enzyme catalysis.

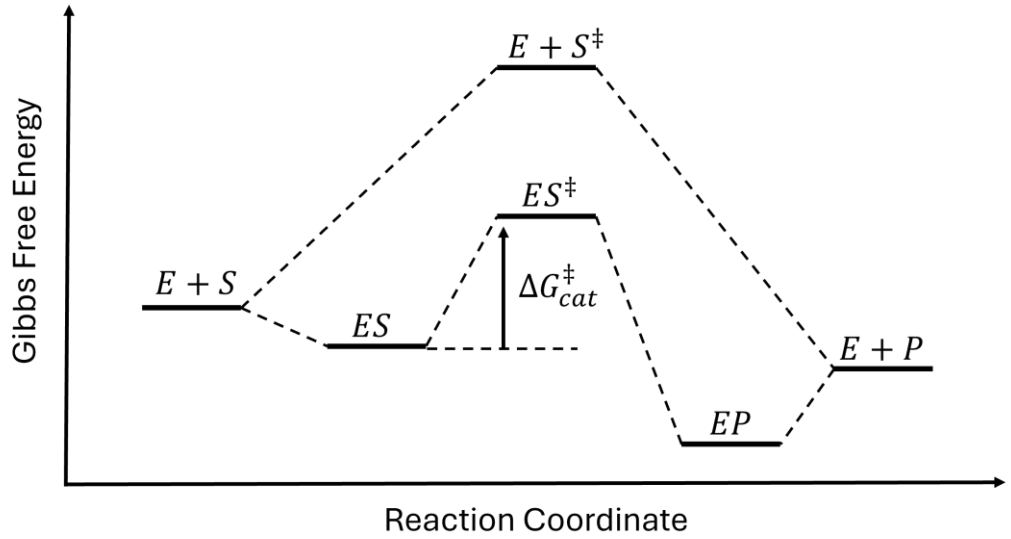

Figure 1. Free energy diagram depicting an enzyme-catalyzed reaction compared to an uncatalyzed reaction in the TSS scheme. The enzyme (E) binds the substrate (S) forming the ES complex before proceeding to the enzyme-bound transition state ($ES^{\ddagger}$), forming the enzyme-bound product (EP), and releasing it (E+P). The upper pathway depicts the uncatalyzed reaction where the reaction occurs in the same system as the enzyme, but without its chemical inclusion. The catalyzed activation energy ($\Delta G^{\ddagger}_{cat}$) can be seen but that of the uncatalyzed reaction ($\Delta G^{\ddagger}$) is omitted for simplicity.

Gibbs Free Energy

In accordance with the equation for the change in Gibbs free energy (G),

$$\Delta G = \Delta H - T\Delta S,$$

the Gibbs free energy of activation can be given by

$$\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger},$$

where ΔH is the change in the system's enthalpy, ΔS is the change in the system's entropy, T is the temperature of the system, and ‡ denotes a thermodynamic property of the transition state of the reaction. From this simple equation, one can easily see that the activation energy of an enzyme catalyzed reaction depends entirely on enthalpy, entropy, and temperature. Per the

Eyring equation, the catalyzed rate of reaction (k_{cat}) using the Gibbs free energy of activation can be calculated as

$$k_{cat} = (\frac{k_B T}{h})\exp{(\frac{-\Delta G^{\ddagger}}{RT})},$$

where k_B is the Boltzmann constant, h is the Plank constant, T is the temperature, and R is the gas constant. The thermodynamic values on which the Gibbs free energy of activation depends accordingly provide insight into the remarkable speed of enzyme catalysis.

Entropic Effects

It has long been argued that enzymes catalyze reactions by maximizing activation entropy relative to that of the reaction in aqueous solution.[5] As the entropy of the transition state complex increases relative to that of the reactant complex, the term $-T\Delta S^{\ddagger}$ becomes increasingly negative. Obviously, this entropic favorability can be achieved by either increasing the entropy of the transition state or decreasing the entropy of the reactant state. As enzymes tightly bind the substrate transition state, the former seems unlikely. Thus, the latter was proposed by William Jencks, who asserted that by binding the substrate while leaving labile the reactive regions thereof, an enzyme will decrease the entropy of the reactant state, lowering the entropic penalty of activation.[6] For some time, this notion seemed critical for elucidating enzyme catalysis. In support of this idea, it was found that by adhering reactants to a scaffold *in situ* which forced them into a presumed reactive geometry, reaction rates could be drastically increased.[7]

Enthalpic Effects

Despite its initial promise, the role of entropy in catalysis was soon shown to be more negligible than believed.[8] Citing studies measuring the temperature dependence of nominally enzyme-driven catalyzed[9] and uncatalyzed[10] reactions, Wolfenden proposed minimization of the enthalpic term to be responsible.[11] In essence, it was concluded that due to the near invariability of enzyme catalyzed reactions in response to temperature fluctuations, $-T\Delta S^{\ddagger}$ must be small relative to $\Delta H^{\ddagger}$.

But what does it mean for enthalpy to drive the activation of an enzyme catalyzed reaction? $\Delta H^{\ddagger}_{cat}$ is then much smaller than $\Delta H^{\ddagger}$ of a reaction in bulk water. As the enzyme-bound substrate is lower in free energy than the unbound substrate (fig. 1), the enthalpy of the $ES^{\ddagger}$ complex must be lower than that of $E+S^{\ddagger}$. In other words, the active site's treatment of the enzyme-bound transition state is such that the enthalpy of the state is lowered, inducing TSS. Enthalpic factors contributing to TSS include van der Waals forces, hydrogen bonding, and electrostatic interactions.[12, 13]

The Electrostatic Theory of Enzymes

The active sites of enzymes often contain more polar and ionizable amino acid residues than other locations.[14-18] One can thus generally find an increase in charge density within active sites relative to other regions within enzymes.[19] Warshel formulated a scheme of enzyme catalysis, now corroborated by these facts, in which the local electrostatic interactions within active sites drive enzyme catalysis.[20] This notion went against the emerging desolvation hypothesis: that enzymes exclude solvent from the active site to effectively reduce the local dielectric, ε and elicit a gas phase-like state favorable for catalysis.[21-23] Warshel argued that rather than simply

eliminating the solvent, the active sites of enzymes create a polar environment by means of substituting the solvent in the active site with charged and polar residues, which promotes catalysis through a variety of means pertaining to electrostatic interactions.[24]

Electrostatic Transition State Stabilization

Warshel deduced that the rate of a reaction can be accomplished by either desolvation of the ground (reactant) state of the substrate or by the solvation of the transition state[24]. Due to the lack of empirical evidence supporting ground state desolvation, and the accompanying lowered binding energy within ES, he reasoned that the transition state must be stabilized by a solvating active site environment. By analyzing the electrostatic potentials within enzyme active sites, Warshel and his colleagues demonstrated how charged residues and polar groups can orient substrates and stabilize transition states through electrostatic interactions, thereby lowering the activation energy barrier and accelerating reaction rates. [25]

Electrostatic Steering and Substrate Specificity

Warshel's ideas contributed to the concept of "electrostatic steering," whereby enzyme active sites use electrostatic interactions to bind substrates and guide the subsequent ES complex towards the transition state geometry. By strategically positioning charged or polar groups within the active site, enzymes can intake and orient substrates and facilitate the formation of transition states, leading to enhanced catalytic efficiency.[26, 27]

Strong Electric Fields in Enzyme Active Sites

A dominant role of electrostatic interactions in enzyme catalysis would bring one to postulate the presence of accompanying strong electric fields (sometimes called "electrostatic

fields") within enzyme active sites. These fields must anchor the substrate and enhance the dipoles corresponding with the transition state to steer ES to ES‡. One would then expect such fields to not only be observable within active sites, but the catalytic efficiency of the enzyme to demonstrably depend upon them.

Electric Field Magnitudes Within Enzyme Active Sites

Steven Boxer has contributed greatly to the study of electric fields within enzyme active sites, specifically their importance in the catalytic efficacy of ketosteroid isomerase (KSI). This enzyme catalyzes the isomerization of 3-oxo-Δ5-steroids by breaking the carbon-hydrogen bond adjacent to a carbonyl stabilized by a tyrosine and aspartate in the active site (fig. 2).[28] The reaction occurs eleven orders of magnitude faster when catalyzed by KSI, making KSI one of the fastest known enzymes.[29]

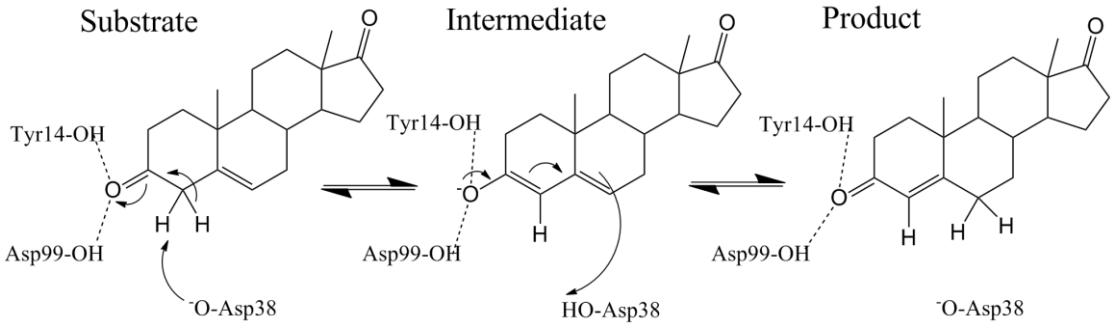

Figure 2. The mechanism of a 3-oxo-Δ5-steroid isomerization to its Δ4-conjugated form within the active site of KSI. Note the electrostatic stabilization of the intermediate by Tyr14 (Boxer's KSI has this tyrosine at position 16) and Asp99.[30]

Using Stark spectroscopy, Boxer measured the net ensemble-average electric field induced by the KSI active site along the substrate's reactive C=O bond to be 0.028 atomic units (a.u.), with the hydrogen bond from the tyrosine alone contributing 0.016 a.u.[31] This, in conjunction

with other work done by Boxer *et al.*, has proven KSI to be extremely dependent on internal electrostatics for catalysis.[32, 33]

Several other enzyme active sites have been shown to exert large electric fields on their substrates. Liver alcohol dehydrogenase (LADH), which reduces an alcohol to an aldehyde or ketone, has been shown to apply fields along the reactive C=O and C-D bonds of its deuterated inhibitor, N-cyclohexylformamide (CXF) of magnitudes 0.032 and 0.019 a.u., respectively.[34] CXF mimics the transition state of the LADH substrate; such strong fields exerted along specific bonds in this molecule coincide well with the concept of electrostatic interaction as a means of TSS.

Uracil-DNA glycosylase is an enzyme responsible for the excision of Uracil from DNA by means of breaking the N-glycosidic bond within DNA. The general bases inculpated in initiating the reaction are Asp145 and His 148.[35, 36] Using MD and QM/MM, internal electric fields of ~0.010 a.u. have been calculated between these residues, and a barrier-lowering electric field of ~0.012 a.u. to be exerted along the N-glycosidic bond of the substrate, DNA.[37]

All heme-iron oxygenases in the E state contain an Fe^{III} complex that is ligated in the ES complex, forming an $Fe^{IV}O$ intermediate. Heme-iron oxidoreductases can be split into functional subclasses by the ligation of this intermediate: P450 oxygenases are Cys ligated, catalases are Tyr ligated, and peroxidases are His ligated; each with their own unique product.[38] In survey of roughly 200 heme-iron oxidoreductases, it was found that the fields along the Fe-O bond not only drive catalysis, but also distinguish these subclasses. Specifically,

> "[…] the protein hosts exert highly specific intramolecular electric fields on the active sites, and there is a strong correlation between the direction and magnitude of this field and the protein function. In all heme proteins, the field is preferentially aligned with the Fe–O bond (**Fz**). The Cys-ligated P450 oxygenases have the

> highest average **Fz** of 28.5 MV cm^{-1} [1 MV/cm = 0.01 V/Å = $2•10^{-4}$ a.u.] i.e., most enhancing the oxyl-radical character of the oxo group, and consistent with the ability of these proteins to activate strong C–H bonds. In contrast, in Tyr-ligated proteins, the average **Fz** is only 3.0 MV cm^{-1}, apparently suppressing single-electron off-pathway oxidations, and in His-ligated proteins, **Fz** is −8.7 MV cm^{-1}. The operational field range is given by the trade-off between the low reactivity of the $Fe^{IV}O$ Compound I at the more negative **Fz**, and the low selectivity at the more positive **Fz**. Consequently, a heme-iron site placed in the field characteristic of another heme-iron protein class loses its canonical function and gains an adverse one. **Thus, electric fields produced by the protein scaffolds, together with the nature of the axial ligand, control all heme-iron chemistry."**[38]

Exploring the development of resistance to penicillin G (PenG) during the evolution of penicillin-binding proteins (PBPs) into TEM β-lactamases, Boxer and Ji found that the ester linkage in PBP–PenG is resistant to hydrolysis due to small electric fields within the active site. In contrast, the same linkage in TEM–PenG, which evolved from PBPs, is subject to larger electric fields that stabilize the charge-separated transition state, thereby lowering the free energy barrier for hydrolysis. Capitalizing on the vibrational stark effect via infrared (IR) spectroscopy, they concluded that during the evolution of TEM–PenG, electric fields on the ester linkage increased from 0.011 to 0.027 a.u., resulting in a rate acceleration by 5 orders of magnitude.[39]

Examples of Enhancing Catalysis by Modification of Electric Fields

Further demonstrating electrostatics as a critical factor in enzyme catalysis, Boxer's group increased the catalytic rate of LADH 50-fold by tailoring the electric fields within the enzyme's active site.[40] In their study, several mutations were used to modify the electric fields on the substrate (measured by proxy of the inhibitor CXF). Notably, by simply substituting the catalytic S48 residue with threonine, they measured (via Stark spectroscopy) a small 0.002 a.u. increase in the electric field magnitude on the reactive C=O bond of CXF, which led to a 6.3-fold increase in the catalytic rate of the enzyme. Conversely, swapping S48 with alanine led to a 0.005 a.u.

decrease in the electric field magnitude on the C=O bond of CXF and a massive 80-fold decrease in catalytic rate. Accordingly, the reaction rate was found to be extremely sensitive to the field exerted by the residue at position 48 of LADH on the substrate.

Another enzyme whose rate has been enhanced electrostatically is Kemp eliminase. True to its name, Kemp eliminase is responsible for Kemp elimination of the substrate (generally benzisoxazole or 5-nitrobenzisoxazole) which occurs via the abstraction of a proton by a catalytic base, leading to a partially negatively charged transition state.[41, 42] It has been calculated that outside of the enzyme, the rate of Kemp elimination is decreased with an increase in the dielectric constant of the solvent, and that an externally applied electric field along the charge transfer axis yields an increase in reaction rate.[41] A study by the Head-Gordon group presented a 43-fold increase in the catalytic rate of the enzyme Kemp eliminase via TSS attributable to a combination of electrostatic optimization and chemical positioning.[43] This was done through combination of four mutations within the enzyme: Ile168Met, Asp130Lys, Tyr167Lys, and Gly199Ala. In the subsequent mutant which demonstrated such a massive increase in catalytic rate, they measured a 0.012 a.u. increase in the electric field contribution of the catalytic base residue to the reactive C-H bond of the substrate.

In addition to showing the reliance of catalysis in these enzymes on electric fields, these enzyme studies offer great examples of the sensitivity of enzyme function to the electric fields within active sites.

Water

The strong electric fields observed within the active sites of enzyme are uniquely interesting when one considers that all enzymes are naturally solvated by the amphoteric molecule known as

water. Moreover, ~44% of side chain oxygen and nitrogen atoms in enzymes are hydrated.[44] Given that enzymes fold within liquid water after expulsion from the ribosome and exist suspended in (or at minimum exposed to) water, one must question whether the fields within enzyme active sites influence the chemical behavior of nearby water molecules. If water is indeed present in the active sites of all enzymes, could the electric fields known to drive enzyme catalysis be unspecific enough to incidentally, but routinely, ionize water? Or could the active sites of enzyme intentionally ionize water to participate in catalysis? Provided below is information regarding water to build the case that these questions are worthwhile.

The Autoionization of Water

The autoionization of water produces the highly reactive acid-base pair hydronium and hydroxide. With the equilibrium constant of the autoionization of water,

$$K_w = [H_3O^+][OH^-] = 1.0 \cdot 10^{-14},$$

hydronium and hydroxide each exist in bulk water at concentrations of $1.0 \cdot 10^{-7}$ M at 25° C. The rate constant associated with this reaction has been found to be $2.6 \cdot 10^{-5}$ s^{-1} by Eigen and Maeyer.[45] Hence, a water molecule in liquid water will ionize approximately once every 10 hours.

The reverse reaction, the recombination of hydronium and hydroxide ions, is among the fastest reactions in the world, occurring with a rate on the order of between 10^{10} and 10^{12} s^{-1}.[45-47] This is a diffusion-controlled reaction, meaning that it is limited by the movement of hydronium and hydroxide through water surrounding them.[48-50]

The Grotthuss Mechanism

But how do hydronium and hydroxide move through liquid water to recombine? In 1806, Theodor Grotthuss authored "Theory of decomposition of liquids by electrical currents", containing a prescient proposal.[51] Grotthuss, who at the time believed the chemical composition of water to be one hydrogen atom and one oxygen atom, envisioned the conductivity of water to occur due to a series of hydrogen exchange between the oxygen in each water. These water molecules are aligned single file in what is known as a water wire.[52] Per the modern incarnation of the Grotthuss mechanism, the placement of hydronium on one end of the wire and hydroxide on the other end yields rapid neutralization through the shuttling of the proton on the hydronium down the water wire towards the hydroxide until both the hydroxide and hydronium are neutralized.[46, 52-54] On account of the microscopic reversibility of the ionization of water, this mechanism occurs in reverse—starting from a neutral water wire, when two water molecules ionize one another, the hydronium and hydroxide formed will diffuse along the wire, albeit at a much slower rate than that of recombination.[53]

The mechanism of recombination along a water wire is both nuanced and well-studied by computational chemists. Natzle and Moore demonstrated that the O-O distance between hydronium and hydroxide is ~6 Å before recombination.[46] Using *ab initio* molecular dynamics, Hassalini *et al.* demonstrated that by placing hydronium and hydroxide on opposite ends of a 6 Å long water wire with two intermittent water molecules, the entire water wire contracts preceding a concerted triple-jump of the proton, yielding a neutralized hypercoordinated hydroxide.[53] The concerted proton jump, they add, occurs in ~0.5 ps. Moreover,

> "Assuming microscopic reversibility, our results provide some exciting perspectives on the mechanisms involved in the ionization of water. The dissociation of water is likely to involve a presolvation mechanism involving the formation of a hypercoordinated water molecule and a collective compression of the nearby hydrogen bonds."[53]

Simulating water wire dynamics, Kale and Herzfield corroborated the notion of bond contraction in water wires coinciding with recombination.[55] They found these "special pairs" of water molecules, with r(O-O) < ~2.5 Å, form in water wires preceding recombination of hydronium and hydroxide placed 6 Å apart. Bai and Herzfield went on to demonstrate that these requisite special pairs are much more likely to form in spaces where 1D chains of water are prevalent:

> "The results presented here suggest that the prevailing and tantalizing picture of concerted proton transfer along an extended chain of H-bonds is only relevant to 1D systems such as those found in channels within proteins and nanotubes. There it is possible to form a continuous series of ultrashort H-bonds that can lower barriers for proton hopping over relatively long distances."[54]

In other words, water wires are not just the natural means by which water ions propagate, but environments which force water wire-like structures elicit the rapid transport of protons through water!

<u>The Electrostatics of the Ionization of Water</u>

What precedes the formation of water ions in a water wire?

In a simulation of 32 periodically spaced water molecules, Geissler *et al.* demonstrated that "Dissociation begins when a fluctuation in solvent electric field causes the cleavage of an oxygen-hydrogen bond on the nascent hydroxide ion."[56] Logically, a strong electric field positioned along the vector of the OH bond of a water molecule will abstract the proton from the molecule. Saitta *et al.* calculated, via the application of various external electric fields strengths

to a box of water molecules in *ab initio* MD, the minimum necessary strength of the field to be 0.005 a.u. in order to deprotonate a water molecule.[57] At an external field of 0.007 a.u., they found they could elicit the propagation of the proton along a chain of water molecules via a Grotthuss-type mechanism. At a field strength of 0.019 a.u., they found ~15% of water molecules to exist at water ions within their simulation. Another study employed similar methods to demonstrate that in the presence of charged particles (aqueous NaCl), the effects observed by Saitta *et al*. at an external field magnitude of 0.007 a.u. can be seen with an external electric field of 0.005 a.u..[58]

Using IR and Raman spectroscopy, Cassone *et al.* discovered that the OH stretching frequency within bulk water has been shown to significantly decrease in response to externally applied electric fields ranging from 0 to 0.005 a.u..[59] This decrease in frequency correlates with a strengthening of hydrogen bonds in water networks, characteristic of an increased proclivity for proton transfer and propagation along water wires, as the hydrogen-bonded protons are generally farther from their own oxygen atoms. Their study further demonstrated the restructuration of the solvation shells around each water molecule during the application of increasing external electric fields and an increase in average molecule dipole moment magnitude from 3.15 to 3.55 D in response a mere 0.005 a.u. applied electric field.

While highly specific in their orientation, the electric fields observed in select active sites[i] (ca. 10^{-2} a.u.) are up to an order of magnitude greater than those calculated to ionize water (ca. 10^{-3} a.u.). This leads to the question of whether enzyme active sites are capable of ionizing water

[i] See "Strong Electric Fields in Enzyme Active Sites" above in this dissertation.

molecules. If so, could protons be propagated along water wires within the many channels[ii,60] known to perforate enzyme structures? Further evidence to support these ideas as ubiquitous truths is given in several studies which have characterized the role of water wires in various enzymes.

Water in Enzymes

Demonstrating that charges within enzymes may effectively regulate proton transfer along internal water wires, Hassan *et al.* simulated a water wire within an enzyme using a carbon nanotube (CNT) with an 8 Å radius and 13.5 Å length containing five water molecules.[61] On one end of the CNT, they placed a basic imidazole molecule. On the other end of the CNT, they placed three water molecules to simulate bulk solvent effects and one hydronium molecule. They then carried out molecular dynamics on copies of this system, each with a different imposed electric field normal to the primary axis of the CNT. It was found that at electric fields lower than $8.7 \cdot 10^{-3}$ a.u., the charge was shuttled back and forth between ends of the CNT, ultimately ending up on the imidazole. At field strengths greater than $9.6 \cdot 10^{-3}$ a.u., the charge became trapped in the middle of the CNT. This study invokes curiosity regarding the use of such a mechanism by enzymes to regulate proton transfer along water wires via internal electric fields generated by their residues.

Such a mechanism seems to be in place within the cytochrome c oxidase proton pump. Using FTIR, Maréchal and Rich discovered that water molecules in the active site of the enzyme reorient in response to the redox state of the binuclear iron-copper center of the enzyme.[62] In an

[ii] Pravda *et al.* found that 64% of enzymes contain two or more channels, with an average length of 28 Å, which terminate at their active sites. See reference adjacent footnote marker.

MD study, Sharma *et al*. added that the reorientation of eight active site water molecules likely facilitates the redirection of proton transfer to the binuclear cluster as needed.[63] Fascinatingly, they observed it also seems to prevent the proton transfer when it would lead to short-circuiting the proton pump through a mechanism of regulation reminiscent of that calculated in the aforementioned study by Hassan *et al*..[61]

Numerous other enzymes have been found to contain water wires which specifically direct proton transfer. For example, a water wire has been heavily characterized in the gramicidin A (GA) channel. In 1996 Pomès and Roux used discretized Feynman path integral MD[64, 65] to analyze proton transfer along the single-file chain of water molecules traversing the enzyme.[66] They approximated that proton diffusion within the GA channel is an order of magnitude faster than the diffusion of water molecules. In contrast, it has been reported that the diffusion of protons is only about 5 times faster than that of water molecules in bulk water.[67] Pomès and Roux also reported that the mechanism of proton transfer in GA was concerted and coincided with contraction of the hydrogen bonds in the water molecules along the wire in rapid response to the presence of a proton on one end of the wire, just as observed outside of the enzyme environment.[53-55, 66]

In many enzymes, water wires have been found to be part of larger “proton wires” consisting of water molecules along with nitrogen and oxygen atoms of side chain residues. In carbonic anhydrase, a crucial enzyme in the regulation of pH and CO2 transports, a zinc-bound water molecule donates a proton to a histidine residue, which then transfers it to the bulk solvent via a chain of hydrogen-bonded water molecules. This relay mechanism ensures a swift and coordinated proton transfer, maintaining the enzyme's catalytic efficiency.[68, 69]

Another example can be found in heme enzyme ascorbate peroxidase[70], where water molecules in a proton transfer pathway from the enzyme's substrate, ascorbate, to its heme-bound oxygen were found to be critical for catalyzing ascorbate's oxidation.[71] In their study, Efimov *et al*. derived that proton transfer from the reactive alcohol group on the substrate is concertedly mediated by two arginine residues and three water molecules along a proton wire within the enzyme.[71] They further noted that these two arginine residues serve as substrate binding sites, effectively binding the substrate before abstracting and transferring a proton from it along the proton wire.

In class Ia ribonucleotide reductases, a tyrosyl radical-diiron cofactor, ligated Asp84, oxidizes cysteine in the active site.[72] These enzymes are inactivated by radical scavengers such as hydroxamic acids and 4-methoxyphenol, making them a valuable antibacterial drug target.[73] Inactivation occurs when the hydroxamic acid inhibitor reduces the tyrosyl radical via proton-coupled electron transfer (PCET).[73] Using reaction-induced Fourier-transform IR spectroscopy[74, 75], Offenbacher and Barry showed that the catalytically critical tyrosine radical is protonated by hydroxamic acids through a proton wire containing four water molecules in conjunction with aspartate and glutamate residues.[73] Curiously, they found evidence suggesting that when 4-methoxyphenol is used as an inhibitor instead of a hydroxamic acid, this proton wire is bypassed during PCET.

When damaged by UVA and UVB light, cyclobutane pyrimidine dimers (CPDs) form within DNA, leading to several deleterious consequences such as skin cancer.[76, 77] In an elegant biological function, visible blue light activates DNA photolyase, which restores a given CPD to two normal base pairs. Lui *et al*. identified the means of repair to be cyclical electron-radical

transfer between the $FADH^-$ cofactor and the CPD, bridged by an adenine.[78] Adding to this discovery, Wang *et al.* conducted a series of classical and QM/MM calculations on the enzyme, finding that three water molecules bridge the $FADH^-$ and the adenine, and that another two bridge the adenine and CPD, comprising a proton wire which facilitates PCET over > 10 Å.[79]

The order of events in the narrative of photolyase research is particularly relevant to this dissertation. In this enzyme, charges were known to be transferred over a long distance and water was known to be present in the space of charge transfer. Given this knowledge, only when explicitly tested was it found that the water directly participated in catalysis. In general, if water is conductive fills the space between two groups which exchange charges, why would the water molecules themselves not participate? For how many more enzymes will such a narrative unfold when subjected to sufficient research?

Showcasing the first ever *ab initio* Born-Oppenheimer MD[80] simulation of an entire protein, Ufimtsev *et al*. demonstrated water wires as a means of diffusing charges from an enzyme into the bulk solvent.[81] When computationally removing a structure of bovine pancreatic trypsin inhibitor protein (BPTI) from solvent—thus emulating a gas phase-like state, they found via electron density mapping a net intraprotein charge transfer of 1 electron which moves from neutral residues to polar residues. This "polarization stress" was found to be relieved when BPTI was again solvated by water; during their 8.8 picosecond (ps) simulation of BPTI after re-solvation, they calculated that 2-3.5 electrons are transferred from bulk solvent to the BPTI. Their letter concludes:

> "Water mediates protein function in many ways. It serves as a buffer for chemical reactions in enzymes, initiates hydrophobic collapse in protein folding, and actively participates in protein recognition. The results presented in this Letter demonstrate another potentially important role of biological water. Water molecules located near

> the protein surface notably affect the protein's electronic structure and stimulate release of polarization stress. This effect occurs in vacuo and leads to significant intraprotein CT. […] Polarization stress could potentially lead to adverse consequences disrupting protein function. Upon solvation, however, BPTI restores its intraprotein charge balance by borrowing two to three electrons from the surrounding water bath, and the neutral residues acquire the expected, that is, neutral, charge. […] Our results reinforce previous findings obtained with more approximate methods for other proteins, [82-85] and one can anticipate that CT is an intrinsic characteristic of large biomolecules, essential for their function."[81]

Another example of charge transfer between a protein and bulk solvent by way of water can be found green fluorescent protein (GFP). GFP presents complex chemical behavior to elicit biofluorescence through the complexation of residues serine 65, glycine 66 and tyrosine 67, which form the chromophore p-hydroxybenzylideneimidazolinone in the presence of oxygen.[86] When irradiated, GFP exhibits excited state proton transfer between the chromophore and glutamate 222. This is mediated by the amusingly titled "privileged water molecule"—so named because of its criticality in the charge transfer process.[87] A second proton wire containing a water molecule has also been uncovered in GFP and is believed to play a role in chromophore biosynthesis.[88] In what is perhaps the most fascinating study cited in this dissertation, Shinobu *et al*. argue for the presence of a "proton antenna" in GFP, they state:

> "'Proton-collecting antenna' are conjectured to consist of several carboxylates within hydrogen-bond (HB) networks on the surface of proteins, which funnel protons to the orifice of an internal proton wire leading to the protein's active site."[89]

Applying a computational algorithm to a 0.9 Å resolution structure of GFP, Shinobu *et al*. uncovered the connection of the Glu222 proton wire with the surface residue Glu5. Moreover, they uncovered a region of negatively charged surface residues connected via a proton wire to Glu5, which itself is connected to Glu222 via a proton wire. This region is argued by the authors

to draw protons from the bulk solvent to the active site of GFP, which are then shuttled out of GFP through a separate hydrophobic region of the protein.

Proposal

Thus far, it has been delineated that enzymes are inherently solvated in water, the fields measured or calculated within enzyme active sites are of magnitudes demonstrated to influence the ionization of water, the propagation of protons along water wires have been reported in select enzymes, and the water wires have been shown to play critical catalytic roles in numerous enzymes. With these facts in mind, could it be that water, the "universal solvent," universally serves as a *chemically requisite* species in enzyme catalysis? This dissertation details the first ever investigation of enzymes operating under the hypothesis that water is a chemical actor in all enzymes.

Possible Mechanisms of Water-Assisted Catalysis

Being such a bold and general suggestion, our hypothesis must be accompanied by a delineation of the possible ways by which water could play a universal chemical role in enzyme catalysis.

Direct Protonation/Deprotonation

Water can directly participate in the catalytic process by acting as a proton donor or acceptor. Of course, these roles are only possible in enzymes where the substrate is protonated or deprotonated at some step of the catalyzed reaction.

A water molecule may donate a proton to the substrate, facilitating a chemical transformation such as hydrolysis. An acidic residue nominally attributed to the protonation of the substrate

may, if separated from the substrate by intervening water, instead protonate the water. The transient and highly acidic hydronium in closer proximity to the substrate would then favorably protonate the substrate instead of the acidic amino acid residue located further away.

Conversely, water can also accept a proton from the substrate, aiding in reactions that require deprotonation. This mechanism is particularly relevant in enzymes where a residue serving as the catalytic base is separated from the substrate by one or more water molecules. Rather than direct deprotonation of the substrate over a long distance by the catalytic base, the basic residue may deprotonate a water molecule. The extremely nucleophilic hydroxide formed may then carry out deprotonation.

Mediated Proton Transfer

If not directly protonating/deprotonating the substrate, water may serve to induce or relieve electrostatic strain within active sites during their catalytic cycles. Water molecules may facilitate the translocation of charges relay mechanism, where a proton is transferred from one molecule to another via one or more intervening water molecules, as discussed above. This proton shuttling may help in the stabilization of transition states and intermediates, thereby lowering the activation energy of the reaction. An extension of this mechanism may apply to electron transfer or PCET, as in DNA Photolyase.[79]

Hydration Shell Stabilization

The hydration shell around the enzyme and the substrate can influence the reaction dynamics. Water molecules can stabilize charged intermediates and transition states through hydrogen bonding and electrostatic interactions. This stabilization can make certain reaction

pathways more favorable, thereby increasing the reaction rate. Notable examples of hydrogen bonds contributing to TSS have been elucidated or reviewed in depth by Perry Frey.[13, 90]

Rationale of this Work

For any combination of these means of water-assisted enzyme catalysis to be universally true, one would find that all enzyme active sites are solvated at some point during their catalytic cycle. Furthermore, one would expect to find that the channels which perforate enzymes—terminating at both their active sites and bulk solvent— would be occupied by clusters of water molecules as an indication of replenishment during catalysis. If these water clusters serve a supplemental role as water wires,[iii] it would be expected that their intermolecular oxygen-oxygen distances are consistent with those found in water wires.

This hypothesis can be tested beyond simply surveying enzyme structures for the locations of water molecules. If water is truly a universal cofactor in enzymes, then the strong fields reportedly experienced by enzyme substrates to be incurred by water molecules as well. Accordingly, the fields experienced by water molecules in the active sites of enzymes would be stronger than those in bulk solvent. Moreover, if water participates in catalysis in the form of hydroxide, it would be true that hydroxide would be favorably formed in enzyme active sites relative to within bulk solvent. Thus, the electric fields incurred by water would be particularly strong as projected along the OH bond in a manner favoring proton abstraction.

The approach taken here is purely computational. To discern whether all enzyme active sites contain water and whether clusters of water connect active sites with bulk solvent, 1013 enzyme

[iii] The differentiating factor between a water cluster and water wire is that a water wire definitively facilitates proton transfer along its length.

structures resolved by XRC were downloaded from the RCSB. In each, water molecules near the active site were identified. Clusters of water molecules containing these molecules were then identified, and their contiguity with the bulk solvent surrounding the enzyme was assessed.

To determine whether enzyme active sites electrostatically influence the ionization or "activation" of water, 1179 enzyme structures from the RSCB PDB were assessed. In each, water molecules near the active site were identified, along with water molecules elsewhere in the enzyme and in bulk solvent. The electric fields experienced by water molecules in each of these partitions of water were collected and compared. To compare the electric fields incurred by water hydrogen between these partitions of water molecules during molecular motion, an experimentally resolved structure of triose phosphate isomerase was subjected to 2000 ns classical molecular dynamics and the electric fields on each water hydrogen were gathered every 10 ns simulation time, then compared on the basis of their proximity to the enzyme's substrate. To measure the proclivity of water molecules to form hydroxide within enzyme active sites during natural molecular motion simulated at a high level of accuracy, eight enzyme active sites were subjected to 100 fs *ab initio* quantum molecular dynamics simulations and the electric fields along the OH bond of each were gathered at 0.1 fs intervals. These fields were then compared to those in a box containing 96 water molecules during a simulation with the same parameters used for the active sites.

CHAPTER TWO

METHODS AND METHODOLOGY

Program Access

The codes used for the collection and analysis of data as described in this chapter can accessed and used via GitHub at https://github.com/cmsindic/Water_Wire_Connectivity and https://github.com/cmsindic/Electrostatic_Mapping.

Water Topology

Summary of Methods

If water is universally involved in the catalysis of all enzymes, it must be present in the active sites of all enzymes. It would be found that no enzyme exists without water molecules near its active site during catalytically relevant conformations of that enzyme. If a Grotthuss-type mechanism necessary for catalysis exists in enzymes, one would further find, contiguous with enzyme active sites, clusters of water molecules within proton transfer distance of one another such that disparate charges could be transferred by this mechanism. Furthermore, if water behaves as a cofactor within enzymes, it would be expected that these clusters trail from the active sites of enzymes to the bulk solvent that surrounds them so that water may be replenished for catalysis.

In order to investigate a hypothesis with such generality and breadth, a computational approach that can assess a large number of enzymes is appropriate. Luckily, tens of thousands of enzyme crystal structures exist freely on the internet, courtesy of the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB). Assisted by a series of in-house and

open-source computer programs, it was practical to explore the topological characteristics of water molecules inside many of these enzyme crystal structures so as to draw inferences about their natural states.

Enzyme structures were obtained in the form of Protein Data Bank (PDB) files from rcsb.org. The selection of PDB files were then filtered down to a smaller sample size on the basis of structural resolution, redundancy, and the method by which their corresponding crystals were resolved. For each of the selected structures, the locations of the structure's active site residues were gathered from uniprot.org. The files were solvated using PACKMOL.[91] During programmatic water cluster analysis, water molecules near active site residues were identified. Importantly, the number of enzyme structures with one or more active site-proximal water molecules were counted. Clusters of water containing active site-proximal water molecules were then characterized and quantified. The number of enzyme structures in which these clusters bridged the active site and bulk solvent were then counted. Figure 3 provides a visual overview of these steps.

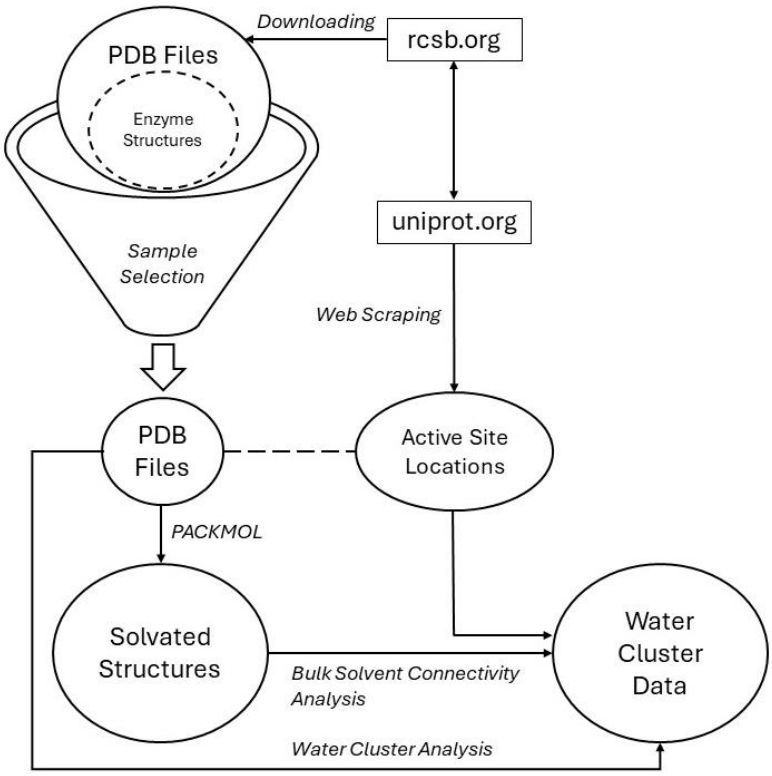

Figure 3. Conceptual diagram depicting the process of obtaining, selecting, modifying, and analyzing water clusters within enzyme structures encoded as PDB files.

Selection of PDB Structures

Initially, ~60,000 enzyme structures were obtained from rcsb.org in the form of PDB files. In order to refine the selection of structures used for water topology analysis, these files were filtered through several criteria. Only structures resolved using X-ray crystallography (XRC) were used. This was done to minimize confounding variables in the subsequent results of data collection, lest one enzyme may have larger water clusters than another due to the differences in the methods by which the two structures were resolved. To assure usage of structures whose crystals were of high precision, all PDBs with crystal resolution > 1.5 Å were discarded.

Many duplicate structures existed which portrayed the same enzyme, enzyme homolog, or mutant. A python program was written and employed to remove duplicate structures. This code

eliminated all but one of each group of structures sharing the same name, e.g., if two PDB files shared compound title THIOREDOXIN REDUCTASE, one was discarded. However, a PDB file with compound title THIOREDOXIN REDUCTASE TRXB-3 was not considered a duplicate of THIOREDOXIN REDUCTASE due to the TRXB-3 tag. Such small differences in compound names were found to significantly dichotomize enzyme structures. In this instance, TRXB-3 denotes *Ss*TrxRB3, a variant of thioredoxin reductase found in thermophile *sulfolobus solfataricus* which can be differentiated from even other thioredoxin reductases found in the same organism (e.g., *Ss*TrxRB2) and shows significant structural differences compared to other thioredoxin reductases from other species.[92, 93]

The PDB files were further selected on the basis of available data on their active sites' locations and their compatibility with the chosen computational method of adding water molecules to the structures. These steps are discussed in detail below.

Gathering Active Site Locations

To identify water molecules near active site atoms in the enzyme structures, we first needed to identify the enzymes' active site residues and associate them with cartesian coordinates in the enzymes' structure files. Due to the large sample of enzymes assessed, doing this manually was unproductive. Thus, a web scraping code segment was written in Python to identify active site residues in each file using data sourced online.[iv]

[iv] This code was originally separate from the primary spatial analysis program (analyze_wires.py) but was later incorporated into this file so that active site web scraping and water wire analysis were conducted directly in sequence.

Every PDB file has a corresponding page on rcsb.org. It was observed that nearly every one of these RCSB pages contains a link to a uniprot.org page for the enzyme represented by that PDB, and that the text (.txt) format of the UniProt page contains the positions of active site residues in a highly legible format. Hence, all that was required of the web scraping code was for it to navigate to these UniProt webpages for each of the PDB files used for analysis and extract the active site positions (indices) therefrom.

To accomplish this, the program first navigated to the rcsb.org page of a given PDB file on the basis of the file's name. For any xxxx.pdb, where xxxx is a PDB ID, the corresponding webpage is https://www.rcsb.org/structure/xxxx. (For example, the rcsb.org page for file 1NEY.pdb is https://www.rcsb.org/structure/1ney.) The program then read the HTML underlying a given RCSB PDB webpage using the *lxml* and *requests* Python packages.[94-96] Links containing the string "uniprot.org" within the RCSB HTML were then opened. If a given RCSB page did not have exactly one link to UniProt, the PDB file was removed from the sample. In addition to the obvious need for active site information, this was because having more than one UniProt link would mean time-consuming comparison of active site positions between the UniProt webpages and literature to assess which page contained the correct information. Within the UniProt .txt pages, digits following the terms ACT_SITE and BINDING in lines containing said terms indicate the positions of active site and binding site residues (fig. 4). For example, an active site residue such as Tyr14 would appear as "ACT_SITE 14" in a UniProt .txt page. Accordingly, these indices were gathered and associated with residues in each of the PDB files.

It was observed that residue entries in UniProt contained both the positions of active site and binding site residues. Many of the PDB files only had binding site residues listed on UniProt.

Initially, this elicited concern that allosteric binding sites may be misinterpreted as active sites by the programs written for this study. However, manual verification established that these binding site residues were bound to the substrate or ligand within the active site pockets of the PDB files used for this study. Because of this, it was deemed that binding site residues, as listed on UniProt, were sufficient to approximate active site locations in the enzyme structures. Thus, both BINDING and ACT_SITE residues, as labeled on the aforementioned uniprot.org .txt pages were included to identify the active site locations within each structure. In this document, they are referenced synonymously as “active site residues” unless explicitly stated otherwise.

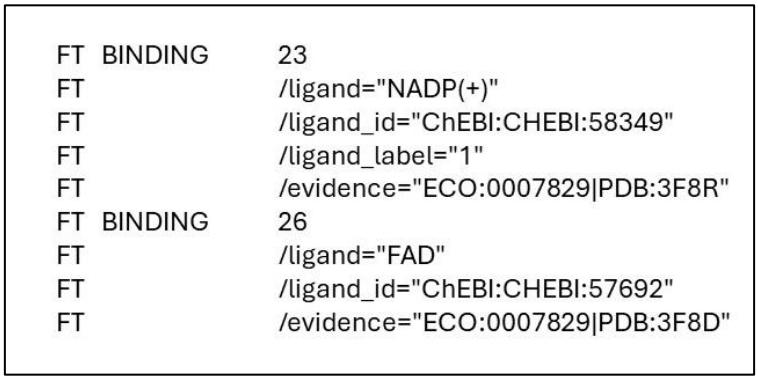

Figure 4. An excerpt from a UniProt .txt webpage for thioredoxin reductase showing that residues 23 and 26 bind NADP(+) and FAD, respectively.

Water Clustering Algorithm

A program, *analyze_wires.py*, was written to conduct the bulk of spatial analysis of water molecules in PDB files gathered from the RCSB PDB. In addition to hosting the web scraping process described in the previous section of this paper, this program served to read PDB files, interpret the locations of atoms within them; and gather information on active site-proximal water molecules, active site-contiguous clusters of water molecules, and such clusters of water molecules also contiguous with bulk solvent. The details of this program's functions are detailed in the remainder of this section.

PDB files were parsed using the *PDBParser* method available in the BioPython[97] package for python. This tool allows for the representation of the data contained in the PDB files as residues, molecules, and atoms. Importantly, the coordinates of a given atom in a given PDB file can be obtained by calling upon the coordinate property of the atom within a python code. Moreover, the positions and names of residues can be obtained in a simple manner using *PDBParser*. These features allowed locating the active site residues and the coordinates of active site atoms in each enzyme structure.

In each enzyme structure, water molecules with oxygen atoms < 5 Å from active site atoms identified and counted. A similar method and classification distance has been used for identifying molecules relevant to active site catalysis in at least one other study.[43] While the number of enzyme structures having any water molecules near active site residues was in itself a valuable metric, *analyze_wires.py* also tested whether these water molecules also formed the initial segment of a cluster of water contiguous with the active site. This was done by first identifying water molecules within 5 Å oxygen-oxygen distance (r(O-O)) of water molecules near the active site. Iteratively, water molecules with r(O-O) < 5 Å from any molecules in the growing cluster of

water were added to the cluster. This was repeated until the cluster of water molecules could no longer be extended.

Solvating Enzyme Structures

As a means of measuring solvent accessibility of enzyme active sites, we sought to determine how many structures had water wires connected to the bulk solvent outside the enzyme. It was noted, however, that most enzyme structures lacked water outside of the boundaries of the enzymes themselves. This external (bulk) solvent was needed for analyzing whether clusters of water reached bulk solvent because of the extreme difficulty of computationally identifying the space outside of an enzyme structure. By adding solvent around each structure, the space outside of the enzymes could be obtained by proxy of the cartesian coordinates of the bulk solvent molecules. Thus, open-source model solvating software PACKMOL[91] was employed to simulate solvation. This program adds water molecules (or any other solvent of choice) in a selected density around existing atoms in a molecular model, such as a PDB file. Accordingly, the end product of solvating a given PDB file enzyme structure was an enzyme structure with naturally occurring water and “artificial” water added by PACKMOL, primarily outside of the bounds of the enzyme structure (fig. 5).

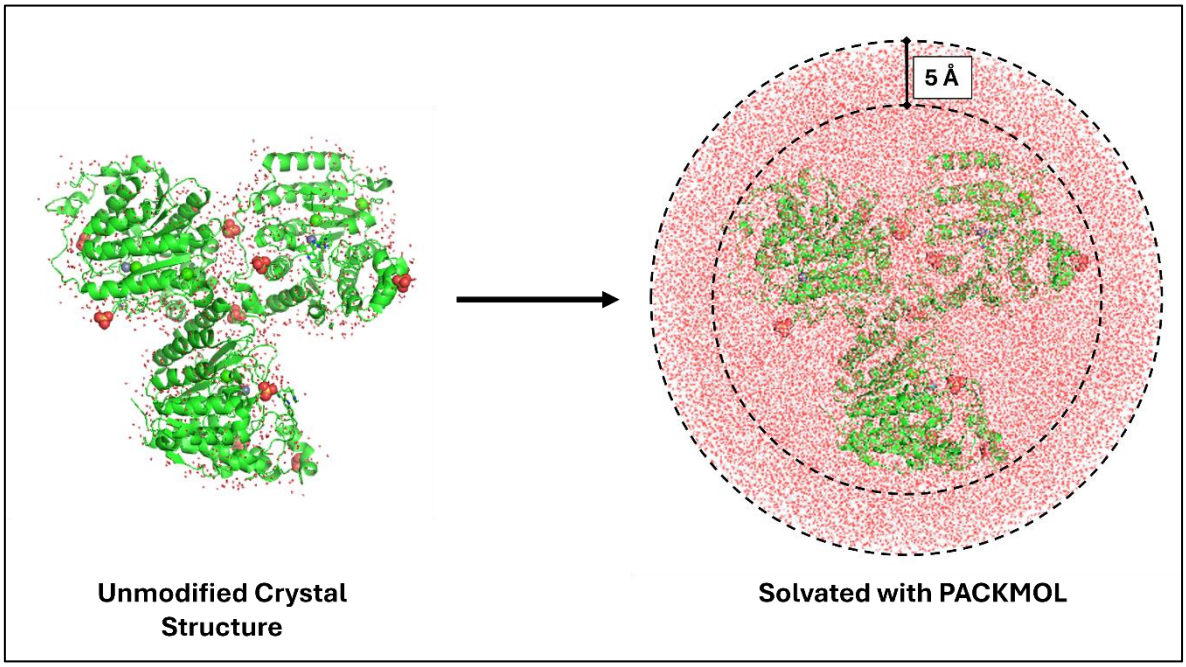

Figure 5. A crystal structure of an enzyme before (left) and after (right) the addition of hydrogen to water oxygen atoms using PyMOL and solvation with water using PACKMOL. The enzyme can be seen in green cartoon representation with water molecules shown as small red points. Other non-enzyme atoms are shown in VDW sphere format. The 5 Å solvent sphere padding from the most distant crystal atom from the crystal's geometric center is explicitly shown. Note the asymmetric distance between the enzyme surface and the boundary of the solvation sphere on account of the enzyme's asymmetry.

A code, *gencsv.py*, was written to gather the data needed to make input (.inp) files which specify solvation conditions for PACKMOL. First, this code generated a spherical boundary with a 5 Å padding from the most distant enzyme atom from the center[v] of each enzyme structure. Using VDW radii of each atom in the structures, sourced via the chemlib[98] Python package, *gencsv.py* calculated the volume occupied by the original structure, including extant water molecules. To find the volume available to the simulated bulk solvent, the structure's volume[vi] was subtracted from the boundary sphere's volume. The resulting volume, in Å^3, was multiplied

[v] Enzyme centers were defined as the 3-dimensional mean of all atomic cartesian coordinates in each enzyme structure.
[vi] The sum of VDW volumes of each atom in the model.

by the density of water at 310 K and 1 atm (0.99 g/cm^3 ≈ 0.03 molecules/Å^3) to determine the number of water molecules that could reasonably fill the space inside of the boundary sphere. For each enzyme structure, these data were stored as a comma-separated value (.csv) file (table 1) for review before being converted to a .inp file for input into PACKMOL.

Directory	.\ligases
Enzyme Volume	2618.02
Total Volume	102160.4
Water Volume	99542.38
Number of Waters	2986
Radius	29

Table 1. An example of data for a specific enzyme structure (1GXT.pdb), stored to a .csv file by *gencsv.py* for conversion into a .inp file for PACKMOL. Volumes and the model's radius are given in units of Å^3 and Å, respectively. Number of Waters refers to the number of water molecules to be added by PACKMOL.

A code, *mkinpt.py*, was then written to convert the data contained in each .csv file made for each enzyme structure (table 1) into a PACKMOL-compatible .inp file (fig. 6).

```
#
# Protein solvated by water and ions
#

tolerance 2.0

structure ../.\ligases\1GXT.pdb
 number 1
 fixed 0. 0. 0. 0. 0. 0.
 centerofmass
end structure

structure water.pdb
 number 2986
 inside sphere 0. 0. 0. 29.
end structure

output .\ligases\1GXT_solvated.pdb
```

Figure 6. An example of a .inp file serving as input for the solvation of enzyme structure 1GXT.pdb. Here, it is specified that the center of mass of the original enzyme structure shall be positioned at origin of cartesian coordinates in the output file, and that a sphere containing the enzyme with radius of 29 Å shall be filled with 2986 water molecules.

Using the input files made by *mkinpt.py*, PACKMOL was employed to solvate each enzyme structure using the TIP3P[99] water model. This water model has a 0.9572 Å O-H bond distance and a 104.52 H-O-H bond angle.

The program previously used for water cluster identification within each of the unsolvated enzyme structures, *analyze_wires.py*, was then applied to the solvated enzyme structures in order to determine whether the active site-contiguous water clusters came in contact with bulk solvent around each enzyme. Within *analyze_wires.py,* a new partition of water was defined: bulk solvent. This partition was classified as containing water with distance r(O-X) > 5 Å, where O is a given water oxygen atom and X is any enzyme atom. This ensured the water in this partition adequately approximated the space occupied by bulk solvent outside of each enzyme. We then

counted the structures with water clusters in contact with bulk solvent. Contact with bulk solvent was defined as having a water cluster containing at least one water molecule with r(O-O) < 5 Å from any bulk solvent water molecule. As a loose measurement of the number of avenues a water molecule can take to access the active sites of enzyme, the number of bulk solvent water molecules contacted by water molecules within each cluster were recorded for each structure. For a given enzyme structure, a high number of bulk solvent water molecules contacted by components of its cluster would indicate a large interface between water in the active site and water outside of the enzyme. Speculatively, this interface may permit water intake and diffusion as well as a possible means of water ion transport. Regardless of the possible implications of a large interface between active site contiguous water clusters and bulk solvent, no such data has been publicly recorded. While these values were available to us, they seemed appropriate to document.

Justification of Cluster Distance Criterion

The use of 5 Å as a cutoff distance for grouping water with respect to other water molecules and enzyme atoms was informed by literature reporting calculated interatomic distances during the autoionization of water. To understand the basis of this literature, the following must be known: While water's autoionization is difficult to simulate, predominately due to the temporal restrictions intrinsic to quantum MD, it is reversible. Thus, the distance at which hydronium and hydroxide recombine to form a pair of water molecules is the distance at which a pair of water molecules will ionize one another.

Using a series of MD trajectories, Bai and Herzfield calculated that the likelihood of recombination of hydronium and hydroxide is roughly equivalent to the likelihood of their

drifting apart at a hydrogen bond distance of ~6 Å.[54] Their study made use of the LEWIS water model, which a) is fully dissociable, b) represents the valence electrons around the central water oxygen (O^{2-}) as four pairs of $2e^-$ objects and the central oxygen itself as "an O^{6+} core", c) more accurately depicts water's reactivity and physical properties than models like TIP3P.[100] Using single-photon excitation, Natzle and Moore experimentally found this recombination distance to be 5.8 ± 0.5 Å.[46] Another study, making use of QM and MD methods, corroborated the distance at which recombination begins to be ~6 Å, adding that the computed r(O-O) in water wires is around 2.6 Å, and that neutralization occurs in ~0.5 ps from the 6 Å separation distance between hydronium and hydroxide.[53] As it would be necessary for a distance less than the autoionization/recombination distance of water/water ions to be found between neighboring water molecules in water clusters presumed to function as water wires by transporting charges via the Grotthuss mechanism, a conservative maximum r(O-O) of 5 Å was chosen to be the cutoff for including water molecules in the clusters of water found within the enzyme structures.

In order to determine whether the average r(O-O) between each water molecule and its nearest neighboring water molecule ($r(O\text{-}O)_{min}$) in each of the enzyme structure's water clusters corroborated the r(O-O) ≈ 2.6 Å profile of a water wire, $r(O\text{-}O)_{min}$ values were collected for each water molecule in each cluster studied. In other words, on the basis of O-O distance, the nearest water molecule to each water molecule was identified and the distance r(O-O) between the two were recorded. These distances, and many other distances between atoms computed in *analyze_wires.py*, were obtained by imputing each atom's cartesian coordinates into the *NearestNeighbors* function within the scikit-learn Python package.[101] This function effectively computes an array of distances between two arrays of coordinates. To find the nearest neighbor

to a given point, the coordinate of the point and an array containing the coordinates of all other points are entered into *NearestNeighbors*. The n^{th} index of the array generated by *NearestNeighbors* yields the distance between the point and the n^{th} point in the array of all other points. This method was found to be significantly faster than pure Python code written to the same end.

Control

As a means of assessing the uniqueness of active site solvent accessibility relative to other parts of enzymes, water wires were generated as seeded by random locations, rather than active sites, within each enzyme structure. A random location was selected in each enzyme structure by selecting a random enzyme residue by its positional index[vii], n, and the residue at position n+1. These residue pairs were treated as *faux* active sites and subjected to the same water cluster analysis as true active sites using *analyze_wires.py*. With this, one prospective active site-contiguous water cluster and one random location-contiguous water cluster were elucidated for each enzyme structure.[viii] It was expected that if active sites do indeed have special solvent accessibility relative to other locations within enzymes, then clusters of water molecules contiguous with active site residues would be larger on average than those contiguous with random adjacent residues. Furthermore, it would be expected that the number of structures having no water near random adjacent residues would be far greater than those with no water near active site residues[ix]. Figure 7 shows these expectations in graphical format.

[vii] e.g. for residue Tyr14, the positional index is 14.
[viii] The word *prospective* is used because a structure with no water near its active site would be imputed during data analysis as having a cluster with zero water molecules
[ix] The number of structures with no water molecules near active site residues would ideally be zero if the hypothesis of universal active site solvent accessibility is perfectly true.

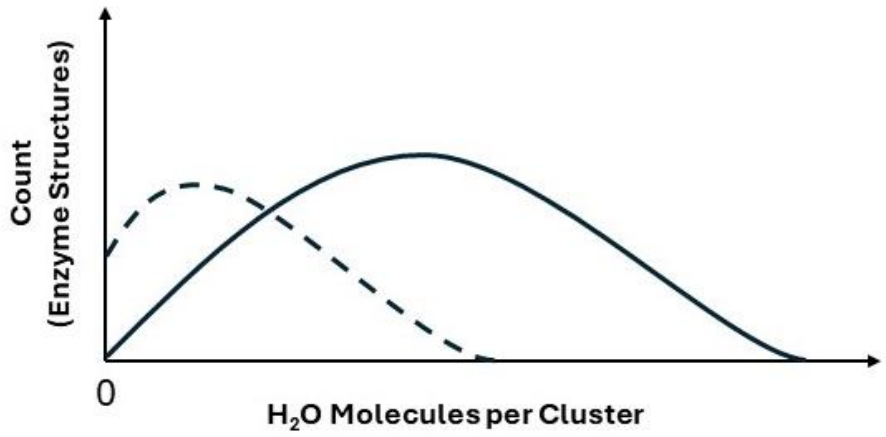

Figure 7. With one water cluster assigned to each enzyme, ideal distributions of the number of enzyme structures having water clusters contiguous with active site residues (solid line) and a location consisting of two random adjacent residues (dashed line) of a given size (number of water molecules).

Electrostatics

Summary of Methods

In this portion of the study, the primary objective was to investigate the potential of enzyme active sites to ionize water molecules or, at minimum, favorably influence their ionization. This was approached by analyzing enzyme structures obtained as PDB files and employing Coulomb's law using point charges within each to assess the strength of electric fields experienced by water molecules in each structure. A key aspect of this methodology involved comparing the electric fields experienced by water molecules within the active sites to those encountered elsewhere within the enzyme structures, thereby evaluating the uniqueness of these fields in the active sites of enzymes. It was expected that, were enzyme active sites to favorably influence the ionization of water by means of electrostatics, the electric fields experienced by water molecules in the active sites of these structures would be greater than those experienced by water molecules in the bulk solvent and elsewhere within the enzymes.

In addition to calculating the electric fields experienced by water molecules in a large sample of static enzyme, structures, the active sites of selected enzyme models were subjected to *ab initio* quantum mechanics by way of atom centered density matrix propagation (ADMP) calculations, enabling the observation of electric fields acting on water hydrogen atoms during routine molecular motions. To comprehensively capture the electric fields on water molecules throughout an entire enzyme during molecular motion, Molecular Dynamics (MD) simulations were conducted on the triosephosphate isomerase (TPI) enzyme.

These electric field calculations serve as a sound means of measuring the proclivity for water molecules to ionize due to the sensitivity of water's ionization to electric fields, as discussed in the introduction chapter of this dissertation.[57-59, 61]

The same PDB files containing the enzyme structures as per the water cluster analysis portion of this work were used for the analysis of electric fields experienced by water molecules. The active site locations of each enzyme were parsed using the same web scraping method used previously[x] as well. Water molecules within each file were split into "partitions" based on their locations: near the active site, elsewhere in the enzyme, and in the bulk solvent outside of the enzyme. A python code was written to obtain the charges of each atom in each enzyme structure using the AMBER99SB-ILDN[102, 103] forcefield. Figure 8 shows a conceptual diagram of this process.

Using Coulomb's law, the electric fields at the point of each water molecule were calculated from these point charges. Because the hydrogen in each PDB enzyme structure were added in an unphysical fashion using PyMOL command *h_add*, the fields on the water molecules were

[x] For the analysis of water clusters, in the program file *analyze_wires.py*.

calculated at the point of their central oxygen, the location of which was experimentally resolved by the author of each PDB file. The fields were then compared among the water molecules in each partition.

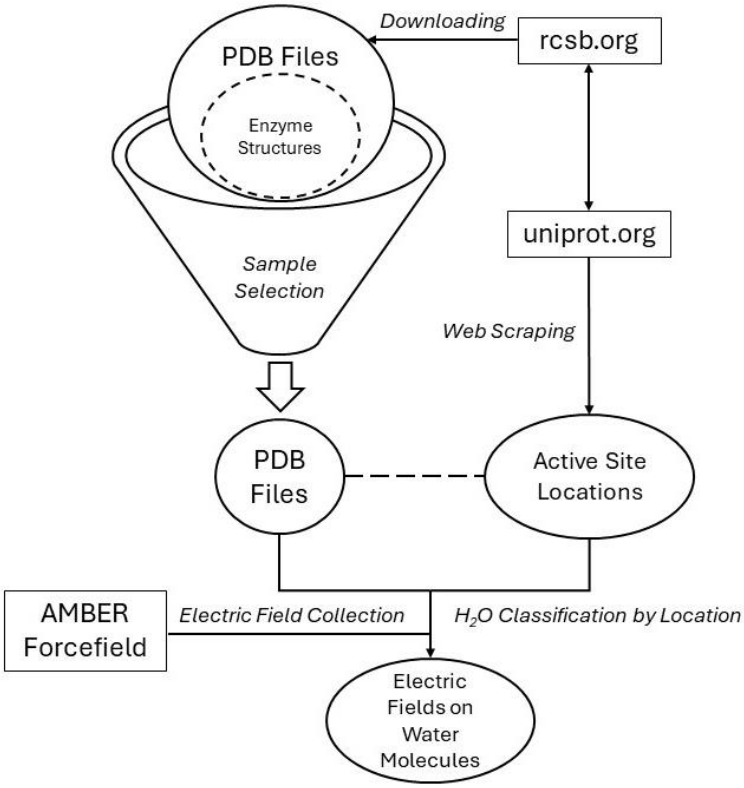

Figure 8. Conceptual process diagram for obtaining and processing PDB files to calculate the electric fields experienced by water molecules within them.

To observe electric fields on water molecules during molecular motion, the active sites of eight enzyme structures were abstracted from their PDB files (~200 atoms per active site) and subjected to *ab initio* QM calculations for 100 fs. The same was done to a box of 96 TIP3P water molecules. These fields were collected at the spatial coordinates of each water molecule's hydrogen, effectively measuring the proclivity for each to be deprotonated within its electrostatic environment. The electric fields experienced by each water molecule's hydrogen atom were

computed every 0.1 fs in each QM trajectory. To do this for a given water hydrogen (H^*), the field contributions of all other point charges ($\vec{E}_{H-x}$) at the coordinate of H^* (other than the O and H of the same molecule as H^*) were summed into a vector representing the net electric field on H^* ($\vec{E}_{net}$). The component of $\vec{E}_{net}$ in the unit vector ($\widehat{OH}$) of the OH^* bond vector ($\overrightarrow{OH}$) yielded the magnitude of the external electric field along the OH^* bond at the point of H^*, $E(H) \cdot \widehat{OH}$. Figure 9 provides a visual for this calculation. These electric fields were then compared between the cumulative active site calculations and the box of bulk water.

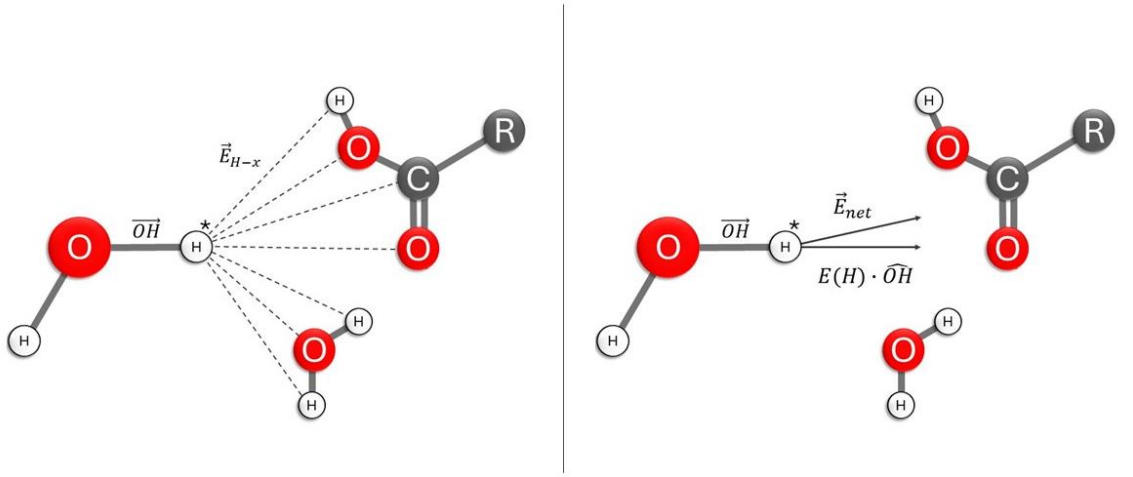

Figure 9. A visualization of the calculation of electric fields experienced by a given water hydrogen atom (H^*) in an enzyme structure, projected along the unit vector of the OH^* bond. Point charge contributions to the electric field on H^* are shown on the left. The net electric field on H^* and the net electric field projected along the OH^* bond are shown on the right.

The same electric field calculations used for the trajectories of enzyme active sites during *ab initio* QM were then applied to water hydrogen atoms during 2000 ns of classical MD simulation of triose phosphate isomerase (PDB: 1NEY).[104] The water molecules in this model were classified into partitions similar to the static PDB enzyme structures studied, with the exception of the partition of water near the active site, which was replaced by the partition of water near the

substrate dihydroxyacetone phosphate (DHAP) in the active site. This was done to assess the electric fields on water specifically near the substrate, which has a strongly electronegative phosphate group. To test whether TPI itself enhanced the electric fields on water molecules near DHAP, DHAP was then placed in a simulation box of water molecules and this calculation was repeated.

Calculating Electric Fields in Static PDB Structures

To broadly measure the electric fields felt by water hydrogen atoms in enzyme active sites relative to bulk solvent and other locations near and within enzymes, all enzyme structures used for the water cluster analysis were reused. In addition, the enzyme structures which were downloaded for the cluster analysis but were not compatible with PACKMOL were used for this study, giving a total of 1179 enzyme structures.

As per the preparatory step taken during water cluster analysis, the *h_add* command within PyMOL was used to add hydrogen atoms bonded to existing atoms in each structure. This method does not protonate basic atoms in charged residues, thus retaining their charge states in the crystal structure of the enzyme. In each model, water molecules were partitioned according to their location: those with the distance between the water oxygen and another atom (r(O-X)) < 4 Å from active site atoms, those with r(O-X) < 4 Å from any non-active site enzyme atom, and those in the bulk solvent: defined by r(O-X) > 4 Å from all enzyme atoms (fig. 10). For each partition, the electric field magnitudes experienced by whole water molecules ($E(H_2O)$) were calculated using the point charges of atoms assigned via the AMBER99SB-ILDN forcefield, which were stored in a GROMACS-compatible .rtp file prior to these calculations.

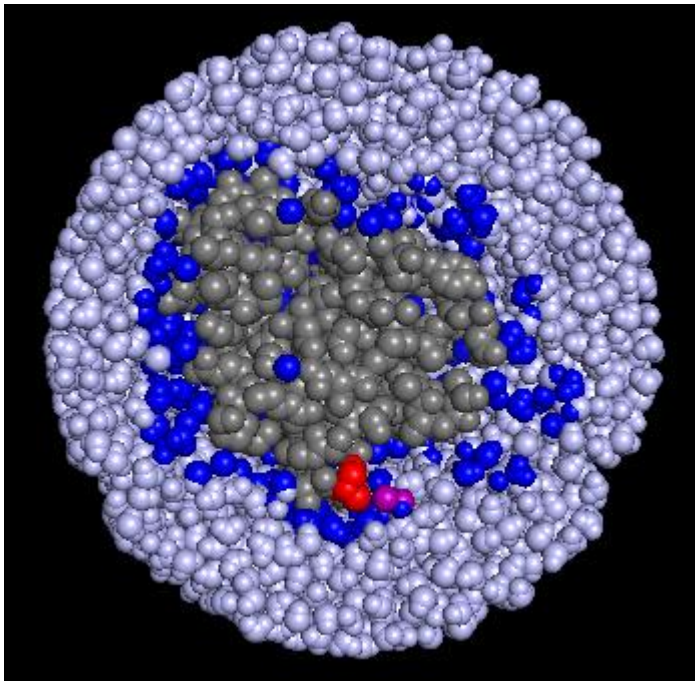

Figure 10. A cross section of an enzyme structure in VDW radii representation showing enzyme atoms (gray), active site atoms (red), H_2O < 4 Å from enzyme atoms (dark blue), H_2O < 4 Å from active site atoms (purple), and bulk solvent: H_2O > 4 Å from enzyme atoms. The electric fields experienced by each of these partitions of water were compared in this study.

E(H_2O) were calculated at each water oxygen atom rather than at each hydrogen atom as done for the outputs of quantum and classical MD later. This was due to the fact that the hydrogen included in each structure via *h_add* were not oriented in a manner informed by any physical features of the structure other than through the fixed OH bond length of ~95 pm and the impermissibility of clashes between the added hydrogen and the VDW radii of other atoms. Assuming that a water molecule will generally reorient such that the positive end of its dipole—the vector bisecting the H-O-H angle, will be closest to nearby negative charges, the fields on entire water molecules approximate a low estimate of the electric fields along their OH bonds. Accordingly, unlike the calculations depicted in figure 9, these fields were taken at the point of

each water molecule's central oxygen atom and were not projected along the OH bond unit vector.

E(H_2O) were calculated for each water molecule in each partition of water in each enzyme structure. The quantitative distributions of these fields were then compared on the basis of spatial partition of the water molecules associated with them.

Calculating the Minimum Electric Field Required to Ionize Water

We sought to establish the minimum external electric field required to ionize a water molecule using the same quantum mechanical calculation scheme later applied to determine such fields in enzyme active sites. A water molecule was placed in a simulation box such that one of the OH bond vectors ($\overrightarrow{OH}$) was parallel to the standard x-axis. Replicate copies of this molecule were then subjected to increasingly negative external electric fields in time-dependent quantum mechanical calculations using a polarizable Hartree Fock (HF) basis set, HF 3-21g* propagated via Atom Centered Density Matrix Propagation (ADMP).[105] Each calculation was conducted for 200 fs simulation time. Initially, the field was increased by 0.01 a.u. (1 a.u. = $5.142 \cdot 10^{11}$ V/m = 51.42 V/Å) between trials. Once a field of sufficient strength to ionize the molecule was found, further trials were conducted wherein the field was decreased by 0.0001 a.u. until the threshold of autoionization, to this precision, was calculated. Lest minor differences in the molecule's geometry upon initialization influence the outcome, this was repeated using a newly initialized water molecule.

Calculating Electric Fields in Enzyme Active Sites; Ab Initio Quantum Mechanics

The electric fields experienced by water molecules in the active sites of eight enzymes were calculated during *ab initio* QM ADMP trajectories in order to measure the electrostatic proclivity for ionization of water within active sites during realistic molecular motion. To do this, enzyme crystal structures were selected from the RCSB PDB. Enzymes were selected on the basis of their subjective popularity in enzymology. If a structure of < ~2.5 Å resolution with no mutations could be found for a given enzyme, it was downloaded for analysis.

The active site residues of each enzyme were identified manually using the in-browser molecular viewer available on rcsb.org. In PyMOL, a water molecule within 5 Å of at least one of these residues was selected. A sphere around this water molecule with ~16 Å diameter was then extracted from the structure. To uphold the electrostatic profile of the active site while minimizing the calculation time during ADMP, partial fragments of residues that were bisected during this process were selectively removed on the basis of perceived electrostatic importance and size; large or charged residue fragments were kept while small or uncharged fragments were removed in PyMOL. The resulting trimmed enzyme structure was then imported into GaussView where it was further trimmed to ~200 atoms on the basis of electrostatic relevance and molecular fragment size. Corrections to the structure were occasionally made to remedy errors caused by importing the file (e.g., a double bond between an oxygen and sp^3 carbon).

The enzymes used for this analysis were triosephosphate isomerase (1NEY),[104] MoFe nitrogenase (7JRF),[106] human carbonic anhydrase II (4YXI),[107] staphylococcal nuclease (2SNS),[108] DNA photolyase (1TEZ),[109] Phosphoribosylamine-glycine ligase (3MJF),[110] apo-form

cationic trypsin (5MOP),[111] human ubiquitin-conjugating enzyme (2ESK),[112] horse liver alcohol dehydrogenase (7RM6),[34] and thioredoxin M1 (7C65).[113]

Each active site was subjected to ADMP over 100 fs of simulation time using the HF 3-21g* basis set. At each frame of the simulation, the electric field on the hydrogen belonging to each water molecule was calculated as projected along ($\overrightarrow{OH}$) from the central oxygen. As a control for the fields experienced by water hydrogen in bulk water, a cubic simulation box with an edge length of 20 Å containing 96 water molecules was subjected to these calculations as well.

Active sites were simulated using implicit solvation in acetone to emulate the dielectric of enzymes. Acetone possesses a dielectric of $\varepsilon \approx 20.5$,[114-116] while proteins are estimated to contain an dielectric constant of $2 \leq \varepsilon \leq 10$ internally, with surface regions estimated as high as $\varepsilon = 30$.[117, 118] As a higher dielectric constant of a solvent corresponds with a smaller electric field incurred by solute within that solvent, the use of acetone was a conservative choice, being at the higher end of the range of estimates for ε within proteins. The box of water was simulated with implicit solvation in water ($\varepsilon \approx 78.4$)[116]. Atomic charges were assigned on the basis Mulliken Population Analysis (MPA), which are influenced by implicit solvation.[119] Given that Mulliken charges were used in the calculation of E(H) and the ε values of implicit solvents altered these charges presumably consistent with the effect of factoring ε into Coulomb's law directly, the dielectric constant used in Coulomb's law for calculating electric fields was left as the permittivity of free space, ε_0.

Calculating Electric Fields on Water Hydrogen in TPI
During Classical MD

Triosephosphate isomerase (TPI) catalyzes the interconversion of dihydroxy acetone phosphate DHAP. DHAP contains a negatively charged phosphate group exposed to bulk solvent in the crystal structure of the enzyme (PDB ID 1NEY).[104] To probe the proclivity of water to be deprotonated near DHAP in the active site, model 1NEY was subjected to 2000 ns of MD simulation at 300K.[xi,120] At each 10 ns frame of the subsequent trajectory, water molecules were partitioned into the same groups as done for other XRC structures apart from the water molecules classified as being near the active site, which were instead qualified by being < 5 Å from the DHAP molecule in the active site rather than active site residue atoms themselves. As a metric of the proclivity for deprotonation, the field experienced by each water hydrogen at each frame was then calculated as projected along the bond shared with its central oxygen using Coulomb's law. Point charge contributions from this oxygen and the sister hydrogen were excluded. To compare the fields experienced by water hydrogen near DHAP in TPI with those experienced by water hydrogen near DHAP in bulk solvent, the DHAP atomic coordinates were copied from 1NEY into a new file and solvated in water. This model of DHAP in water was then subjected to the same MD parameters and water hydrogen field analysis as for the TPI structure.

To further validate results of deprotonating a water molecule under applied fields during ADMP, we conducted a series of Self-Consistent Field (SCF) scans along the hydroxyl bond of the water molecule. These scans were conducted under varying external electric fields, as done

[xi] This MD simulation was created by former undergraduate student Max Yates in our research group. The subsequent work described is a significantly modified extension and validation of his own. See attached reference.

for ADMP calculations of deprotonation, and calculations were performed using the HF 3-21G* basis set to maintain consistency with active site simulations.

In each SCF scan, we plotted the energy profile of the hydroxyl bond as a function of its length. The activation energy required for bond breaking was determined by subtracting the energy at the minimum point (preceding the transition state) from the energy at the maximum point (transition state). The bond length was scanned in increments of 0.05 Å from 0.96 Å until, at minimum, a decrease in the energy of the system was seen following the visible peak in energy associated with the transition state.

The activation energy of hydroxyl bond breakage in each SCF scan was then plotted against the electric field applied along the hydroxl bond. The regression function from these data was then used to extrapolate the electric field at which the activation energy was 0 Hartrees. This value could then be used to validate any water-deprotonating events in the ADMP calculations of enzyme active sites and argue for the case that any given water molecules in each could be deprotonated given sufficient simulation time.

CHAPTER THREE

IS WATER A COFACTOR FOR ALL ENZYMES? SURVEYING ENZYME CRYSTAL STRUCTURES REVEALS THE COMMONALITY OF ACTIVE SITE SOLVENT ACCESSIBILITY AND ENZYMATIC WATER NETWORKS

Contribution of Authors and Co-Authors

Manuscript in Chapter 3

Author: Caleb M. T. Sindic

Contributions: Majority of text in all sections, creation of all data and figures other than 11 & 12.

Co-Author: Pedro L. Muiño

Contributions: Figures 1 & 2, data for figures 1 & 2, corrections to text.

Co-Author: Patrik R. Callis

Contributions: Figures 1 & 2, data for figures 1 & 2, corrections to text.

Manuscript Information

Caleb M.T. Sindic, Pedro L. Muiño, Patrik R. Callis

Status of Manuscript:
☒ Prepared for submission to a peer-reviewed journal
☐ Officially submitted to a peer-reviewed journal
☐ Accepted by a peer-reviewed journal
☐ Published in a peer-reviewed journal

Abstract

Despite the demonstrable dependence of enzyme functionality on solvation, the notion of water being directly chemically required for catalysis inside active sites remains unexplored. Here we report that over 99% of 1013 enzyme crystals obtained by X-ray crystallography with high resolution (< 1.5 Å) contain tunnels of water linking residues within the active site to bulk water. Also reported are the findings which inspired this study—that electric fields experienced by water hydrogen atoms are on average twice as strong in the active sites of two enzyme structures compared to those in bulk water. These results point to the possibility that water molecules within active sites may be paramount to the immense catalytic power of enzymes.

Introduction

The Grotthuss mechanism is a process wherein rapid proton and electron transfer across water is attributed to the reversible formation of unique structures known as water wires.[52] These water wires consist of chains of water molecules all within proton transfer distance of one another and all oriented in a manner that lowers the activation energy barrier gating proton transfer. Hydronium or hydroxide found on one end of the wire can be readily transferred to the other end via rapid proton transfer occurring sequentially between each water pair.[121] Water wires have been observed to form in both bulk water and complex molecular environments.[122] In the context of biology, water wires were first observed to play critical roles in transmembrane proteins. They have been especially characterized in the “model membrane protein”, the gramicidin A channel.[66, 123] In one particularly interesting study, a redox-responsive water wire was calculated to function as a means to prevent the short-circuiting of the cytochrome c oxidase

proton pump.[63] Many examples of water wires in non-transmembrane enzymes exist as well. One such example is the heme enzyme ascorbate peroxidase, where the involvement of an intermediary water molecule in a concerted proton transfer pathway from the enzyme's substrate to its heme-bound oxygen has been found to be critical for oxidation.[71] In Green Fluorescent Protein, negatively charged surface residues draw protons from the bulk solvent to be shuttled to the active site via a water wire.[87, 89] In Ribonucleotide Reductase, evidence exists that the catalytically critical tyrosine radical is protonated by hydroxamic acid through a proton wire containing four water molecules.[73] In photolyase, QM/MM calculations have demonstrated the assistance of a water wire to facilitate DNA damage repair through proton-coupled electron transport over >10 Å.[79] Alongside these individual findings, X-ray crystallographic (XRC) studies have demonstrated that 44% of side chain oxygen and nitrogen atoms in enzymes are hydrated.[44]

In addition, computational work by Warshel, *et al*. has provided evidence that enzymes are efficient catalysts because their structural rigidity and protection from bulk water maximizes the electrostatic interaction from active site charged residues, thereby lowering activation energy.[124] Water has been calculated via high-level *ab initio* computations to ionize under fields as low as 0.35 V/Å ($6.8 \cdot 10^{-3}$ a.u.).[57] The products of ionization, hydronium and hydroxide are routinely proposed as participants in published mechanisms of enzyme action involving bond breaking and making because of the requirement of strong, rapid acting nucleophiles or electrophiles; there are occasional reports of their observation within active sites of enzymes.[125, 126] The production of these species has been associated with a means of transporting charge across large distances within enzymes where reactive components are disparate.[88, 127-129] In GTPase, for example,

networks of water recruit charges from distant residues to enhance the nucleophilicity of an active site water molecule which then hydrolyzes guanosine triphosphate.[130, 131] The Boxer group has reported findings connecting strong electric fields within the active site of the enzyme ketosteroid isomerase (KSI) along with evidence that the field is vital to its catalytic power.[132] The importance of water in enzymatic action is further underscored by observations that enzyme reaction rate increases in non-aqueous solvents as the concentration of water increases for solvents that do not significantly change the enzyme's structure.[133]

Preliminary work by our group on two enzymes has revealed that electric fields promoting water ionization are twice as strong on water in their active sites than in bulk water. We therefore believe that water should demonstrate an increased proclivity towards autoionization when located in these enzyme active sites. From crystal structure of polynucleotide kinase, 4QM6[134], wherein the O2'-H2' bond of the riboguanosine (RG) nucleotide is H-bonded to its 3'phosphate, we have previously computed the strength of electric fields acting along this O2'-H2' bond and along water hydroxyl bonds in the active site to promote ionization. Electric field magnitudes were compared with those along the hydroxyl bonds of bulk water lying outside the surface of the enzyme. These results are shown in figures 11 and 12.

Exceptionally strong electric fields were calculated along the O2'−H2' bond of the RG nucleotides in the active sites of subunits C and D in bacterial polynucleotide kinase (fig. 11). The electric field magnitudes experienced by water molecules along their hydroxyl bonds were then calculated in bulk water, as well as among water molecules within 5 Å of guanosine triphosphate (GTP) and the phosphate molecule within the active site of bacterial polynucleotide kinase. A noticeable increase in these fields was observed among water near the active site bound

GTP and active site bound phosphate relative to those in bulk water (fig. 12). We thus postulate that the ribose O2' proton transfer in the proximity of the active site would almost certainly generate hydroxide that would be the strong base required for nucleophilic attack of the scissile bond. With these findings corroborating the numerous examples of catalytically pertinent water wires in enzymes given above, we propose that water is a universal chemical cofactor in biological catalysis.

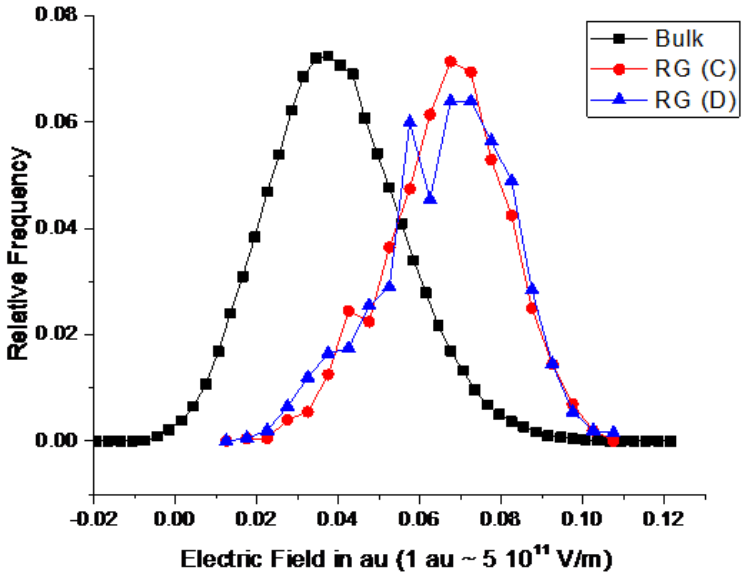

Figure 11. A histogram of the exceptionally high electric fields felt by the O2'−H2' bond of the RG nucleotide in the active site of bacterial polynucleotide kinase.

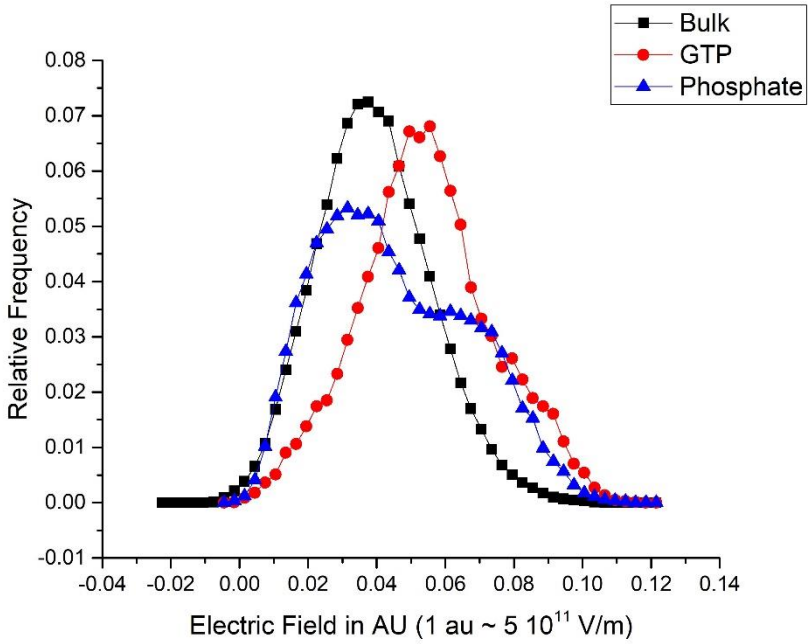

Figure 12. A histogram of the exceptionally high electric fields felt by water molecules near the active site bound GTP (red circles) and phosphate (blue triangles) in bacterial polynucleotide kinase, and in bulk water (black squares).

If water is chemically required for catalysis, an obvious expectation is that most enzyme crystal structures will reveal water within the active site and a mechanism, e.g., a water wire, providing access to bulk water. This manuscript describes a procedure and the results from 1013 published enzyme XRC structures for which active site atoms are identified and for which we identified those that exhibit both water within their active sites and water wires linking the active site to bulk water. These are features that portend the existence of a continuous supply of water to the active site and has been shown to initiate the regeneration of the active site's charge state via the diffusion of charges in multiple enzymes.[68, 135] The number of water molecules in each water wire was then compared to that in bulk solvent to characterize the feasibility of autoionization within them.

Materials And Methods

1013 enzyme structures were procured in the form of protein database (PDB) files from rcsb.org. Only high-resolution X-ray crystallography (resolution < 1.5 Å) structures were used. Duplicate enzymes were removed from the study to ensure that each structure corresponded with a unique structure. Because the goal of this study was to determine the spatial relationship between water and the active sites of these structures, the locations of the active site residues in each were obtained. This was done via a custom web scraping algorithm written in Python. This algorithm searched the HTML of a given structure's rscb.org page for links to uniprot.org. From the uniprot.org link associated with each enzyme, a text file corresponding to the locations of critical components of the enzyme was parsed for the residue identities belonging to active site atoms. These were denoted in the text file by the term "active site" followed by an index corresponding with the residue's position in the polypeptide chain(s) comprising the enzyme. These indices were then associated with the cartesian coordinates of active site atoms in the PDB file stored on the operating computer.

A dilemma was encountered when identifying active site residues using the aforementioned process. For many structures, namely those of the oxidoreductase class, the text file corresponding to the structure on uniprot.org did not possess residues denoted by the "active site" keyword but did possess residues denoted by the "binding site" keyword. It was manually verified that these binding site residues were not members of allosteric binding sites but were located in the active sites of the structures. In other cases, structures possessed both "active site" and "binding site" residues in their associated uniprot.org text files which were manually found near one another within the structures' active site pockets. For example, the residue GLU168 in

glutaminyl cyclase (PDB ID 4MHN), responsible for coordinating substrate binding and denoted as a “binding site” residue on uniprot.org, is located immediately next to the catalytic residue GLU183, a proton acceptor in the active site, which is denoted as an “active site” residue on uniprot.org.[136] Given their proximity, both residues were considered to approximate the active site location within the structure. Accordingly, both residues deemed “binding site” and “active site” on uniprot.org were considered to identify the active site locations within each structure used in this study.

In each structure, water molecules whose oxygen was less than 5 Å of any of the active site atoms were identified and counted. These waters composed a nascent water wire grounded at the active site, which was extended by iteratively adding water molecules within a 5 Å oxygen-oxygen distance (r(O-O)) of the growing water wire (fig. 13). After characterizing the water wires within each structure, we sought to assess the number of structures which possessed water wires contiguous with bulk solvent outside of the enzyme. Because the structures largely lacked water outside of the enzyme crystal, we employed PACKMOL[91] to simulate solvation. To prepare PDB files for solvation, a spherical boundary with a padding of 5 Å from the most distant enzyme atom from the structure center was constructed around each structure. Using the atomic radii of the atoms in the PDB file, assigned via the chemlib[98] python package, the volume occupied by the original structure was ascertained. To calculate the volume to be occupied by the simulated bulk solvent, the volume of the original structure was subtracted from the volume of the boundary sphere. The resulting volume was multiplied by the density of water at 310 Kelvin and 1 atm (~0.03 molecules / $Å^3$) to determine the number of water molecules to be added. The files were then solvated by PACKMOL using the TIP3P[99] water structure. In each of the

subsequent structures, a new partition of water was defined: bulk solvent, defined by water greater than 5 Å from all enzyme atoms, ensuring it was outside of the enzyme. The structures possessing water wires having at least one water molecule with an r(O-O) < 5 Å of any of these bulk solvent molecules (bulk solvent contiguity) were then counted.

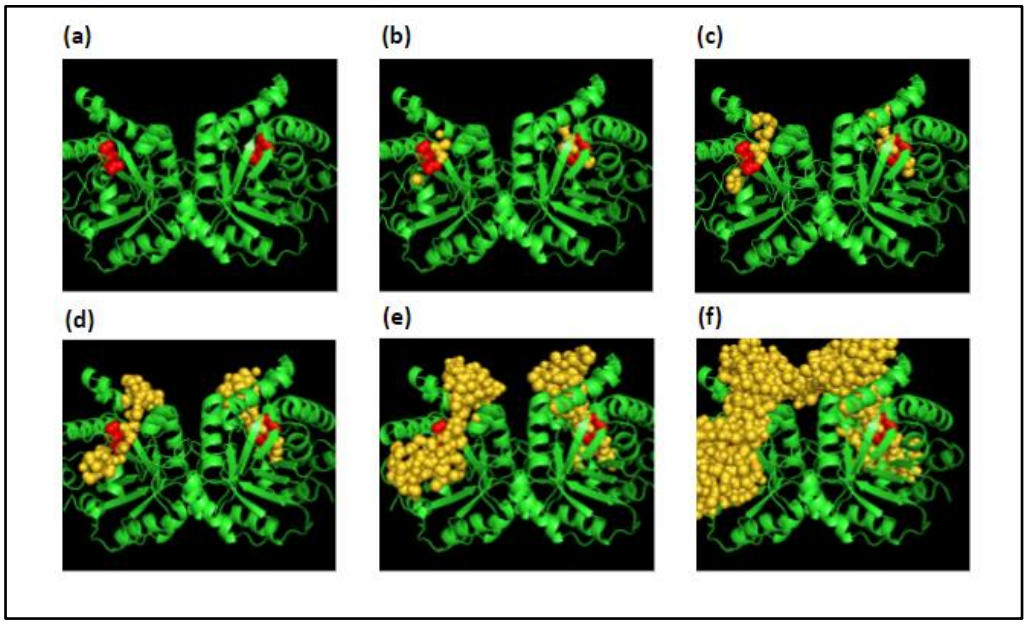

Figure 13. An illustrated example of elucidating a water wire for the active site of triose phosphate isomerase (XRC crystal structure resolution: 1.2 Å, PDB code: 1NEY).[104] The atoms of active site residues Glu 165 and His 95 can be seen as red spheres, the enzyme structure as a green cartoon depiction, and the water atoms belonging to the water wire as yellow spheres. Figure (a) shows the structure before probing for water near the active site residues. Figure (b) shows water molecules within 5 Å of active site atoms. Figures (c) through (f) show the recursive expansion of the water wire by adding water molecules whose oxygen are within 5 Å of the current water wire oxygen atoms. In figures (e) and (f), the water wire has reached the bulk solvent and can be seen extending far from the enzyme. By this point, the wire-generating algorithm would stop as the solvent has been contacted.

The use of 5 Å as a cutoff distance for grouping water with respect to other water molecules and enzyme atoms was informed by previous calculations on the autoionization of water. Using a series of LEWIS[100] water recombination MD trajectories, it has been shown that the likelihood of recombination of hydronium and hydroxide is roughly equivalent to the likelihood of their

drifting apart at a hydrogen bond distance of ~6 Å.[54] Other MD studies show that below this distance, recombination is highly favorable.[46, 53] While the forward process of autoionization in pure water is difficult to simulate over the timescales permitted by MD without biasing parameterization, it is reversible. Thus, the recombination distance is a proper metric for the distance at which water ion pairs can be expected immediately following autoionization. Accordingly, to ensure confidence in the distance between water and other bodies in the structures being sufficient for ionization, and to accommodate for the resolution of the structures (up to 1.5 Å), a 5 Å distance was chosen. More information on this value and its connotations is given in the discussion section below.

As a control, the number of water molecules near target (active site) residues, the number of water molecules in wires emanating from these, and the number of solvent molecules contacted by the wires were obtained for each structure using a random location in the enzyme as a comparative “active site”. To this end, two random adjacent residues in each structure were selected to obtain a target point of reasonable size and subjected to the same water topology analysis conducted for the active site. Accordingly, each structure was assigned one active site contiguous water cluster and one random site contiguous water cluster.

Because the proclivity for proton transfer in a homogenous liquid is partially dependent on the distance between molecules, the densities of the water wire and bulk solvent were recorded. The chosen metric of density for each partition was the distance between each water oxygen and its closest neighboring water oxygen, calculated using the scikit-learn NearestNeighbors[137] method for Python.

Results And Discussion

Of the 1013 enzyme structures studied, 1005 were found to have water within 5 Å of active site atoms. This finding establishes a structural precedence for solvent-assisted catalysis in enzymes because the majority contain water within their active site pocket. When water wires were generated from water molecules in the active sites of each structure, most were found to host a large network of water molecules contiguous with their active sites. Across the water wires of the 1013 enzyme structures studied before the addition of solvent molecules, there was an average of 333 water molecules per wire when seeded from active sites and 242 water molecules when seeded from two random adjacent residues (table 3, fig. 14). Note that the areas under both curves in figure 14 are equal on account of there being one active site contiguous water cluster and one random site contiguous water cluster assigned to each enzyme structure.

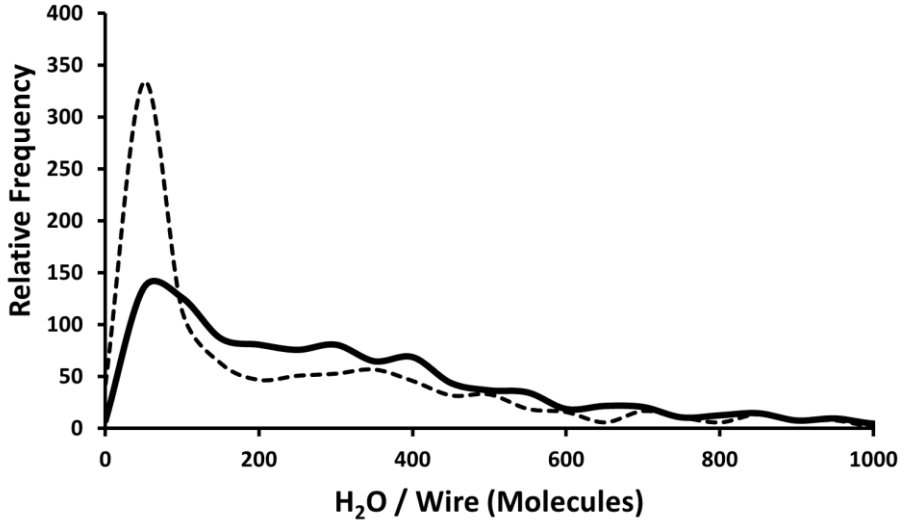

Figure 14. Histograms of water molecules per water wire originating from the active site (solid black line) and two random adjacent residues (dashed line) in each of 1013 crystal structures. Each water wire was grouped by the criterion of r(O-O) < 5 Å.

After the solvation of the structures using PACKMOL, contiguity between the active site and bulk solvent via water wires was observed in 993 of the 1013 structures studied. This feature spatially permits a supply of water to the active site and suggests that a Grotthuss-type mechanism of charge transport between water near the active sites and bulk solvent within these structures could be involved. The mean number of bulk solvent water molecules within the parameterized r(O-O) < 5 Å proton transfer distance of any member of each water wire was 1482, with a standard deviation of 740. This value was drastically increased by large structures whose water wires extended into grooves on the surface of the enzyme, with a large contact area for interaction with the bulk solvent. Irrespective of this subgroup, there are many avenues for water to infiltrate the active sites of enzymes. Likewise, in the instance of an ionizing event on the enzyme surface, there are many water molecules capable of receiving or initiating ionizing events to and from the active site of most enzymes via a Grotthuss-type mechanism.

In 15 structures, water was present within 5 Å of active site atoms but these waters did not form wires that reached the bulk solvent. The average number of water molecules in the active site contiguous water wires found in these structures was 24, significantly lower than the average value of 333 water molecules/wire found for all structures studied. Unless these enzymes were folded with their active sites entombed from the space around the protein, which would hinder substrate access, this finding is likely due to constrictions in the buried tunnels containing the wires which bifurcate them and subsequently prevent contact with bulk solvent. If this is the case, it would be expected that the enlargement of these constrictions during routine molecular motion would elicit a larger water wire size and the contiguity of the wire with bulk solvent.

When two random adjacent residues were selected in each structure as targets in lieu of active site residues, the number of structures that did not exhibit water molecules within 5 Å of target residues was significantly less than when active site residues were selected (48 and 8 structures, respectively). The number of structures that did not have target residue-adjoining water wires that reached the bulk solvent was 94 and 20 for structures after targeting random adjacent residues and true active site residues, respectively. Furthermore, smaller water wires of ca. 50 water molecules seeded from random adjacent residues were much more frequent than those seeded from true active site residues (fig. 14). These findings implicate the unique solvent accessibility of the active site of enzymes relative to other locations. This is as expected given the requirements of the active sites to be accessed by substrates through the bulk solvent, but it also demonstrates the enhanced ability for active sites to be contacted by the solvent itself.

All enzymes studied belonged to one of seven different classes: hydrolases, oxidoreductases, isomerases, lyases, ligases, transferases, and translocases. The hydrolases composed the largest class in the study (table 2). The distributions of water wire sizes based on enzyme class show similar features, with a large portion of enzymes falling under a curve at approximately 100-500 water molecules per wire (fig. 15).

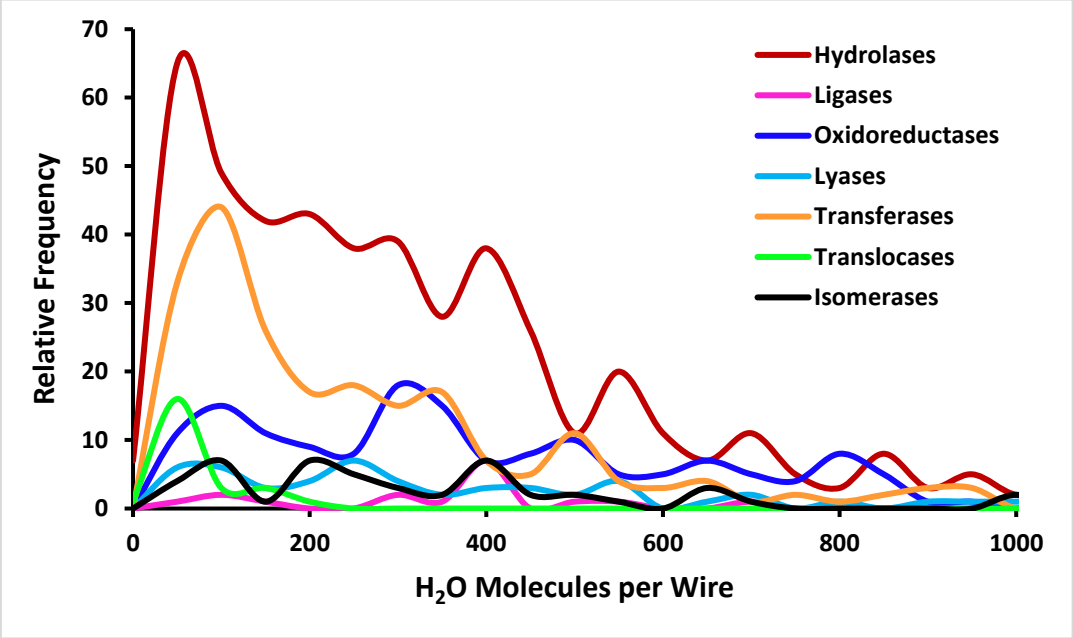

Figure 15. On the basis of enzyme class, histograms of water molecules per wire originating from the active site among 1013 enzyme crystal structures. Each water wire was grouped by the criterion of r(O-O) < 5 Å and required to have one member whose oxygen was this distance from any active site atom.

Of the eight structures that showed no water near their active site, seven belonged to the hydrolase class and one belonged to the translocase class. These structures were Klebsiella Pneumoniae carbapenemase 2 beta-lactamase[138], PI-SceI[139], Fluoroacetate Dehalogenase[140], human "a disintegrin and metalloproteinase with thrombospondin motifs" 4[141], phage T4 lysozyme[16], V84D mutant of S. solfataricus acylphosphatase[142], bovine trypsin[143], and 2-hydroxy-dATP diphosphatase.[144] While many of these structures seemed to be united only by their lack of water within 5 Å of their active site residues, the structure of phage T4 lysozyme was resolved by the authors to demonstrate the impermissibility of water within nonpolar pockets in the enzyme.[16]

	H_2O MOLECULES PER WIRE (MEAN)	H_2O MOLECULES PER WIRE (STD. DEV.)	NUMBER OF STRUCTURES WITH BULK SOLVENT CONTACTED	NUMBER OF STRUCTURES IN CLASS
HYDROLASES	330	353	470	482
ISOMERASES	300	224	45	47
LIGASES	458	370	20	20
LYASES	457	465	57	58
OXIDOREDUCTASES	444	487	151	162
TRANSFERASES	248	243	216	219
TRANSLOCASES	87	218	24	25
ALL STRUCTURES	333	368	993	1013

Table 2. Descriptive statistics for water wire sizes (count of water molecules/wire) and the number of structures having a water wire that contacts the bulk solvent for each class of enzyme.

The average water hydrogen-oxygen bond distance found in literature is 0.96 Å.[145] Assuming water may rotate to minimize the proximity of the positively charged hydrogen atoms to the nearest neighboring water oxygen, this distance, r(O-H*), may be within 4.04 Å of the nearest water oxygen at the maximum allowed r(O-O) of 5 Å for waters in these wires. Within the wires, the average r(O-O) between a water and its nearest neighbor, $r(O\text{-}O)_{min}$, was 2.85 Å (fig. 16). Provided again that a hydrogen atom may rotate around its central oxygen atom in water, this value corresponds with a possible range of $1.89 < r(O\text{-}H^*) < 3.81$ Å on average. These distances are well the r(O-H*) values observed in liquid water, where hydrogen bonding is frequent and proton transfers occur on the order of 10^{-5} sec^{-1} (about once per 10 hours).[56] Moreover, the $r(O\text{-}O) < 5$ Å parameter in the water wires is less than the calculated 6 Å separation distance of hydronium and hydroxide in liquid water immediately after ionization.[56]

In nitrogenase, which is responsible for the hydrogenation of N_2 to produce NH_3, a catalytically critical hydronium-shuttling pathway has been predicted via density functional theory with water-water distances fluctuating between 2.45 and 3.5 Å and hydrogen bonding characterized by r(O-O) ≈ 2.8 Å.[146] This correlates well with our average $r(O\text{-}O)_{min}$ = 2.85 Å found in water wires. On the basis of distance, water molecules in these networks are sufficient for hydrogen bonding and charge transfer via the Grotthuss mechanism.[56, 147]

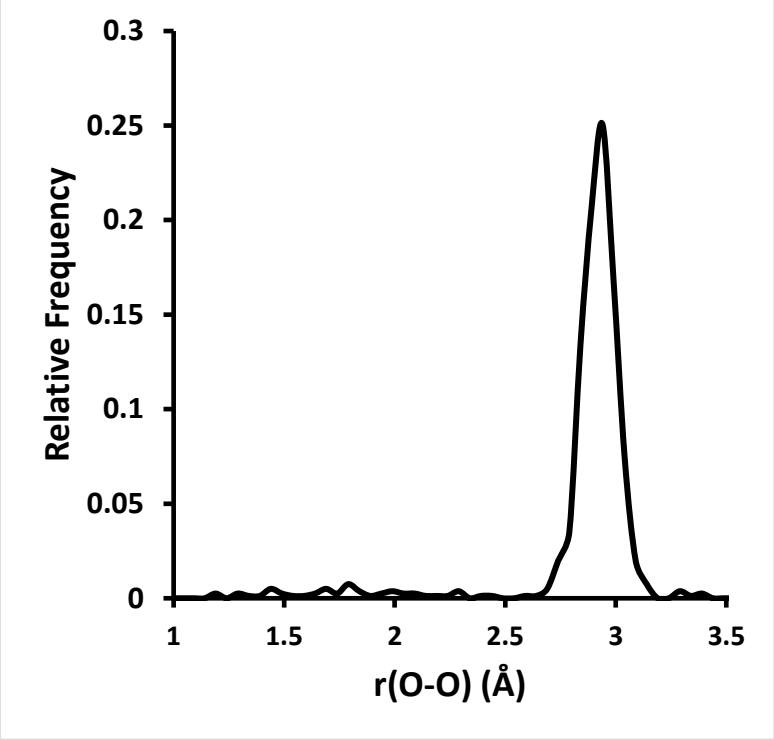

Figure 16. Oxygen-oxygen distances between each water molecule and its nearest neighboring water molecule within the active site-contiguous water wires of 1013 enzyme structures.

Conclusion

The results from our procedure strongly support the postulate that all enzymes require water as a cofactor. Water wires emanating from the active sites of the enzymes, clustered by a ≤ 5 Å oxygen-oxygen distance, were found in 99% of the 1013 x-ray crystal structures surveyed, with

an average of 333 water molecules per enzyme residing in water wires. This large value appears from our data to be attributable to water wires that connect water in active sites to water at the surface (bulk water). Because each water molecule in these wires was inherently selected by the condition of it being within a certain distance of its neighbors suitable for autoionization, the wires found here are capable of propagating hydronium or hydroxide ions to the active site, in addition to the finding that electric fields within the active sites are computed to be higher on average than in bulk water. Given that each water molecule in a system has a probability of initiating a Grotthuss-type charge transfer reaction, the size of these water wires demonstrates a vast number of locations from which a hydronium or hydroxide ion may be generated and conducted via a Grotthuss-type mechanism to the active site of the enzyme, depending on the local electric fields. Accordingly, most enzymes seem to structurally permit water exchange between the active site and bulk solvent.

Because enzymes are plastic and typically captured in only one conformation by x-ray crystallography, sampling numerous conformations of each structure in this study would permit a better understanding of water wire dynamics. One computationally intensive but powerful means to this end would be an extension of our analysis using temporal frames extracted from classical molecular dynamics trajectories created from the PDB structures of the enzymes studied. Such an approach would also allow for the visualization of the movement of water within the water wires found in this study if random fluctuations of the protein provide adequate chemical potential differences for water along the wires. Another viable method for confirming our findings is CAVER[148], a computational tool used to determine whether a given molecule can access a selected point in an enzyme from the bulk solvent.

Acknowledgments

James Vivian, PhD., Sydney Austad, Max Yates.

CHAPTER FOUR

CALCULATING ELECTRIC FIELDS IN ENZYME CRYSTAL STRUCTURES UNVEILS THE RELATIVE FAVORABILITY OF THE IONIZATION OF WATER MOLECULES IN ENZYME ACTIVE SITES.

Contribution of Authors and Co-Authors

Manuscript in Chapter 4

Author: Caleb M. T. Sindic

Contributions: Majority of text in all sections, generation of all data.

Co-Author: Patrik R. Callis

Contributions: Editing.

Manuscript Information

Caleb M.T. Sindic, Patrik R. Callis

Status of Manuscript:
☒ Prepared for submission to a peer-reviewed journal
☐ Officially submitted to a peer-reviewed journal
☐ Accepted by a peer-reviewed journal
☐ Published in a peer-reviewed journal

Abstract

Our group has previously conducted a comprehensive analysis of 1013 X-ray crystallography (XRC) enzyme structures, revealing that a striking 99% harbor water molecules situated within 5 Å of their active site. This observation is particularly intriguing considering water's pivotal role as the precursor to the highly reactive acid-base pair hydronium and hydroxide. Given that active sites within enzymes commonly exhibit heightened polarity and stronger electric fields compared to other regions within the enzyme, it stands to reason that water molecules in their vicinity are predisposed to ionization. Building upon this premise, we propose that electric fields within enzyme active sites routinely induce the ionization of water molecules, resulting in the generation of water ions that actively participate in biocatalysis. If our hypothesis proves to be substantiated, it underscores the imperative of considering the ionization of water and the consequential presence of hydronium and hydroxide ions in future enzyme research endeavors, highlighting a previously overlooked aspect with profound implications for understanding enzyme function and manipulating biocatalytic processes.

Introduction

Enzymes, the remarkable biological catalysts essential for countless cellular processes, have long captivated researchers with their unparalleled efficiency and specificity. Among the various theories proposed to elucidate the mechanisms behind enzyme catalysis, the electrostatic theory stands out as fundamental. This theory underscores the significance of electrostatic forces in substrate recognition, transition state stabilization, proton transfer, and the overall catalytic efficiency of enzymes.[8, 24, 25]

There is much evidence to support this theory. For instance, the electrostatic profiles of lysozyme active sites have been shown to remain conserved, even when the active site residues are not.[149] Similarly, the antibiotic resistance of β-Lactamases has been linked to the adaptive strengthening of electric fields within the active site.[39] Numerous successful efforts to improve catalytic rates by tailoring electric fields within enzyme active sites have been made.[43] Notably, a computed 43-fold increase in the productivity of kemp eliminase has been elicited through optimization of the fields exerted by the active site on the substrate.[150] Tailoring the fields within liver alcohol dehydrogenase has yielded a similar, experimentally validated, 50-fold increase in catalytic rate.[40]

Electrostatic interactions within enzyme active sites are driven by uniquely strong fields within the local environment. For example, the C=O bond of the ketosteroid isomerase substrate has been measured via Stark spectroscopy to experience ensemble average field magnitudes of 0.028 a.u. (1 a.u. = $5.142 \bullet 10^{11}$ V/m = 51.42 V/Å).[31] Similarly, liver alcohol dehydrogenase has been shown to exert fields along the reactive C=O and C-D aldehyde bonds of its natural inhibitor of magnitudes 0.032 and 0.019 a.u., respectively.[34] In the active site of uracil-DNA glycosylase, electric fields of 0.012 a.u. have been shown to drive catalysis.[37] In fact, the electric fields within active sites are of such significance that peaks in the electrostatic potential profiles of enzymes are predictive of active site locations.[151]

Our previous work has demonstrated that active sites are almost unanimously populated with water molecules. Of 1013 enzyme XRC structures studied, 99% were found to have water within 5 Å of active site residues. Remarkably, 98% of these structures were found to contain clusters of water molecules within 5 Å (less than the calculated 6 Å proton transfer distance)[56] of one

another, linking their active sites to external bulk solvent. We believe these water clusters may hypostatize Grotthuss-type[52, 71, 100, 121, 146] proton transfer and solvent induction in enzymes. Given that water has been calculated to autoionize above external electric field strengths as low as $7 \cdot 10^{-3}$ a.u.,[57] a value observed to be surpassed within select active sites,[31, 39, 152, 153] this leads to the question of whether the electrostatic environments of active sites may ubiquitously ionize water molecules. If true, the markedly reactive water ions hydronium and hydroxide are routinely generated within active sites an abundant chemical involvement of water in enzyme catalysis may be implicated. To probe whether the electrostatic environments of enzyme active sites encourage the ionization of water molecules, this work examines the absolute and relative electric fields experienced by water in static XRC enzyme structures, and during *ab initio* quantum mechanics and molecular dynamics (MD) simulations of enzyme active sites. It follows that if our hypothesis is proven correct, it has radical implications for our understanding of enzyme catalysis.

Methods

Electric Fields on Water in Static Enzyme Structures

To measure *en masse* the electric fields incurred by water molecules in or near active sites relative to those experienced by water molecules a) elsewhere in enzymes and b) in bulk solvent, 1179 Protein Data Bank (PDB) files of resolution < 1.5 Å were obtained from the RCSB PDB. A web scraping algorithm was employed to locate the active site residues of each enzyme structure within the PDB files using the UniProt webpages associated with each.

Using the *h_add* command within molecular viewing software PyMOL,[154, 155] hydrogen atoms bonded to extant atoms were initialized in each structure if excluded by the authors of the

structure. *h_add* does not protonate basic atoms in charged residues. Thus, molecular charges of amino acid residues and non-enzyme entities were retained.

In each enzyme structure, water molecules were classified into "partitions" according to their location: those with oxygen < 5 Å from active site atoms (near active site), those with oxygen < 5 Å from any other enzyme atom (in enzyme), and those > 5 Å from any enzyme atom (bulk solvent). For each partition, the electric field magnitudes experienced by water ($E(H_2O)$) were calculated at the oxygen coordinates of each using AMBER99SB-ILDN forcefields.[102, 103] These fields were calculated at the position of each water molecule's central oxygen as an approximation of its rotational center of mass.

Electric Fields on Hydroxyl Bonds During ADMP

Once a minimum electric field required to ionize a water molecule was established, the electric fields experienced by water molecules in the active sites of select enzymes were computed during Atom Centered Density Matrix Propagation (ADMP)[105] calculations at the HF 3-21g* level at 300K. For each of the enzymes chosen, a PDB file was obtained from rcsb.org. The active site of the enzyme structure was then manually identified and a water molecule within 5 Å of at least one catalytic residue was selected. A sphere centered around this water molecule with a ~16 Å diameter was then extracted from the structure and trimmed to a system of ~229 atoms (mean number) using PyMOL. Because the snipping process in PyMOL did not retain the whole residues on the perimeter of the sphere, the missing portions of these residues were either restored or simplified according to their likelihood of electrostatically influencing nearby water. For instance, a missing saturated hydrocarbon chain would be approximated by capping the previously sp^2 carbon on the surface of the active site sphere with hydrogen to become an sp^3

terminal carbon. Partial snippets of uncharged residues extending into the active site sphere were removed unless found within ~5 Å of a water molecule. This was done to maintain the electrostatic profile of the active site while drastically reducing the CPU time of subsequent calculations.

The enzymes whose active sites were used for the computation of fields experienced by water molecules during ADMP calculations were triosephosphate isomerase (PDB 1NEY),[104] MoFe nitrogenase (PDB 7JRF),[106] human carbonic anhydrase II (HCAII) (PDB 4YXI),[107] staphylococcal nuclease (PDB 2SNS),[108] DNA photolyase (PDB 1TEZ),[109] Phosphoribosylamine-glycine ligase (PDB 3MJF),[110] human ubiquitin-conjugating enzyme (PDB 2ESK),[112] horse liver alcohol dehydrogenase (PDB 7RM6),[34] and thioredoxin M1 (PDB 7C65).[113] These active sites, including non-enzyme atoms, contained an average of 229 total atoms and 9 water molecules (table 3). We used a particularly large active site excerpt (402 atoms) for DNA photolyase to examine the water molecules between the enzyme active site residues and the DNA substrate.

PDB ID	TOTAL ATOMS	TOTAL WATER MOLECULES
1NEY	231	8
7JRF	227	15
4YXI	203	7
2SNS	111	9
1TEZ	402	18
3MJF	254	11
2ESK	198	4
7RM6	222	5
7C65	209	4

Table 3. Total atom counts, including water and non-enzyme atoms, and water molecules counts for each enzyme active site fragment subjected to 100 fs ADMP calculations.

Proteins are estimated to primarily possess an internal dielectric constant of $2 \leq \varepsilon \leq 10$, with surface regions as high as $\varepsilon = 30$.[117, 118] We sought to use a solvent with a value of ε near the conservative end of these estimates, and thus chose acetone ($\varepsilon \approx 20.5$).[114-116] This conservative ε was chosen so as not to bias our calculations towards increased charge-charge interactions in favor of our hypothesis that the fields experienced by water are higher in enzyme active sites

than bulk solvent. The active sites were subjected to ADMP over 100 fs of simulation time using the HF 3-21g* basis set. This basis set was chosen for the computational time associated with such large systems—each active site calculation took approximately two weeks to complete with 512 GB of allocated memory.

At each 0.1 fs frame of the ADMP trajectories, the electric fields on the hydrogen belonging to each water molecule were calculated as projected along the unit vector $\widehat{OH}$ of the bond vector $\overrightarrow{OH}$ for each. These fields were calculated as the sum of point charges through Coulomb's law using the Mulliken[119] charges assigned to each atom by Gaussian at the onset of each ADMP calculation. There was no cutoff distance for including charges from other atoms in the simulations—contributions from every atom were included. As a control for the fields experienced by water hydrogen in bulk water, a cubic system with an edge length of 20 Å containing exclusively 96 TIP3P[99] water molecules with implicit solvation in water ($\varepsilon \approx 78.4$)[116] was also subjected to these calculations, including measurement of the electric fields along the hydroxyl bonds of each water molecule. Implicit solvation was carried out using the Polarizable Continuum Model (PCM).[156]

Ionization Under an External Electric Field

We sought to establish the minimum external electric field required to ionize a water molecule using the same quantum mechanical calculation scheme later applied to determine such fields in enzyme active sites. A water molecule was placed in a simulation box such that one of its OH bond vectors was parallel to the standard x-axis. Replicate copies of this molecule were then subjected to increasingly negative external electric fields in time-dependent quantum mechanical calculations using a polarizable Hartree Fock (HF) basis set, HF 3-21g* propagated

via ADMP. Each calculation was conducted for 200 fs simulation time. Initially, the field was increased by 0.01 a.u. between trials. Once a field of sufficient strength to ionize the molecule was found, further trials were conducted wherein the field was decreased by 0.0001 a.u. until the threshold of autoionization, to this precision, was calculated. Lest minor differences in the molecule's geometry upon initialization influence the outcome, this was repeated using a newly initialized water molecule.

Measuring Electric Fields on Water Molecules Near DHAP in TPI

Triosephosphate isomerase (TPI) catalyzes the interconversion of dihydroxy acetone phosphate (DHAP) and D-glyceraldehyde 3-phosphate. The substrate, DHAP, contains a negatively charged phosphate group exposed to bulk solvent in the crystal structure (PDB ID 1NEY).[104] To probe the proclivity of water to be deprotonated near DHAP in the active site, model 1NEY was subjected to 2000 ns of classical MD simulation at a temperature of 300 K. At each 10 ns frame of the subsequent trajectory, water molecules were partitioned into the same groups as done for other XRC structures, apart from the water molecules nominally classified as being near the active site residues, which were instead qualified by being < 5 Å from the DHAP substrate molecule in the TPI active site. This was done because we wanted to explicitly examine the fields on water molecules near DHAP. The electric field experienced by each water hydrogen at each frame was then calculated as projected along the bond shared with its central oxygen using Coulomb's law. Intramolecular point charge contributions were discarded from the calculation to measure exclusively external electric fields on each water hydrogen atom. All charges used for Coulomb's law were obtained via the AMBER99SF-ILDN forcefield as per the analysis of static enzyme structures. To test whether the active site of TPI enhances the electric

fields experienced by water hydrogen atoms near DHAP, the atomic coordinates of DHAP were copied from 1NEY into a new file and solvated by water. This model of DHAP in water was then subjected to the same MD conditions, simulation time, and water hydrogen electric field analysis as for the TPI structure.

All *ab initio* QM calculations were conducted using Gaussian 16.[157] The enzyme structures used for these calculations were modified and visualized using PyMOL and GaussView.[154, 158] MD calculations used for the analysis of TPI and DHAP in water were accomplished using GROMACS.[159]

Results and Discussion

Electric Fields on Water Molecules in Static Enzyme Structures

Among the 1179 < 1.5 Å resolution enzyme XRC models studied, 3146 water molecules were found near active site residues, 180585 were found elsewhere near the enzyme, and 8293 were found > 5 Å away from all enzyme residues in the bulk solvent. The mean electric field experienced by water molecules in the active site was $5 \cdot 10^{-3}$ a.u. higher than that of water elsewhere in the enzyme and $9 \cdot 10^{-3}$ a.u. higher than that of bulk water (Table 4, Figure 17). The finding of heightened fields experienced by water in the active site relative to those elsewhere within the enzyme is in striking corroboration of the electrostatic theory of catalysis, whereby strong fields in the active sites of enzymes facilitate transition state stabilization.

	BULK SOLVENT	NEAR ENZYME	ACTIVE SITE
FIELD STRENGTH (MEAN) / A.U.	$2.2 \cdot 10^{-2}$	$2.6 \cdot 10^{-2}$	$3.1 \cdot 10^{-2}$
FIELD STRENGTH (STD. DEV.) / A.U.	$2.1 \cdot 10^{-2}$	$5.1 \cdot 10^{-2}$	$2.1 \cdot 10^{-2}$
COUNT / MOLECULES	8293	180585	3146

Table 4. Descriptive statistics for electric fields calculated to be experienced by water molecules in different locations of each enzyme model studied. The category of water molecules denoted as “near enzyme” excludes those near the active site of each enzyme.

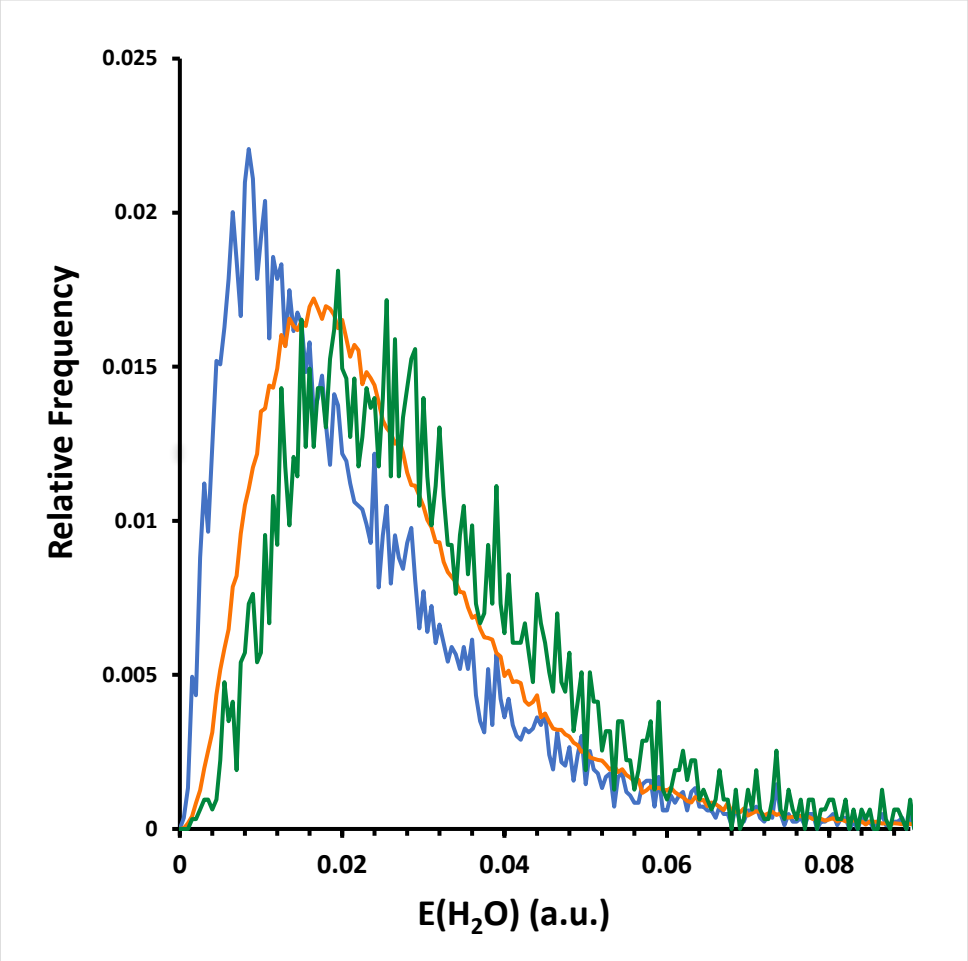

Figure 17. Calculated electric field magnitudes experienced by water molecules at their oxygen coordinates in 1179 enzyme XRC models of < 1.5 Å resolution. Water molecules were partitioned by location with respect to the enzyme and the active site of each model. Water molecules near the active site (green) were classified as being less than 5 Å from any active site atom. Other water molecules within 5 Å of the enzyme but not near the active site were classified as being near the enzyme (orange). All other water molecules were classified as bulk solvent (blue).

While 58.5% of water molecules in bulk solvent experienced electric fields less than 0.02 a.u. in magnitude, only 43.9% and 31.7% of water molecules in each enzyme (excluding the active site) and in the active site of each enzyme experienced electric fields of a magnitude less than this value, respectively (table 5). Noticeable differences could be observed between water partitions at higher field strengths as well. 12.0% of bulk water molecules were found to incur

electric fields stronger than 0.04 a.u.. For water molecules within the active site of each structure and elsewhere within each, 14.6% and 22.8% were found to incur electric fields stronger than 0.04 a.u., respectively.

RANGE (A.U.)	BULK (%)	ENZYME (%)	ACTIVE SITE (%)
0.0-0.02	58.5	43.9	31.7
0.02-0.04	29.5	41.6	45.6
0.04-0.06	8.4	10.8	16.5
0.06-0.08	2.2	2.5	4.1
0.08-0.1	0.7	0.7	1.2
> 0.1	0.7	0.6	1.0

Table 5. Percentages of each partition of water molecules within 1179 enzyme structures experiencing electric fields within given ranges of magnitude.

Electric Fields on Water Hydroxyl Bonds in TPI

During 2000 ns of MD simulation of TPI structure 1NEY, water hydrogen within 5 Å of non-active site residues (H_{enz}) were calculated to incur only slightly greater fields, projected along $\widehat{OH}$, when compared to those in the bulk solvent around the enzyme (H_{bulk}) (fig. 18). This presents a curious discrepancy from the much greater fields experienced by whole water molecules near enzyme residues compared to those in bulk solvent among the 1179 static XRC structures studied (fig. 17). However, fields on water hydrogen < 5 Å from DHAP in the TPI active site (H_{DHAP}) were calculated to be significantly greater than both H_{enz} and H_{bulk} (fig. 18). A distinguishing feature in the distribution of H_{DHAP} fields was observed at around 0.07 a.u. electric field strength. Unsurprisingly, these high incurred fields corresponded with water hydrogen near the electronegative phosphate group on DHAP (fig. 19).

During 2000 ns of simulation of DHAP in bulk water, the fields incurred by H_{DHAP} were found to be significantly lower than H_{DHAP} in TPI (fig. 20). The peak around 0.07 a.u. in the latter distribution is remarkably not present in the former, implicating a population of H_{DHAP}

experiencing massive electric fields in the active site of TPI which was drastically diminished among H_{DHAP} in bulk water. We initially inculpated the disproportionate exposure of the DHAP phosphate group to water in TPI for this observation. This explanation was negated, however, upon observing that much of the DHAP molecule is indeed accessible when the inhibitor is in the enzyme-substrate complex, excluding the terminal hydroxyl group opposite the phosphate. Accordingly, the enhanced fields incurred by H_{DHAP} in TPI relative to H_{DHAP} in bulk water indicate that the electrostatic environment of the TPI active site enhances fields favoring the deprotonation of water molecules near the phosphate head group of DHAP.

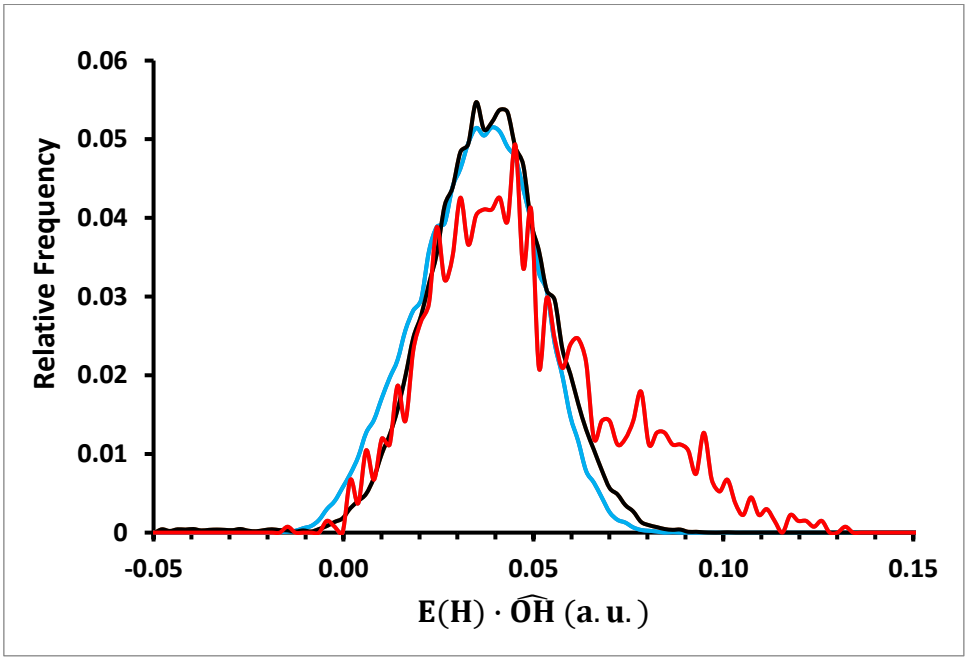

Figure 18. Histograms of electric fields experienced by water hydrogen over 2000 ns of MD simulation time in the triose phosphate isomerase PDB model 1NEY. Water hydrogens were classified as being part of the bulk solvent surrounding the enzyme (blue), in the enzyme but not near the substrate (black), or near the substrate, DHAP (red). Each data point represents a given hydrogen's incurred field at a given frame of the trajectory. Field strengths were sampled at intervals of 10 ns during the simulated trajectory. The field at each water hydrogen was calculated as projected along $\widehat{OH}$ and excluded charge contributions from the central oxygen and secondary hydrogen. Bin size=0.01 a.u.. Relative frequency was calculated as the number of water hydrogen experiencing fields within a bin divided by the product of the number of water hydrogen and sample frames.

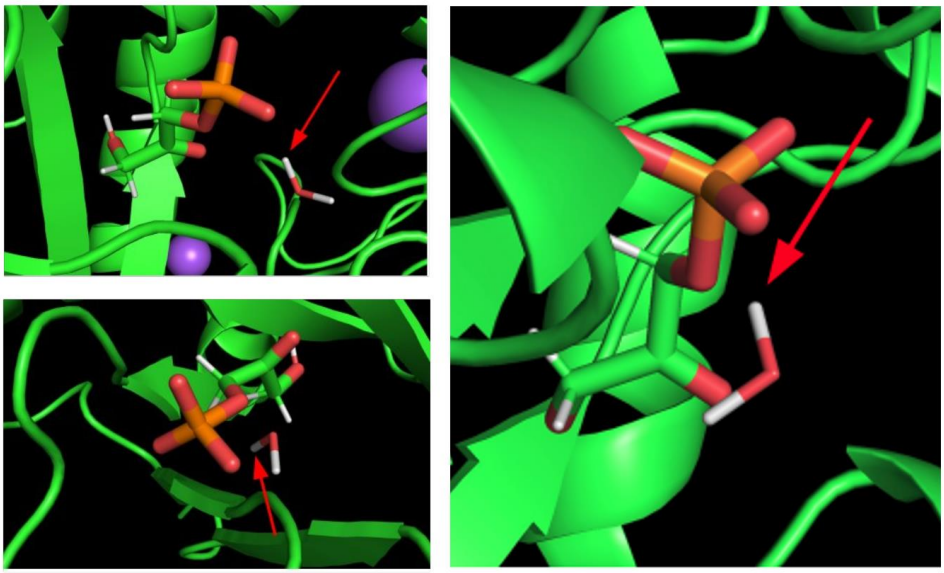

Figure 19. Water molecules (stick representation, indicated by red arrows) whose hydrogen experienced the highest electric fields during 2000 ns of MD simulation of TPI (PDB structure 1NEY). The green cartoon depicts the active site of TPI. The substrate DHAP can be seen in stick representation, with the phosphate group shown as an orange phosphorous atom surrounded by four red oxygen atoms, attached to the green carbon backbone of the molecule. Purple spheres depict Na^+ ions in the model.

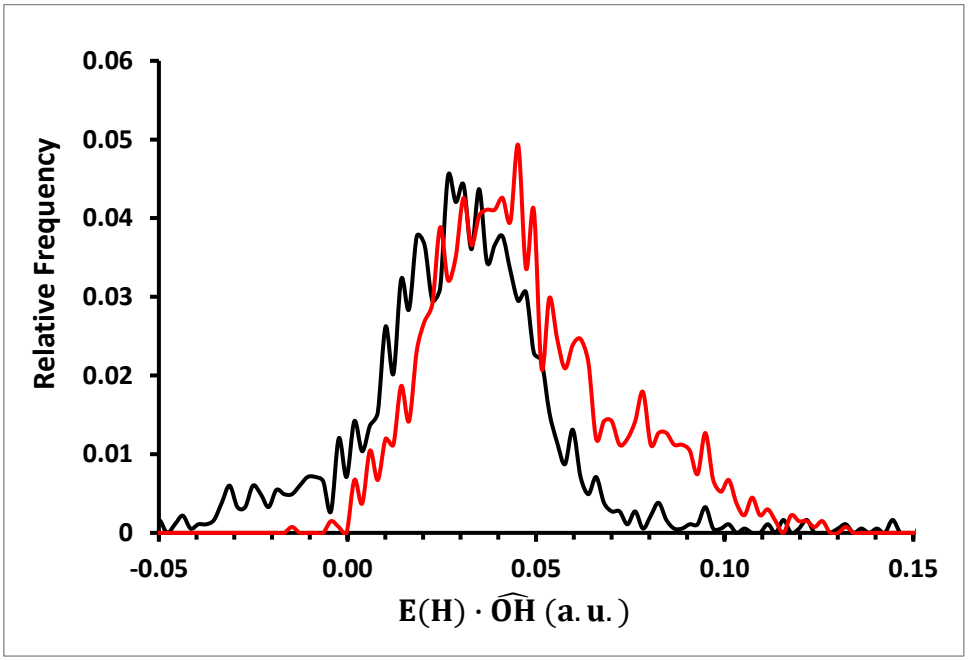

Figure 20. Sampled at 10 ns simulation time intervals over 2000 ns of MD simulation time, fields experienced by water hydrogen within 5 Å of DHAP in TPI (red) and within 5 Å of DHAP in bulk water (black). The field at each water hydrogen was calculated as projected along $\widehat{OH}$ and excluded charge contributions from the central oxygen and secondary hydrogen. Bin size=0.01 a.u.. Relative frequency was calculated as the number of water hydrogen experiencing fields within a bin divided by the product of the number of water hydrogen and sample frames.

It is mentionable that there is quality, yet indefinite, research indicating that the closure of loop 6 in TPI displaces water molecules from the active site.[160-162] Our work utilized a structure of TPI in which this loop was in an open conformation, permitting solvent. As such, we do not intend our results regarding the fields experienced by H_{DHAP} in the active site of TPI to support or contradict the hypothesis that the closure of loop 6 excludes solvent. Rather, these results serve at the most basic level to demonstrate the heightened electric fields incurred by water molecules near DHAP in the active site of the enzyme. In correspondence with this, our results show that the likelihood of water's ionization near DHAP in the active site of TPI is increased relative to that of water in bulk, water molecules elsewhere within TPI, and water molecules near DHAP in bulk.

Electric Fields on Water Hydroxyl Bonds in Active Sites During ADMP

To measure the fields along the hydroxyl bonds of water molecules within enzyme active sites during molecular motion simulated at a higher level of theory than classical mechanics, *ab initio* QM calculations were applied to a box of 96 water molecules and individual eight active sites containing using ADMP and the HF 3-21G* basis set. The simulations were carried out for 100 fs simulation time. Every 0.1 fs, the fields experienced by each water hydrogen projected along its corresponding hydroxyl bond, $E(H) \cdot \widehat{OH}$, were calculated in each trajectory.

Our analysis of these electric fields revealed a notable difference in $E(H) \cdot \widehat{OH}$ calculated during these simulations between the bulk water and the active sites (fig. 21). Specifically, we observed a broader distribution in $E(H) \cdot \widehat{OH}$ within the active sites (σ=0.038 a.u.) relative to bulk water (σ=0.025 a.u.), indicating a propensity toward ionization of water molecules. This breadth further suggests that, as experienced by water molecule hydrogen about their hydroxyl bonds, the electrostatic environments generated by enzyme active sites are highly heterogenous relative to bulk solvent.

30% of $E(H) \cdot \widehat{OH}$ in the active sites and 8% in bulk water were less than 0 a.u. during ADMP. Because of the ~104.5° bond angle in water, a strong electric field directly pulling one hydrogen along hydroxyl bond axis would have the opposite effect on other bonded hydrogen—pulling it towards their shared oxygen. Thus, these unique fields observed during active site simulations were in part due to electronegative amino acid groups opposite hydroxyl bonds on water molecules. Another phenomenon responsible was the momentary orientations of water molecules such that their hydroxyl bonds faced positively-charge species during the simulation. Two respective examples of this were found in HCAII, where a water molecule was positioned

such that the hydroxyl bond was directed away from an electronegative imidazole nitrogen with a water oxygen-nitrogen distance of ~2 Å, yielding $E(H) \cdot \widehat{OH}$ of ca. -0.05 a.u.; and in DNA photolyase, where a water molecule associating with a ligated Mg^{2+} cation briefly oriented with hydroxyl bonds facing the cation, eliciting $E(H) \cdot \widehat{OH}$ of ca. -0.06 a.u.. Because of the negative $E(H) \cdot \widehat{OH}$ values corresponding with such events in the active sites, the mean $E(H) \cdot \widehat{OH}$ of 0.022 a.u. within these active sites during ADMP was much lower than that of the bulk solvent, 0.034 a.u..

Despite the increased frequency of negative $E(H) \cdot \widehat{OH}$ in the active sites relative to bulk solvent, there was another remarkable discrepancy in these two distributions. was the presence of fields exceeding ~0.15 a.u. in a small number of instances within enzyme active sites, which were not present during the ADMP trajectory of bulk water molecules. The significance of this is discussed below with respect to the results of deprotonating water under an applied external electric field.

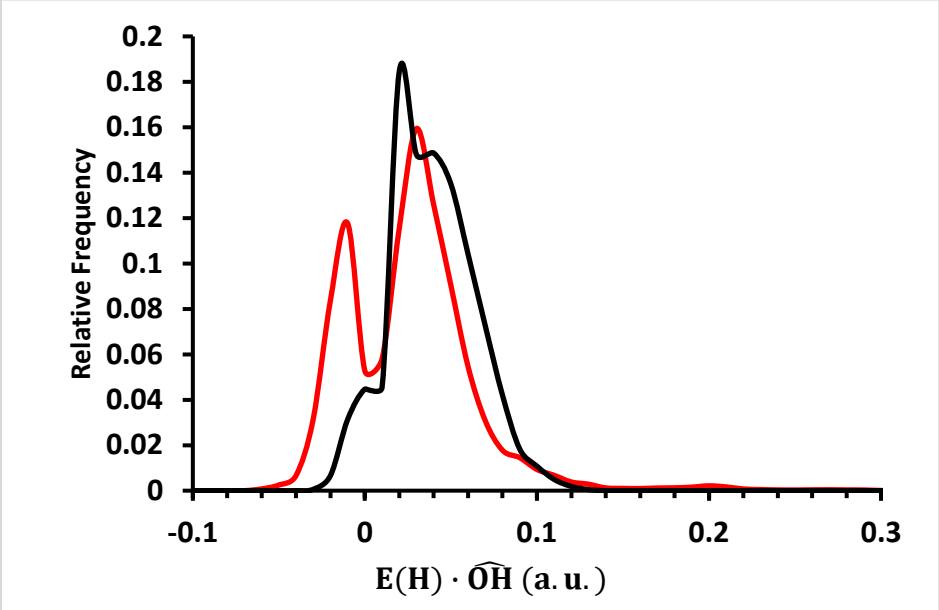

Figure 21. Sampled at 0.1 fs simulation time intervals, distributions of electric fields experienced by water hydrogen projected along $\widehat{OH}$ among water molecules within a model of 96 water molecules (black) and 9 different active sites (red) during 100 fs ADMP. Each data point is the electric field experienced by a water single hydrogen at a frame of the simulation. Bin size=0.01 a.u.. Relative frequency was calculated as the number of water hydrogen experiencing fields within a bin divided by the product of the number of water hydrogen and sample frames.

Hydroxyl Bond Lengths in ADMP Trajectories

Nuanced but meaningful differences existed in water hydroxyl bond lengths between the bulk water and active sites during 100 fs ADMP (fig. 22). While the means of these bond lengths were nearly identical—0.988 Å in the active sites and 0.982 Å in bulk, the standard deviations of hydroxyl bond lengths were 0.080 Å in the enzyme active sites and 0.036 Å for bulk. No bond lengths were at any simulation frame less than 0.9 Å in bulk water. 0.26% of bond lengths were found between 0.78 Å and 0.9 Å in the active sites.

Moreover, no hydroxyl bond length in bulk water was at any moment greater than 1.15 Å, corresponding with an expected lack of ionizing events in the bulk water simulation. In the active

site simulations, 3% of hydroxyl bonds were longer than 1.15 Å across all simulation frames, with a maximum distance of 1.75 Å. These belonged to three water molecules in the active site of TPI, two in staph nuclease, and two in HCAII. In TPI, one of these corresponded with hydrogen bonding with Glu165, one associated with the phosphate group of DHAP, and one with another water molecule on the periphery of the simulation. In staph nuclease, both hydroxyl bonds longer than 1.15 Å were found to be engaged in hydrogen bonding with separate carboxyl oxygen of Glu43. In HCAII, these bonds were correlated with a water molecules hydrogen bonding with the amide nitrogen linking Phe95 and His94, and one interacting with the phenyl alcohol of Tyr7, discussed later. Observation of these phenomena within a short 100 fs time period in two of eight active site simulations is highly intriguing and implicates the capacity of charged residues and non-enzyme molecules to engage in hydrogen bonding and acid-base chemistry with water molecules.

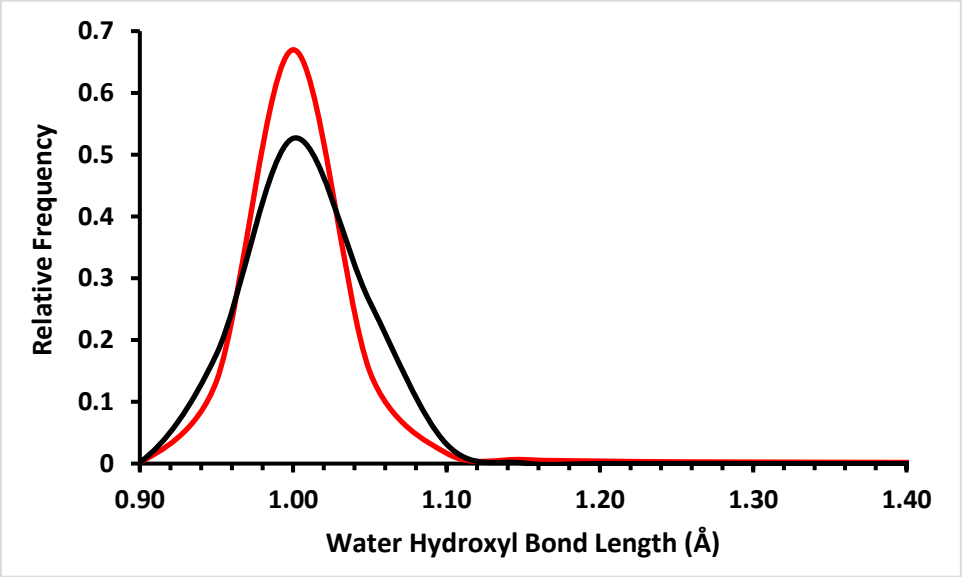

Figure 22. Smoothed histogram comparing water oxygen-hydrogen distances among eight different enzyme active sites (red) with those in a simulation box containing 96 water molecules (black) during 100 fs ADMP calculations at a sample rate of 10 fs^{-1}. Bin sizes are 0.05 Å.

External Electric Field Applied to Water During ADMP

Through the application of varied external electric fields perpendicular to the hydroxyl bond of a single water molecule during 100 fs ADMP treated with HF 3-21g* at 300 K (nuclear kinetic energy: 1400 microHartrees), we found that an external field of approximately 0.1561 a.u. is required to deprotonate insular water molecules, which occurred at 74 fs in the simulation. This was corroborated by trials with applied electric fields bordering this magnitude. For example, at 0.1565 a.u. external electric field strength, the water molecule was deprotonated in 50 fs simulation time. After observing no deprotonation at 0.1560 a.u. external electric field strength, the trial was extended to 200 fs. Deprotonation subsequently occurred at 110 fs with a 0.1560 a.u. external electric field magnitude. For these simulations, deprotonation was defined as the water hydrogen monotonically moving away from the central oxygen of the molecule in the direction of the fictitious source of the negative external electric field after reaching a hydroxyl bond distance of 2 Å. Repeating this test with implicit solvation in water yielded extremely similar results, corresponding with a dielectric-invariant external electric field effects in the Gaussian software. Thus, the water molecule studied was simulated in free space in all trials. Each ADMP calculation was ran in duplicate with a new-initialized water molecule, yielding identical results.

SCF Scans of Hydroxyl Bonds

Seeking to further validate the 0.1561 a.u. external electric field magnitude required to deprotonate a water molecule in under 100 fs, we conducted a series of SCF scans along the hydroxyl bond of this water molecule under varying external electric fields, treated with HF 3-21G* for consistency with our ADMP calculations. In each, the energy of the hydroxyl bond

could be plotted with respect to its length, and the activation energy of the bond breaking was calculated by subtracting the energy minimum preceding the energy maximum (transition state) from the energy maximum (fig. 23). Energies were scanned at 0.05 Å increments of the bond length from 0.96 Å to 3.46 Å, but in many cases, the scan could not be completed for large distances, leading to truncation of this range. This was not a significant error in that the ground state and transition state energies of the hydroxyl bond breakage were visually validated to be captured in each scan range. Scans were conducted at field strengths incremented by 0.01 a.u. from 0.01 a.u. to 0.14 a.u.. We were able to generate a tightly correlated (R^2=0.99) 3rd order polynomial equation of the activation energy of the hydroxyl bond as a function of external electric field strength (fig. 24). Per this equation, the external electric field magnitude at which the activation energy of a water hydroxyl bond is 0 Hartrees was extrapolated to be 0.1350 a.u.. This represents the electric field needed to be experienced by a water hydrogen molecule projected along its hydroxyl bond in order for deprotonation to be energetically neutral, with any value exceeding this making it favorable. The reason for this approach in supplement to ADMP runs of water exposed to varying external electric fields is that the result is not time dependent—0.1350 a.u. is not the field strength required for deprotonation within 100 fs as per the ADMP trials, it is the minimum field strength required for deprotonation over any period of time.

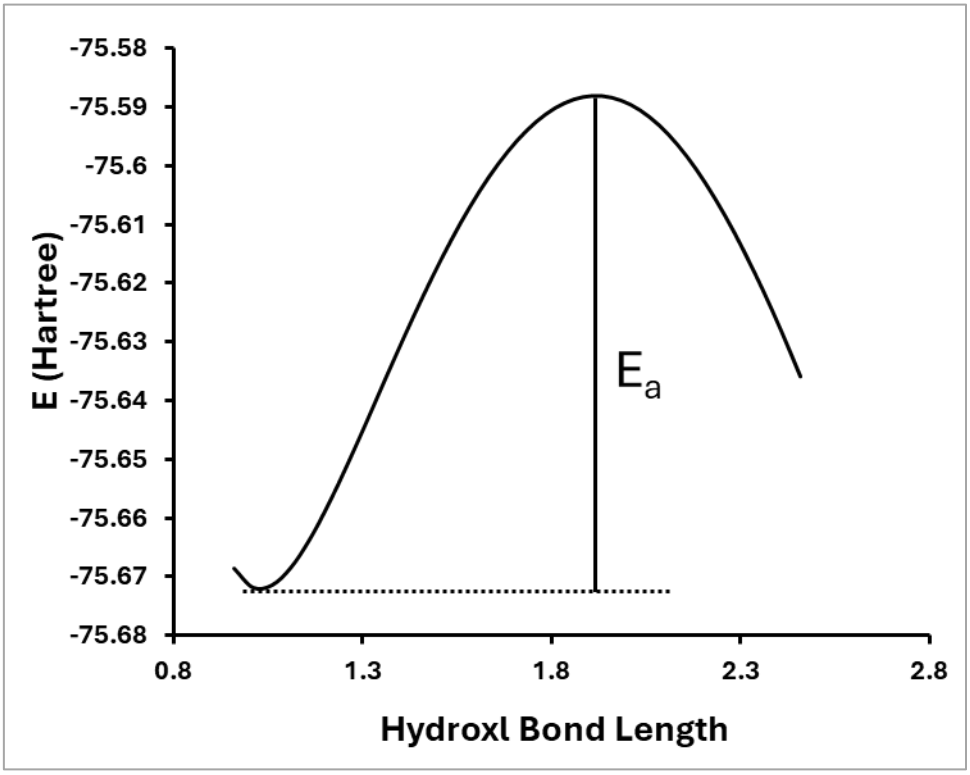

Figure 23. An example of the energy of the hydroxyl bond of a water molecule with respect to its length under an external electric field of 0.9 a.u.. along $\overrightarrow{OH}$. The activation energy of this bond is shown as E_a and was calculated by subtracting the energy minimum preceding the energy maximum from the energy maximum.

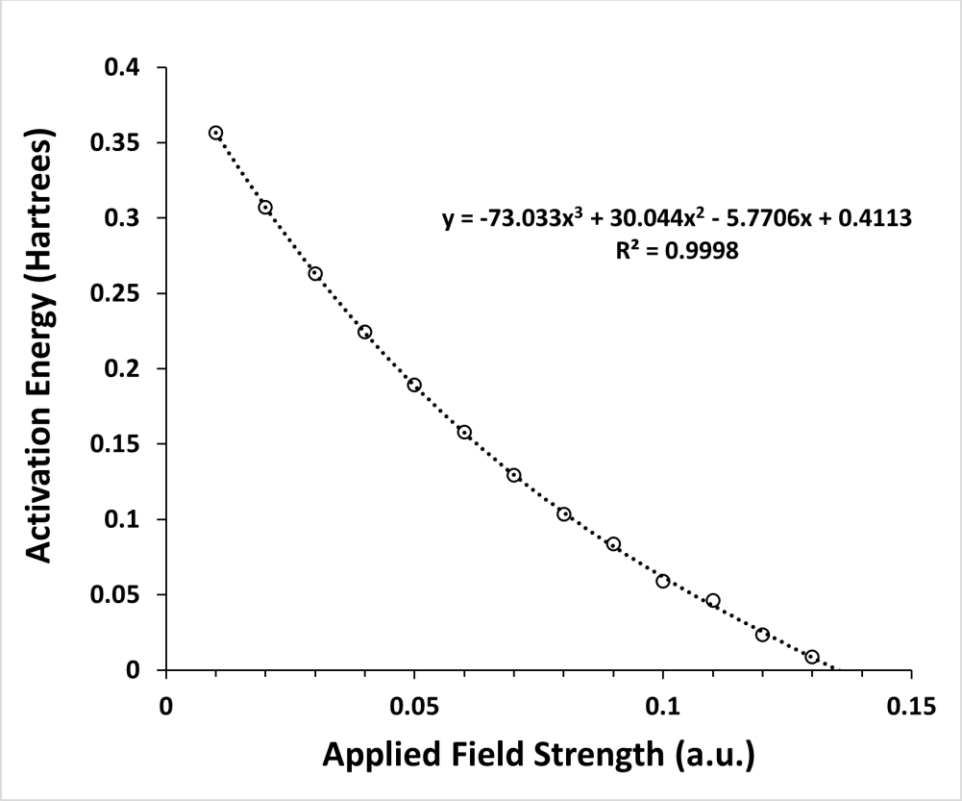

Figure 24. A plot of the activation energy of the hydroxyl bond of a water molecule subjected to various external electric field strengths along the hydroxyl bond axis, obtained by SCF of the hydroxyl bond at each field magnitude. A scan was conducted for 0.14 a.u. electric field magnitude but was not included in this plot on account of its hydroxyl bond activation energy of 0 a.u. introducing bias into the polynomial regression.

Comparison with Active Site ADMP Results

These results validate water ionization events which occurred within two of the eight active site simulations during the 100 fs that were simulated, both of which corresponded with water hydrogen experiencing electric fields greater than 0.1561 a.u.. These were within the active sites were of TPI and HCAII. In TPI, hydrogen bonding between the amide nitrogen linking Ser96 and His95, Wat5014, and the non-reactive carboxylate oxygen of Glu165 led to the deprotonation of the amide nitrogen, forming a transient hydronium on Wat5014, which was neutralized by protonating Glu165. The $E(H) \cdot \widehat{OH}$ value of this hydrogen reached a massive 0.3

a.u. at this point. The hydrogen transferred to Glu165 was bonded to Wat5014 at the onset of the simulation, rather than transferring from the Ser96 amide before protonating Glu165.

In HCAII, the alcohol on the phenol hydroxyl group of Tyr7 participated in hydrogen bonding with Wat443, whose own hydrogen then engaged in apparent hydrogen bonding with the Zn^{2+}-bound His96 imidazole nitrogen. This hydrogen atom momentarily incurred $E(H) \cdot \widehat{OH}$ of 0.28 during this interaction. The observation of these events within a mere 100 fs in two of eight active sites is remarkable in light of the 0.1 hr^{-1} ionization rate of bulk water. These water hydrogen of interest all exceeded $E(H) \cdot \widehat{OH}$ of 0.1350 a.u. in both active sites, corroborating the value obtained from SCF data of the hydroxyl bond.

Conclusion

The findings presented in this study provide compelling evidence supporting the hypothesis that electric fields within enzyme active sites induce a favorable electrostatic environment for the ionization of water molecules. This ionization leads to the generation of water ions that actively participate in biocatalysis. Through a comprehensive analysis of X-ray crystallography enzyme structures, we have demonstrated that water molecules near enzyme active sites generally experience significantly higher electric fields compared to those in bulk solvent and elsewhere within enzymes. By simulating the entirety of TPI using classical MD, we have found that the electric fields experienced by water hydrogen along their hydroxyl bonds near the substrate DHAP are significantly greater than those both the bulk solvent and elsewhere within the enzyme. Moreover, we have shown that the active site of TPI enhances the fields on these hydroxyl bonds—evidenced by increased field strengths incurred by water hydroxyl bonds near DHAP in the active site of TPI relative to those near DHAP simulated in bulk solvent. We have

additionally found that the electric field magnitude required to deprotonate water at the level of HF 3-21G* is ~0.1561 within 100 fs simulation time and calculated to be 0.1350 a.u. independent of simulation time. We subsequently found that water hydroxyl bonds in two of eight enzyme active sites, initialized from their crystal structures, experienced fields greater than these magnitudes during 100 fs ADMP, corresponding with ionization events within an extremely short timeframe.

One disadvantage of our approach is the calculation of electric fields using atomic charges assigned by AMBER99SB-ILDN and Mulliken charges, which limits the quantitative accuracy of the fields due to the inherent drawbacks of each method. While AMBER99SB-ILDN is optimized for enzyme calculations, it does not consider local polarization effects on a case-by-case basis when assigning atomic charges.[103] This limitation precludes the calculation of quantitatively exact electric fields in both immobile enzyme structures and those calculated from the TPI trajectory obtained via MD. Mulliken charges, when used with a polarizable basis set, indirectly include polarization in the assignment of atomic charges but not as effectively as other available methods. Generally, Mulliken populations are known to produce workable but unphysical charges.[163] Furthermore, in the studies of electric fields on water molecules from the trajectory frames of MD and ADMP calculations, recalculating the charges on each atom in response to changes in the electrostatic environment at each time step would have yielded more accurate reported field magnitudes. Despite these sources of error, they are consistent within this work. Thus, the comparisons of electric field magnitudes experienced by water molecules and water hydrogen atoms between regions of enzyme structures and bulk water presented here are reliable.

Our research has shed light on the importance of considering the ionization of water and the presence of water ions in enzyme research endeavors. By demonstrating that a) the electric field magnitudes incurred by water molecules are significantly greater in static enzyme structures' active sites than both bulk solvent and elsewhere within enzymes, b) the active site of TPI enhances the electric fields experienced by water hydrogen near DHAP, and c) water O-H bond lengths and water hydrogen-incurred electric fields deviate strongly in enzyme active sites relative to bulk water during ab initio quantum mechanics, we have presented a case for the favorability of water's ionization in enzyme active sites.

These findings have profound implications for our understanding of enzyme function and the manipulation of biocatalytic processes. Moving forward, further investigation into the role of water ions in enzyme catalysis is warranted. Future research efforts in this regard should focus on illuminating the specific mechanisms by which water ions contribute to enzyme reactions and exploring potential applications in biotechnology and drug discovery. By continuing to unravel the intricate interplay between electrostatic forces and water molecules in enzyme active sites, we can unlock new avenues for the design and optimization of enzyme catalysis with enhanced efficiency and specificity.

CHAPTER FIVE

CONCLUSION

Limitations and Improvements

While this research provides compelling evidence supporting the role of water as an essential cofactor in enzyme catalysis, there are areas where the study could be improved:

Hydroxyl Bond Electric Field Calculations

During *ab initio* quantum MD simulations, the electric fields were calculated along the hydroxyl (OH) bonds of water molecules within enzyme active sites. This was done as a metric of the proclivity for these bonds to break, leading to the formation of hydroxide. What this method neglects, however, is measurement of the proclivity for water to produce the hydronium ion. Per the scope of this work, when a water molecule is deprotonated in a simulation, the identity of the nucleophile involved is unknown unless visually observed in the simulation trajectory.

To quantitatively measure the proclivity for water to form hydronium during molecular dynamics on the basis of electrostatics calculations, the electric field generated by water oxygen on nearby hydrogen within enzyme active sites would need to be calculated. This would be a much more complex undertaking than measuring the fields favoring deprotonation because the electrophilic hydrogen atoms nearby each water oxygen would need to be identified, the strengths of their bonds to other atoms would need to be calculated, and the angle at which they may approach nucleophilic water would need to be calculated. The analysis of hydronium ion formation could involve supplementing electrostatics calculations with an analysis of orbital

overlap between the oxygen of water molecules and nearby electrophilic hydrogen atoms in different active site conformations. Sufficient overlap between the highest occupied molecular orbital (HOMO) of the water oxygen and the lowest unoccupied molecular orbital (LUMO) of the electrophilic hydrogen would be indicative of a high likelihood of bond formation.[xii]

Basis Sets

The HF 3-21G* basis set is a split-valence Gaussian extension of the 3-21G basis set with added polarization functions. HF 3-21G* has several flaws and was used for the ADMP calculations of enzyme active sites exclusively for its reduced computational time relative to other basis sets.[164] This basis set lacks diffuse functions, which are important for describing electron density far from the nucleus. As this research examined the electrostatic behavior of systems during ADMP, such a lack of accurately modeled long-range electrostatic interactions may have led to an underrepresentation of ionizing events and hydrogen bonding involving water molecules. For more accurate and reliable quantum chemistry calculations, especially for electrostatics using MPA, larger and more sophisticated basis sets would be better choices than the HF 3-21G* basis set.

One such option is HF 6-31G*. This basis set is another split-valence basis set, but with three more core Gaussian primitives and one more primitive in the first of the two valence sets, providing a more detailed description of the electron distribution.[165] The 6-31G* basis set includes polarization functions, which allow the electron cloud to adapt to molecular geometries more accurately, improving the reliability of electrostatic potential calculations.

[xii] Acknowledgment: some of these ideas regarding characterizing hydronium formation came about in conversation with Dr. Shawn Kathmann of PNNL.

Another superior option is the cc-pVDZ (correlation-consistent polarized valence double-zeta) basis set.[166] This basis set is designed to systematically converge towards the complete basis set limit, meaning it is constructed to improve the accuracy of correlation energy calculations. The cc-pVDZ basis set includes multiple polarization functions, providing a highly flexible description of the electron density and making it particularly suitable for correlated methods like MP2 or coupled-cluster calculations, which can improve the accuracy of electrostatic interactions within systems, as well as MPA charge assignment.

For even greater accuracy of ADMP calculations, especially for properties like electrostatic potentials, the aug-cc-pVDZ basis set is a better choice. This basis set extends cc-pVDZ by adding diffuse functions, remedying the lack of long range-electrostatics interactions with HF 3-21G*.[166-168] However, employing aug-cc-pVDZ on systems of several hundred atoms, as done for the active sites in our study, would likely be unfeasible for many years to come on account of computational time requirements.

MPA and AMBER99SB-ILDN

To calculate the electric fields on water molecules during quantum and classical MD simulations, charges were obtained via MPA and the AMBER99SB-ILDN forcefield, respectively. These methods have several limitations when it comes to accurately assigning atomic charges, resulting in inaccurate calculations of electric fields.

MPA is highly sensitive to the basis set used in quantum chemical calculations, and the HF 3-21G* basis set employed during ADMP has several aforementioned flaws. The dependence of MPA on the basis set used leads to significant variations in the calculated charge distributions, making the results less reliable. Moreover, Mulliken charges often do not accurately represent

the electrostatic potential around molecules, particularly for systems with delocalized electrons or significant polarization effects. This can result in an inaccurate depiction of electrostatic interactions, which are critical for MD simulations. While natural population analysis (NPA) has limitations, it would be useful to repeat electrostatics calculations with charges obtained by this method.[163]

The AMBER99SB-ILDN force field, while widely used and respected for its balance of computational efficiency and accuracy in simulating biomolecular systems, has its own set of limitations. One issue is that it relies on fixed atomic charges, which do not adapt to the changing electronic environment during MD. This static nature can lead to inaccuracies in the representation of dynamic electrostatic interactions.

Lack of Experimental Validation

While robust, the methods employed in this project were purely computational. Offering some component of experimental empiricism, all enzyme structures used were resolved experimentally via XRC. Accordingly, the topological component of this research is metanalytical in that it gathers information from a host of experimental findings reported by other research groups. The electrostatics component of this work, however, strays further into the realm of theory through the application of molecular dynamics, the use of charges derived from MPA & AMBER99SB-ILDN, implicit solvation, and many other computational methods. Naturally, experimental validation would significantly enhance confidence in any and all results reported in this dissertation. Suggested means of experimental validation are discussed later in this chapter.

Timescales of ADMP Calculations

The molecular dynamics simulations used for electrostatics calculations, while informative, could benefit from longer time scales. The simulations of enzyme active sites via quantum MD were carried out over 100 fs of simulation time. Ufimtsev *et al*. found in their quantum MD simulation of BPTI protein an electronic equilibrium between the enzyme and bulk solvent after 8.8 ps, 88 times the simulation times used for enzyme active sites in this work.[81] Longer simulation times would allow for sampling electric fields in a more diverse range of molecular configurations. While the ADMP approach was used to capture these fields while the molecular systems were in motion, it was hoped, but not expected, that the active sites' trajectories would trend towards equilibrium. This equilibrium would enable the sampling of electric fields at near-optimized molecular configurations at 300 K. Intuitively, it is highly unlikely that 100 fs time is sufficient for 100 fs ADMP simulations to approach an optimized state. To calculate electric fields incurred by water at ideal molecular configurations within enzyme active site, optimization calculations or (speculatively) much longer ADMP calculations would be needed.

Before it was decided that implicit solvation of the systems would yield more accurate descriptions of their motion and MPA-assigned charges during ADMP, each was subjected to ADMP without implicit solvation. The data from these calculations was unreported, but valuable insight into the computational time added to each calculation upon the introduction of implicit solvation was gleaned. When implicit solvation was used for ADMP via PCM, yielding the data reported for the electrostatics portion of this study, it was observed that the use of PCM approximately doubled the computational time of each ADMP calculation. It is likely that the application of "Our own N-layered Integrated Molecular Orbital and Molecular Mechanics" (ONIOM)[169] instead of implicit solvation would likely allow for longer simulation times at less

computational cost. ONIOM allows for classical treatment of select molecules and atoms while applying quantum mechanical treatment to others. Because classical MD simulations are significantly faster than quantum MD, the number of atoms simulated per unit simulation time with fixed available computer memory could be increased with ONIOM. This would likely only be a faster approach were the immediate environment around the active sites to be treated classically with ONIOM and the rest of the enzyme being excluded from the system. However, if more realistic depictions of the enzyme systems were desired at the sacrifice of computational time, ONIOM could be used to simulate entire enzymes, with only their active sites treated quantum mechanically.[170]

Accuracy of Implicit Solvation

Capitalizing on the ONIOM method for our ADMP calculations would allow for the simulation of whole enzymes while retaining accurate depiction of their active sites. This would negate the need for implicit solvation, as the dielectric of the enzyme around each active site would be emulated explicitly by the surrounding enzyme atoms. Implicit solvation was an imperfect component of the ADMP calculations of enzyme active sites.[xiii] While the PCM method of implicit solvation allows one to approximate the dielectric environment of a system, it is applied homogenously—the trajectories of the atoms within the active sites were effectively calculated as if the dielectric ghost of acetone were omnipresent throughout the system.[156] Not a single one of the active sites simulated in this study are solvated by acetone in nature!

[xiii] Implicit solvation in water was also used for the ADMP calculation of liquid water.

The dielectric constants of enzymes are heterogenous and complex.[117] Only through the simulation of whole enzymes could the dielectric environment be earnestly approximated.

Addressing Detracting Research

In exploring the essential role of water in enzymatic function, significant research has shown enzymes maintaining activity in non-aqueous environments, thereby challenging our hypothesis. Generally, these enzymes show reduced activity outside of water, but some can function nevertheless.[171] For example, Rogers and Seddon demonstrated that enzymes can thrive in ionic liquids, environments traditionally devoid of bulk water but capable of stabilizing enzyme structures and enhancing catalytic efficiency.[172]

The studies demonstrating enzyme activity in non-aqueous solvents, such as organic solvents and ionic liquids, raise intriguing questions about the role of water in enzymatic catalysis. While these solvents lack the bulk water traditionally considered essential, they may possess properties that emulate key aspects of water's influence on enzyme function. For instance, organic solvents can provide a hydrophobic environment that may stabilize enzyme structures akin to hydrophobic interactions observed in aqueous environments. Similarly, ionic liquids, despite their low water content, can exhibit unique solvation properties that mimic certain aspects of hydrogen bonding and electrostatic interactions crucial for substrate binding and catalytic activity.

However, it is essential to recognize that these solvent properties, while enabling enzymatic activity, do not fully replicate the multifaceted roles of water in biological systems. Water not only serves as a solvent but also participates in enzyme-substrate interactions through hydrogen bonding, mediates conformational changes crucial for catalysis, and facilitates proton transfer

reactions. Therefore, while enzymes can adapt to non-aqueous environments under specific conditions, the distinct biochemical roles of water in enzyme function underscore its irreplaceability in biological catalysis.

Computational and Experimental Validation

This research provides compelling evidence that water plays an indispensable chemical role in enzyme catalysis. The consistent presence of water near active sites, the enhanced electric fields experienced by these water molecules, and their dynamic behavior all point to a fundamental integration of water into the fabric of enzymatic activity. Nevertheless, the computational nature of this work merely opens an avenue for further exploration of our hypothesis. Future steps must be taken for our theoretical implications to form the framework of a widely accepted theory.

Characterizing Channels Accessible to Water with CAVER

The findings presented in this dissertation open several avenues for future research. One promising direction is the use of CAVER, a software tool that allows for the analysis of solvent accessibility within enzyme structures.[148] CAVER can be employed to test whether molecules of user-defined dimensions can access specific locations within enzyme structures. This capability could be leveraged to confirm the solvent accessibility of enzyme active sites, specifically by quantifying the number of enzyme structures where active sites are contiguous with bulk solvent via water clusters.

Additionally, CAVER is capable of analyzing these pathways of solvent access during classical MD simulations. Combining CAVER analysis with MD simulations could provide a

more dynamic picture of solvent accessibility and its impact on enzyme function. This combination would allow for visualizing the amount of time that clusters of water molecules maintain contiguity between active sites and bulk solvent. More importantly, it would facilitate the determination of the duration that water molecules linger in enzyme active sites. Longer residence times would correspond to tighter binding of water molecules, while shorter residence times would indicate rapid exchange of solvent molecules. This rapid exchange might suggest a role for water as a transient cofactor, essential for the catalytic mechanism but not permanently bound. By providing a detailed temporal and spatial analysis of solvent pathways and water dynamics, this integrated approach enhances our understanding of the functional role of water in enzymatic processes.

Hydrogen-Deuterium Exchange

Hydrogen-Deuterium Exchange (HDX), in the context of biology, is the process by which deuterated solvent exchanges deuterium with the amide nitrogen backbone of a protein.[173] When coupled with Mass Spectrometry (MS), HDX-MS is a useful tool for mapping the structure of proteins. After placing the protein in heavy water, the mass analysis of fragments of the protein allows for the determination of which amino acid residues are more solvent-exposed than others on the basis of their deuterium uptake; the deuterium content of each the fragment belies extent to which it has interacted with the bulk solvent.[174]

In this regard, HDX-MS offers perhaps the best experimental method of validating our findings of active site-contiguous clusters of water in the majority of enzymes. While it would be extremely costly and time consuming, each of the 1013 enzymes studied in our water topology work could be expressed, purified, and subjected to HDX-MS. If high levels of exchange are

observed in the active sites of the enzymes after short periods of exposure to deuterated water, it would demonstrate that the solvent readily accesses the active site of each. Shorter enzyme incubation periods in the presence of their substrates may also correspond with functional criticality of solvent intake; if the intake rate of deuterium corresponds with the catalytic rate of each enzyme, it would offer exceptionally strong evidence for the necessity of water in catalysis.

Deuterium Exchange with Substrates and Catalytic Bases

Two specific examples of the numerous manners through which water could be involved in enzyme catalysis (and go thus far unnoticed) pertain to enzyme-catalyzed reactions involving the protonation or deprotonation of the substrate. In a reaction involving the protonation of the substrate, it is possible to envision the foremost protonation of a nearby water molecule, forming a transient and highly acidic hydronium which subsequently protonates the substrate. In the latter example, rather than direct deprotonation of the substrate by a catalytic base, the catalytic base deprotonates a water molecule, forming a transient and highly basic hydroxide that in turn deprotonates the substrate. Both of these scenarios can be tested through isotopic labelling.

For measuring whether an amino acid base deprotonates water prior as a first step in the deprotonation of a substrate, HDX-MS applied to an enzyme would result in the deuterium-labeled conjugate base of the amino acid. Moreover, one would find the proton from the substrate in the bulk solvent. This could be determined by the use of IR Spectroscopy, with unique stretching frequencies between HOD, H_2O, and D_2O.[175, 176]

To measure whether water is responsible for the protonation of the substrate of an enzyme, a more elegant HDX-MS experiment could be conducted. Quite simply, incubating the enzyme in heavy water in the presence of its substrate would yield a deuterated product. This could be

contrasted with the product of conducting the reaction incubated in heavy water in the absence of the enzyme[xiv] and again measuring the deuteration of the product. If the enzyme assists the deuteration of the substrate by water, one will find more deuterium in the product of the enzyme-catalyzed reaction relative to the reaction absent from the enzyme.

Summary

This dissertation details a comprehensive computational investigation to test the hypothesis that water is an essential cofactor for all enzymes. Firstly, our study utilized an extensive dataset of 1013 enzyme structures and employed various computational methodologies to elucidate the role and behavior of water molecules in enzyme active sites. We discovered that in 1005 out of 1013 enzyme structures, water molecules were located within 5 Å of active site residues, demonstrating a near-universal presence of water in proximity to enzyme active sites. Furthermore, in 993 of these structures, clusters of water molecules formed a contiguous pathway connecting the bulk solvent to the active site. These observations demonstrate that in most enzymes, water exists within reactive distance of the active site and has a means of replenishment. Ultimately, these findings instantiated the possibility that water is universally involved in enzyme catalysis.

We then investigated the electric fields experienced by water molecules within enzyme active sites and compared them to those in the bulk solvent and elsewhere in each enzyme. Using a dataset of 1179 enzyme structures containing a total of 192024 water molecules, our analysis revealed that water molecules in enzyme active sites are subjected to electric fields of greater

[xiv] One would need to find a reaction that occurs reasonably quickly without enzyme assistance.

magnitude (mean 0.031 a.u.) than those elsewhere in the enzyme (mean 0.026 a.u.) and in the bulk solvent (mean 0.021 a.u.). This heightened electric field environment suggests the possibility of a functional adaptation of the active site to modulate the properties of water electrostatically, potentially for the enhancement of catalytic efficiency. At minimum, it implicates the general inadvertent capability of enzyme active sites to influence water in favor of ionization.

To delve deeper into the dynamics of water in enzyme catalysis, we conducted a 2000 ns classical MD simulation on the enzyme triose phosphate isomerase (TPI). Our results showed that the hydroxyl bonds of water molecules near the substrate DHAP experienced electric fields favoring deprotonation, particularly near the electronegative phosphate group. This effect was notably less pronounced when DHAP was placed in a box of water, indicating that the TPI active site amplifies the electric fields experienced by water molecules, thereby favoring their deprotonation.

This finding may offer insight into a novel catalytic mechanism of the enzyme. Glu165 in TPI is commonly believed to deprotonate the sp^3 carbon of the substrate DHAP during catalysis, but these strong fields favoring the formation of hydroxide near DHAP in the active site of TPI raise the question of whether hydroxide is instead responsible for this step.

Moreover, quantum molecular dynamics simulations provided further insight into the behavior of water in enzyme active sites. By subjecting a box of 96 water molecules and eight individual enzyme active sites containing water molecules to 100 fs MD simulations at the HF 3-21G* level using ADMP, we observed that the electric fields on water hydroxyl bonds were far more heterogeneous in active sites compared to bulk solvent. This observation reinforces the

notion that enzyme active sites are finely tuned environments where water molecules are primed to participate in catalytic processes.

Water ionizing events occurred in two of eight active sites simulated, corresponding with electric fields along their hydroxyl bonds exceeding 0.135 a.u., the ionization threshold predicted from a series of SCF scans of the hydroxyl bond of a single water molecule subjected to various external electric fields. Furthermore, elongated hydroxyl bonds greater than 1.15 Å were associated with these ionization events. Additionally, there was significant overlap in the distributions of electric fields and hydroxyl bond lengths between the active sites and bulk solvent simulations. This finding is expected, as not every water molecule is ionized within an active site, especially within the short 100 fs simulation time used.

If enzymes indeed utilize water as a cofactor, this understanding could revolutionize bioorganic applications, particularly in the field of enzyme engineering for synthesis. Several steps can be taken to enhance the catalytic rates of enzymes with the understanding of water's involvement in their mechanisms. If specific cofactor water molecules and their binding sites can be identified on a per-enzyme basis, designing enzymes that better bind water molecules in place could lower the $\Delta G^{\ddagger}$ for that enzyme. Such an effect could also be achieved by identifying specific residues responsible for the activation of a cofactor water molecule and enhancing the electric field they exert on it. Through these methods, the work presented in this dissertation may lead to more effective industrial biocatalysts. This work also opens the possibility of novel drug design. If an enzyme within the body is identified as a novel drug target, and it is found to operate through the involvement of a cofactor water molecule, then the residues that participate in binding and activating that water molecule may be targeted for pharmaceutical intervention.

Through the integration of molecular dynamics and rigorous structural analysis, our research implicates a universal catalytic role of water in enzymatic reactions. This dissertation not only establishes a solid foundation for future investigations into the molecular intricacies of enzyme catalysis but also underscores the indispensable role of water in enzyme chemistry. By beginning to elucidate the dynamic interactions between enzymes and water molecules, our findings pave the way for deeper insights into enzymatic mechanisms and all the practical benefits to humanity guided by such understanding.

REFERENCES CITED

(1) Lewis, C. A.; Wolfenden, R. Uroporphyrinogen decarboxylation as a benchmark for the catalytic proficiency of enzymes. *Proceedings of the National Academy of Sciences* **2008**, *105* (45), 17328-17333. DOI: doi:10.1073/pnas.0809838105.

(2) Lovrien, R. Activation diagrams in enzyme catalysis. *Journal of Theoretical Biology* **1969**, *24* (3), 247-265.

(3) Schramm, V. L. Transition states and transition state analogue interactions with enzymes. *Accounts of Chemical Research* **2015**, *48* (4), 1032-1039.

(4) Giraldo, J.; Roche, D.; Rovira, X.; Serra, J. The catalytic power of enzymes: Conformational selection or transition state stabilization? *FEBS letters* **2006**, *580* (9), 2170-2177.

(5) Åqvist, J.; Kazemi, M.; Isaksen, G. V.; Brandsdal, B. O. Entropy and Enzyme Catalysis. *Accounts of Chemical Research* **2017**, *50* (2), 199-207. DOI: 10.1021/acs.accounts.6b00321.

(6) Jencks, W. P. Binding Energy, Specificity, and Enzymic Catalysis: The Circe Effect. In *Advances in Enzymology and Related Areas of Molecular Biology*, 1975; pp 219-410.

(7) Bruice, T. C.; Pandit, U. K. The effect of geminal substitution ring size and rotamer distribution on the intramolecular nucleophilic catalysis of the hydrolysis of monophenyl esters of dibasic acids and the solvolysis of the intermediate anhydrides. *Journal of the American Chemical Society* **1960**, *82* (22), 5858-5865.

(8) Shurki, A.; Štrajbl, M.; Villa, J.; Warshel, A. How much do enzymes really gain by restraining their reacting fragments? *Journal of the American Chemical Society* **2002**, *124* (15), 4097-4107.

(9) Haldane, J. Enzymes *Longmans Green & Co.* **1930**.

(10) Mabey, W.; Mill, T. Critical review of hydrolysis of organic compounds in water under environmental conditions. *Journal of Physical and Chemical Reference Data* **1978**, *7* (2), 383-415. DOI: 10.1063/1.555572 (acccessed 4/29/2024).

(11) Wolfenden, R.; Snider, M. J. The Depth of Chemical Time and the Power of Enzymes as Catalysts. *Accounts of Chemical Research* **2001**, *34* (12), 938-945. DOI: 10.1021/ar000058i.

(12) Snider, M. J.; Wolfenden, R. Site-bound water and the shortcomings of a less than perfect transition state analogue. *Biochemistry* **2001**, *40* (38), 11364-11371.

(13) Frey, P. A. Review: Strong hydrogen bonding in molecules and enzymatic complexes. *Magnetic Resonance in Chemistry* **2001**, *39* (S1), S190-S198. DOI: https://doi.org/10.1002/mrc.953.

(14) Hosler, J. P.; Shapleigh, J. P.; Mitchell, D. M.; Kim, Y.; Pressler, M. A.; Georgiou, C.; Babcock, G. T.; Alben, J. O.; Ferguson-Miller, S.; Gennis, R. B. Polar residues in helix VIII of subunit I of cytochrome c oxidase influence the activity and the structure of the active site. *Biochemistry* **1996**, *35* (33), 10776-10783.

(15) Kallos, J.; Avatis, K. Study of the polarity of the active site of chymotrypsin. *Biochemistry* **1966**, *5* (6), 1979-1983.

(16) Liu, L.; Quillin, M. L.; Matthews, B. W. Use of experimental crystallographic phases to examine the hydration of polar and nonpolar cavities in T4 lysozyme. *Proceedings of the National Academy of Sciences* **2008**, *105* (38), 14406-14411.

(17) Meng, M.; Bagdasarian, M.; Zeikus, J. G. The role of active-site aromatic and polar residues in catalysis and substrate discrimination by xylose isomerase. *Proceedings of the National Academy of Sciences* **1993**, *90* (18), 8459-8463.

(18) Ko, J.; Murga, L. F.; Wei, Y.; Ondrechen, M. J. Prediction of active sites for protein structures from computed chemical properties. *Bioinformatics* **2005**, *21* (suppl_1), i258-i265.

(19) Bate, P.; Warwicker, J. Enzyme/non-enzyme discrimination and prediction of enzyme active site location using charge-based methods. *Journal of molecular biology* **2004**, *340* (2), 263-276.

(20) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of molecular biology* **1976**, *103* (2), 227-249.

(21) Dewar, M. J.; Storch, D. M. Alternative view of enzyme reactions. *Proceedings of the National Academy of Sciences* **1985**, *82* (8), 2225-2229. DOI: doi:10.1073/pnas.82.8.2225.

(22) Cohen, S. G.; Vaidya, V. M.; Schultz, R. M. Active site of α-chymotrypsin activation by association-desolvation. *Proceedings of the National Academy of Sciences* **1970**, *66* (2), 249-256.

(23) Wolfenden, R. Waterlogged molecules. *Science* **1983**, *222* (4628), 1087-1093.

(24) Warshel, A.; Aqvist, J.; Creighton, S. Enzymes work by solvation substitution rather than by desolvation. *Proceedings of the National Academy of Sciences* **1989**, *86* (15), 5820-5824. DOI: doi:10.1073/pnas.86.15.5820.

(25) Warshel, A. Electrostatic Origin of the Catalytic Power of Enzymes and the Role of Preorganized Active Sites*. *The Journal of Biological Chemistry* **1998**, *273*, 27035 - 27038.

(26) Antosiewicz, J.; McCammon, J. A. Electrostatic and hydrodynamic orientational steering effects in enzyme-substrate association. *Biophysical journal* **1995**, *69* (1), 57-65.

(27) Wade, R. C.; Gabdoulline, R. R.; Lüdemann, S. K.; Lounnas, V. Electrostatic steering and ionic tethering in enzyme–ligand binding: Insights from simulations. *Proceedings of the National Academy of Sciences* **1998**, *95* (11), 5942-5949.

(28) Pollack, R. M. Enzymatic mechanisms for catalysis of enolization: ketosteroid isomerase. *Bioorganic Chemistry* **2004**, *32* (5), 341-353. DOI: https://doi.org/10.1016/j.bioorg.2004.06.005.

(29) Radzicka, A.; Wolfenden, R. A Proficient Enzyme. *Science* **1995**, *267* (5194), 90-93. DOI: doi:10.1126/science.7809611.

(30) Ohamto, W. u. KSI Mechanism. Wikipedia, 2013; p This is the mechanism of the reaction catalyzed by C. testosteroni KSI.

(31) Fried, S. D.; Bagchi, S.; Boxer, S. G. Extreme electric fields power catalysis in the active site of ketosteroid isomerase. *Science* **2014**, *346* (6216), 1510-1514.

(32) Jha, S. K.; Ji, M.; Gaffney, K. J.; Boxer, S. G. Direct measurement of the protein response to an electrostatic perturbation that mimics the catalytic cycle in ketosteroid isomerase. *Proceedings of the National Academy of Sciences* **2011**, *108* (40), 16612-16617.

(33) Wu, Y.; Fried, S. D.; Boxer, S. G. A preorganized electric field leads to minimal geometrical reorientation in the catalytic reaction of ketosteroid isomerase. *Journal of the American Chemical Society* **2020**, *142* (22), 9993-9998.

(34) Zheng, C.; Mao, Y.; Kozuch, J.; Atsango, A. O.; Ji, Z.; Markland, T. E.; Boxer, S. G. A two-directional vibrational probe reveals different electric field orientations in solution and an enzyme active site. *Nature chemistry* **2022**, *14* (8), 891-897.

(35) Naydenova, E.; Roßbach, S.; Ochsenfeld, C. QM/MM study of the uracil DNA glycosylase reaction mechanism: A competition between Asp145 and His148. *Journal of Chemical Theory and Computation* **2019**, *15* (8), 4344-4350.

(36) Przybylski, J. L.; Wetmore, S. D. A QM/QM investigation of the hUNG2 reaction surface: the untold tale of a catalytic residue. *Biochemistry* **2011**, *50* (19), 4218-4227.

(37) Diao, W.; Yan, S.; Farrell, J. D.; Wang, B.; Ye, F.; Wang, Z. Preorganized Internal Electric Field Powers Catalysis in the Active Site of Uracil-DNA Glycosylase. *ACS Catalysis* **2022**, *12* (20), 12488-12499.

(38) Bím, D.; Alexandrova, A. N. Local Electric Fields As a Natural Switch of Heme-Iron Protein Reactivity. *ACS Catalysis* **2021**, *11* (11), 6534-6546. DOI: 10.1021/acscatal.1c00687.

(39) Ji, Z.; Boxer, S. G. β-Lactamases evolve against antibiotics by acquiring large active-site electric fields. *Journal of the American Chemical Society* **2022**, *144* (48), 22289-22294.

(40) Zheng, C.; Ji, Z.; Mathews, I. I.; Boxer, S. G. Enhanced active-site electric field accelerates enzyme catalysis. *Nature Chemistry* **2023**, *15* (12), 1715-1721.

(41) Acosta-Silva, C.; Bertran, J.; Branchadell, V.; Oliva, A. Kemp Elimination Reaction Catalyzed by Electric Fields. *Chemphyschem* **2020**, *21* (4), 295-306. DOI: 10.1002/cphc.201901155 From NLM.

(42) Klijn, J. E.; Engberts, J. B. F. N. Kemp Elimination in Membrane Mimetic Reaction Media: Probing Catalytic Properties of Catanionic Vesicles Formed from Double-Tailed Amphiphiles. *Journal of the American Chemical Society* **2003**, *125* (7), 1825-1833. DOI: 10.1021/ja027772a.

(43) Welborn, V. V.; Ruiz Pestana, L.; Head-Gordon, T. Computational optimization of electric fields for better catalysis design. *Nature Catalysis* **2018**, *1* (9), 649-655. DOI: 10.1038/s41929-018-0109-2.

(44) Baker, E. N.; Hubbard, R. E. Hydrogen bonding in globular proteins. *Progress in biophysics and molecular biology* **1984**, *44* (2), 97-179.

(45) Eigen, M.; De Maeyer, L. Untersuchungen über die Kinetik der Neutralisation. I. *Zeitschrift für Elektrochemie, Berichte der Bunsengesellschaft für physikalische Chemie* **1955**, *59* (10), 986-993. DOI: https://doi.org/10.1002/bbpc.19550591020.

(46) Natzle, W. C.; Moore, C. B. Recombination of hydrogen ion (H+) and hydroxide in pure liquid water. *The Journal of Physical Chemistry* **1985**, *89* (12), 2605-2612.

(47) Bannister, J.; Holzwarth, J.; Gormally, J.; King, T. Iodine laser and fast reactions. *Chemistry in Britain* **1984**, *20*, 227.

(48) Marx, D.; Chandra, A.; Tuckerman, M. E. Aqueous Basic Solutions: Hydroxide Solvation, Structural Diffusion, and Comparison to the Hydrated Proton. *Chemical Reviews* **2010**, *110* (4), 2174-2216. DOI: 10.1021/cr900233f.

(49) North, A. M. Diffusion-controlled reactions. *Quarterly Reviews, Chemical Society* **1966**, *20* (3), 421-440.

(50) Eigen, M.; De Maeyer, L.; Bernal, J. D. Self-dissociation and protonic charge transport in water and. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1958**, *247* (1251), 505-533. DOI: doi:10.1098/rspa.1958.0208.

(51) de Grotthuss, C. Sur la dcomposition de leau et des corps quelle tient en dissolution laide de llectricit galvanique. *Ann Chim* **1806**, *58*, 5473.

(52) Agmon, N. The grotthuss mechanism. *Chemical Physics Letters* **1995**, *244* (5-6), 456-462.

(53) Hassanali, A.; Prakash, M. K.; Eshet, H.; Parrinello, M. On the recombination of hydronium and hydroxide ions in water. *Proceedings of the National Academy of Sciences* **2011**, *108* (51), 20410-20415.

(54) Bai, C.; Herzfeld, J. Special pairs are decisive in the autoionization and recombination of water. *The Journal of Physical Chemistry B* **2017**, *121* (16), 4213-4219.

(55) Kale, S.; Herzfeld, J. Proton Defect Solvation and Dynamics in Aqueous Acid and Base. *Angewandte Chemie International Edition* **2012**, *51* (44), 11029-11032. DOI: https://doi.org/10.1002/anie.201203568.

(56) Geissler, P. L.; Dellago, C.; Chandler, D.; Hutter, J.; Parrinello, M. Autoionization in liquid water. *Science* **2001**, *291* (5511), 2121-2124.

(57) Saitta, A. M.; Saija, F.; Giaquinta, P. V. Ab initio molecular dynamics study of dissociation of water under an electric field. *Physical review letters* **2012**, *108* (20), 207801.

(58) Cassone, G.; Creazzo, F.; Giaquinta, P. V.; Saija, F.; Saitta, A. M. Ab initio molecular dynamics study of an aqueous NaCl solution under an electric field. *Physical Chemistry Chemical Physics* **2016**, *18* (33), 23164-23173.

(59) Cassone, G.; Sponer, J.; Trusso, S.; Saija, F. Ab initio spectroscopy of water under electric fields. *Physical Chemistry Chemical Physics* **2019**, *21* (38), 21205-21212.

(60) Pravda, L.; Berka, K.; Svobodová Vařeková, R.; Sehnal, D.; Banáš, P.; Laskowski, R. A.; Koča, J.; Otyepka, M. Anatomy of enzyme channels. *BMC Bioinformatics* **2014**, *15* (1), 379. DOI: 10.1186/s12859-014-0379-x.

(61) Hassan, S. A.; Hummer, G.; Lee, Y.-S. Effects of electric fields on proton transport through water chains. *The Journal of Chemical Physics* **2006**, *124* (20). DOI: 10.1063/1.2198820 (acccessed 6/4/2024).

(62) Maréchal, A.; Rich, P. R. Water molecule reorganization in cytochrome c oxidase revealed by FTIR spectroscopy. *Proceedings of the National Academy of Sciences* **2011**, *108* (21), 8634-8638.

(63) Sharma, V.; Enkavi, G.; Vattulainen, I.; Róg, T.; Wikström, M. Proton-coupled electron transfer and the role of water molecules in proton pumping by cytochrome c oxidase. *Proceedings of the National Academy of Sciences* **2015**, *112* (7), 2040-2045.

(64) Feynman, R.; Hibbs, A. The path integral formulation of quantum mechanics. *McGraw-Hiill, New York* **1965**.

(65) Schweizer, K. S.; Stratt, R. M.; Chandler, D.; Wolynes, P. G. Convenient and accurate discretized path integral methods for equilibrium quantum mechanical calculations. *The Journal of Chemical Physics* **1981**, *75* (3), 1347-1364.

(66) Pomes, R.; Roux, B. Structure and dynamics of a proton wire: a theoretical study of H+ translocation along the single-file water chain in the gramicidin A channel. *Biophysical journal* **1996**, *71* (1), 19-39.

(67) Hille, B. v. Ion channels of excitable cells. 1992.

(68) Cui, Q.; Karplus, M. Is a “proton wire” concerted or stepwise? A model study of proton transfer in carbonic anhydrase. *The Journal of Physical Chemistry B* **2003**, *107* (4), 1071-1078.

(69) Supuran, C. T. Structure and function of carbonic anhydrases. *Biochemical Journal* **2016**, *473* (14), 2023-2032.

(70) Asada, K. Ascorbate peroxidase – a hydrogen peroxide-scavenging enzyme in plants. *Physiologia Plantarum* **1992**, *85* (2), 235-241. DOI: https://doi.org/10.1111/j.1399-3054.1992.tb04728.x.

(71) Efimov, I.; Badyal, S. K.; Metcalfe, C. L.; Macdonald, I.; Gumiero, A.; Raven, E. L.; Moody, P. C. Proton delivery to ferryl heme in a heme peroxidase: enzymatic use of the Grotthuss mechanism. *Journal of the American Chemical Society* **2011**, *133* (39), 15376-15383.

(72) Levitz, T. S.; Drennan, C. L. Starting a new chapter on class Ia ribonucleotide reductases. *Current Opinion in Structural Biology* **2022**, *77*, 102489. DOI: https://doi.org/10.1016/j.sbi.2022.102489.

(73) Offenbacher, A. R.; Barry, B. A. A proton wire mediates proton coupled electron transfer from hydroxyurea and other hydroxamic acids to tyrosyl radical in class Ia ribonucleotide reductase. *The Journal of Physical Chemistry B* **2020**, *124* (2), 345-354.

(74) Mäntele, W. Infrared and Fourier-Transform Infrared Spectroscopy. In *Biophysical Techniques in Photosynthesis*, Amesz, J., Hoff, A. J. Eds.; Springer Netherlands, 1996; pp 137-160.

(75) Zscherp, C.; Barth, A. Reaction-Induced Infrared Difference Spectroscopy for the Study of Protein Reaction Mechanisms. *Biochemistry* **2001**, *40* (7), 1875-1883. DOI: 10.1021/bi002567y.

(76) Mouret, S.; Baudouin, C.; Charveron, M.; Favier, A.; Cadet, J.; Douki, T. Cyclobutane pyrimidine dimers are predominant DNA lesions in whole human skin exposed to UVA radiation. *Proceedings of the National Academy of Sciences* **2006**, *103* (37), 13765-13770. DOI: doi:10.1073/pnas.0604213103.

(77) Setlow, R. B. Cyclobutane-Type Pyrimidine Dimers in Polynucleotides. *Science* **1966**, *153* (3734), 379-386. DOI: doi:10.1126/science.153.3734.379.

(78) Liu, Z.; Tan, C.; Guo, X.; Kao, Y.-T.; Li, J.; Wang, L.; Sancar, A.; Zhong, D. Dynamics and mechanism of cyclobutane pyrimidine dimer repair by DNA photolyase. *Proceedings of the National Academy of Sciences* **2011**, *108* (36), 14831-14836. DOI: doi:10.1073/pnas.1110927108.

(79) Wang, H.; Chen, X.; Fang, W. Excited-state proton coupled electron transfer between photolyase and the damaged DNA through water wire: a photo-repair mechanism. *Physical Chemistry Chemical Physics* **2014**, *16* (46), 25432-25441.

(80) Marion, A.; Gokcan, H.; Monard, G. Semi-Empirical Born–Oppenheimer Molecular Dynamics (SEBOMD) within the Amber Biomolecular Package. *Journal of Chemical Information and Modeling* **2019**, *59* (1), 206-214. DOI: 10.1021/acs.jcim.8b00605.

(81) Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Charge Transfer and Polarization in Solvated Proteins from Ab Initio Molecular Dynamics. *The Journal of Physical Chemistry Letters* **2011**, *2* (14), 1789-1793. DOI: 10.1021/jz200697c.

(82) Nadig, G.; Van Zant, L. C.; Dixon, S. L.; Merz, K. M. Charge-transfer interactions in macromolecular systems: a new view of the protein/water interface. *Journal of the American Chemical Society* **1998**, *120* (22), 5593-5594.

(83) van der Vaart, A.; Merz, K. M. The role of polarization and charge transfer in the solvation of biomolecules. *Journal of the American Chemical Society* **1999**, *121* (39), 9182-9190.

(84) Khandogin, J.; York, D. M. Quantum descriptors for biological macromolecules from linear-scaling electronic structure methods. *Proteins: Structure, Function, and Bioinformatics* **2004**, *56* (4), 724-737.

(85) Komeiji, Y.; Ishida, T.; Fedorov, D. G.; Kitaura, K. Change in a protein's electronic structure induced by an explicit solvent: an ab initio fragment molecular orbital study of ubiquitin. *Journal of Computational Chemistry* **2007**, *28* (10), 1750-1762.

(86) Cantrill, S. Green fluorescent protein. *Nature Chemistry* **2008**. DOI: 10.1038/nchem.75.

(87) Shinobu, A.; Agmon, N. Proton wire dynamics in the green fluorescent protein. *Journal of Chemical Theory and Computation* **2017**, *13* (1), 353-369.

(88) Shinobu, A.; Agmon, N. Mapping proton wires in proteins: carbonic anhydrase and GFP chromophore biosynthesis. *The Journal of Physical Chemistry A* **2009**, *113* (26), 7253-7266.

(89) Shinobu, A.; Palm, G. J.; Schierbeek, A. J.; Agmon, N. Visualizing proton antenna in a high-resolution green fluorescent protein structure. *Journal of the American Chemical Society* **2010**, *132* (32), 11093-11102.

(90) Frey, P. A. Strong hydrogen bonding in chymotrypsin and other serine proteases. *Journal of physical organic chemistry* **2004**, *17* (6-7), 511-520.

(91) Martínez, L.; Andrade, R.; Birgin, E.; PACKMOL, J. M. A package for building initial configurations for molecular dynamics simulations., 2009, 30. *DOI: https://doi.org/10.1002/jcc 21224*, 2157-2164.

(92) Arcari, P.; Masullo, L.; Masullo, M.; Catanzano, F.; Bocchini, V. A NAD(P)H Oxidase Isolated from the Archaeon Sulfolobus solfataricus Is Not Homologous with Another NADH Oxidase Present in the Same Microorganism: BIOCHEMICAL CHARACTERIZATION OF THE ENZYME AND CLONING OF THE ENCODING GENE*. *Journal of Biological Chemistry* **2000**, *275* (2), 895-900. DOI: https://doi.org/10.1074/jbc.275.2.895.

(93) Ruggiero, A.; Masullo, M.; Ruocco, M. R.; Grimaldi, P.; Lanzotti, M. A.; Arcari, P.; Zagari, A.; Vitagliano, L. Structure and stability of a thioredoxin reductase from Sulfolobus solfataricus: A thermostable protein with two functions. *Biochimica et Biophysica Acta (BBA) -*

Proteins and Proteomics **2009**, *1794* (3), 554-562. DOI: https://doi.org/10.1016/j.bbapap.2008.11.011.

(94) Ogbuji, U. Python and XML. In *a conference on XML*, 2007; p 53.

(95) Chandra, R. V.; Varanasi, B. S. *Python requests essentials*; Packt Publishing Birmingham, UK, 2015.

(96) Uzun, E.; Yerlikaya, T.; Kirat, O. Comparison of python libraries used for web data extraction. *Journal of the Technical University-Sofia Plovdiv Branch, Bulgaria* **2018**, *24*, 87-92.

(97) Cock, P. J.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **2009**, *25* (11), 1422.

(98) Ambethkar, H. *Chemlib*. https://github.com/harirakul/chemlib (accessed 2024.)

(99) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **1983**, *79* (2), 926-935.

(100) Kale, S.; Herzfeld, J.; Dai, S.; Blank, M. Lewis-inspired representation of dissociable water in clusters and Grotthuss chains. *Journal of biological physics* **2012**, *38*, 49-59.

(101) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* **2011**, *12*, 2825-2830.

(102) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **2006**, *65* (3), 712-725. DOI: https://doi.org/10.1002/prot.21123.

(103) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78* (8), 1950-1958. DOI: 10.1002/prot.22711 From NLM.

(104) Jogl, G.; Rozovsky, S.; McDermott, A. E.; Tong, L. Optimal alignment for enzymatic proton transfer: structure of the Michaelis complex of triosephosphate isomerase at 1.2-Å resolution. *Proceedings of the National Academy of Sciences* **2003**, *100* (1), 50-55.

(105) Iyengar, S. S.; Schlegel, H. B.; Voth, G. A. Atom-Centered Density Matrix Propagation (ADMP): Generalizations Using Bohmian Mechanics. *The Journal of Physical Chemistry A* **2003**, *107* (37), 7269-7277. DOI: 10.1021/jp034633m.

(106) Buscagan, T. M.; Perez, K. A.; Maggiolo, A. O.; Rees, D. C.; Spatzal, T. Structural Characterization of Two CO Molecules Bound to the Nitrogenase Active Site. *Angewandte*

Chemie International Edition **2021**, *60* (11), 5704-5707. DOI: https://doi.org/10.1002/anie.202015751.

(107) Gaspari, R.; Rechlin, C.; Heine, A.; Bottegoni, G.; Rocchia, W.; Schwarz, D.; Bomke, J.; Gerber, H.-D.; Klebe, G.; Cavalli, A. Kinetic and Structural Insights into the Mechanism of Binding of Sulfonamides to Human Carbonic Anhydrase by Computational and Experimental Studies. *Journal of Medicinal Chemistry* **2016**, *59* (9), 4245-4256. DOI: 10.1021/acs.jmedchem.5b01643.

(108) Cotton, F. A.; Hazen Jr, E. E.; Legg, M. J. Staphylococcal nuclease: Proposed mechanism of action based on structure of enzyme—thymidine 3′, 5′-bisphosphate—calcium ion complex at 1.5-Å resolution. *Proceedings of the National Academy of Sciences* **1979**, *76* (6), 2551-2555.

(109) Mees, A.; Klar, T.; Gnau, P.; Hennecke, U.; Eker, A. P. M.; Carell, T.; Essen, L.-O. Crystal Structure of a Photolyase Bound to a CPD-Like DNA Lesion After in Situ Repair. *Science* **2004**, *306* (5702), 1789-1793. DOI: doi:10.1126/science.1101598.

(110) Osipiuk, J., Zhou, M., Papazisi, L., Anderson, W.F., Joachimiak, A. Phosphoribosylamine-glycine ligase from Yersinia pestis. Center for Structural Genomics of Infectious Diseases (CSGID): www.wwpdb.org, 2010.

(111) Schiebel, J.; Gaspari, R.; Wulsdorf, T.; Ngo, K.; Sohn, C.; Schrader, T. E.; Cavalli, A.; Ostermann, A.; Heine, A.; Klebe, G. Intriguing role of water in protein-ligand binding studied by neutron crystallography on trypsin complexes. *Nature Communications* **2018**, *9* (1), 3559. DOI: 10.1038/s41467-018-05769-2.

(112) Özkan, E.; Yu, H.; Deisenhofer, J. Mechanistic insight into the allosteric activation of a ubiquitin-conjugating enzyme by RING-type ubiquitin ligases. *Proceedings of the National Academy of Sciences* **2005**, *102* (52), 18890-18895. DOI: doi:10.1073/pnas.0509418102.

(113) Juniar, L.; Tanaka, H.; Yoshida, K.; Hisabori, T.; Kurisu, G. Structural basis for thioredoxin isoform-based fine-tuning of ferredoxin-thioredoxin reductase activity. *Protein Science* **2020**, *29* (12), 2538-2545. DOI: https://doi.org/10.1002/pro.3964.

(114) Mohsen-Nia, M.; Amiri, H.; Jazi, B. Dielectric constants of water, methanol, ethanol, butanol and acetone: measurement and computational study. *Journal of Solution Chemistry* **2010**, *39*, 701-708.

(115) Maryott, A. A.; Smith, E. R. *Table of dielectric constants of pure liquids*; US Government Printing Office, 1951.

(116) *Solvents List SCRF*. https://gaussian.com/scrf/?tabid=7 (accessed 2024)

(117) Simonson, T.; Brooks, C. L. Charge screening and the dielectric constant of proteins: insights from molecular dynamics. *Journal of the American Chemical Society* **1996**, *118* (35), 8452-8458.

(118) Schutz, C. N.; Warshel, A. What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins: Structure, Function, and Bioinformatics* **2001**, *44* (4), 400-417.

(119) Mulliken, R. S. Electronic population analysis on LCAO–MO molecular wave functions. I. *The Journal of chemical physics* **1955**, *23* (10), 1833-1840.

(120) Yates, M.; Callis, P. R. Water as a reactant in the first step of triosephosphate isomerase catalysis. *bioRxiv* **2021**, 2021.2001. 2030.427993.

(121) Marx, D. Proton transfer 200 years after von Grotthuss: Insights from ab initio simulations. *ChemPhysChem* **2006**, *7* (9), 1848-1870.

(122) Brewer, M. L.; Schmitt, U. W.; Voth, G. A. The formation and dynamics of proton wires in channel environments. *Biophysical Journal* **2001**, *80* (4), 1691-1702.

(123) Kelkar, D. A.; Chattopadhyay, A. The gramicidin ion channel: a model membrane protein. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2007**, *1768* (9), 2011-2025.

(124) Warshel, A. Energetics of enzyme catalysis. *Proceedings of the National Academy of Sciences* **1978**, *75* (11), 5250-5254.

(125) Kovalevsky, A. Y.; Hanson, B.; Mason, S.; Yoshida, T.; Fisher, S.; Mustyakimov, M.; Forsyth, V.; Blakeley, M.; Keen, D.; Langan, P. Identification of the elusive hydronium ion exchanging roles with a proton in an enzyme at lower pH values. *Angewandte Chemie* **2011**, *123* (33), 7662-7665.

(126) Ikeda, T.; Saito, K.; Hasegawa, R.; Ishikita, H. The existence of an isolated hydronium ion in the interior of proteins. *Angewandte Chemie International Edition* **2017**, *56* (31), 9151-9154.

(127) Li, T.-R.; Huck, F.; Piccini, G.; Tiefenbacher, K. Mimicry of the proton wire mechanism of enzymes inside a supramolecular capsule enables β-selective O-glycosylations. *Nature Chemistry* **2022**, *14* (9), 985-994.

(128) Frank, R. A.; Titman, C. M.; Pratap, J. V.; Luisi, B. F.; Perham, R. N. A molecular switch and proton wire synchronize the active sites in thiamine enzymes. *Science* **2004**, *306* (5697), 872-876.

(129) Dai, S.; Funk, L.-M.; von Pappenheim, F. R.; Sautner, V.; Paulikat, M.; Schröder, B.; Uranga, J.; Mata, R. A.; Tittmann, K. Low-barrier hydrogen bonds in enzyme cooperativity. *Nature* **2019**, *573* (7775), 609-613.

(130) Tripathi, R.; Noetzel, J.; Marx, D. Exposing catalytic versatility of GTPases: taking reaction detours in mutants of hGBP1 enzyme without additional energetic cost. *Physical Chemistry Chemical Physics* **2019**, *21* (2), 859-867.

(131) Tripathi, R.; Glaves, R.; Marx, D. The GTPase hGBP1 converts GTP to GMP in two steps via proton shuttle mechanisms. *Chemical Science* **2017**, *8* (1), 371-380.

(132) Fried, S. D.; Boxer, S. G. Electric fields and enzyme catalysis. *Annual review of biochemistry* **2017**, *86*, 387-415.

(133) Yang, L.; Dordick, J. S.; Garde, S. Hydration of enzyme in nonaqueous media is consistent with solvent dependence of its activity. *Biophysical journal* **2004**, *87* (2), 812-821.

(134) Das, U.; Wang, L. K.; Smith, P.; Munir, A.; Shuman, S. Structures of Bacterial Polynucleotide Kinase in a Michaelis Complex with Nucleoside Triphosphate (NTP)-Mg2+ and 5′-OH RNA and a Mixed Substrate-Product Complex with NTP-Mg2+ and a 5′-Phosphorylated Oligonucleotide. *Journal of Bacteriology* **2014**, *196* (24), 4285-4292. DOI: doi:10.1128/jb.02197-14.

(135) Riccardi, D.; König, P.; Prat-Resina, X.; Yu, H.; Elstner, M.; Frauenheim, T.; Cui, Q. "Proton holes" in long-range proton transfer reactions in solution and enzymes: A theoretical analysis. *Journal of the American Chemical Society* **2006**, *128* (50), 16302-16311.

(136) Huang, K.-F.; Hsu, H.-L.; Karim, S.; Wang, A.-J. Structural and functional analyses of a glutaminyl cyclase from Ixodes scapularis reveal metal-independent catalysis and inhibitor binding. *Acta Crystallographica Section D: Biological Crystallography* **2014**, *70* (3), 789-801.

(137) *Nearest Neighbors — scikit-learn 0.21.3 documentation*; scikit-learn.org, https://scikit-learn.org/stable/modules/neighbors.html

(138) Papp-Wallace, K. M.; Nguyen, N. Q.; Jacobs, M. R.; Bethel, C. R.; Barnes, M. D.; Kumar, V.; Bajaksouzian, S.; Rudin, S. D.; Rather, P. N.; Bhavsar, S. Strategic approaches to overcome resistance against Gram-negative pathogens using β-lactamase inhibitors and β-lactam enhancers: activity of three novel diazabicyclooctanes WCK 5153, zidebactam (WCK 5107), and WCK 4234. *Journal of medicinal chemistry* **2018**, *61* (9), 4067-4086.

(139) Duan, X.; Gimble, F. S.; Quiocho, F. A. Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. *Cell* **1997**, *89* (4), 555-564.

(140) Chan, P. W.; Yakunin, A. F.; Edwards, E. A.; Pai, E. F. Mapping the reaction coordinates of enzymatic defluorination. *Journal of the American Chemical Society* **2011**, *133* (19), 7461-7468.

(141) Durham, T. B.; Klimkowski, V. J.; Rito, C. J.; Marimuthu, J.; Toth, J. L.; Liu, C.; Durbin, J. D.; Stout, S. L.; Adams, L.; Swearingen, C. Identification of potent and selective hydantoin inhibitors of aggrecanase-1 and aggrecanase-2 that are efficacious in both chemical and surgical models of osteoarthritis. *Journal of Medicinal Chemistry* **2014**, *57* (24), 10476-10485.

(142) de Rosa, M.; Bemporad, F.; Pellegrino, S.; Chiti, F.; Bolognesi, M.; Ricagno, S. Edge strand engineering prevents native-like aggregation in S ulfolobus solfataricus acylphosphatase. *The FEBS journal* **2014**, *281* (18), 4072-4084.

(143) Krzywda, S.; Jaskolski, M.; Rolka, K.; Stawikowski, M. J. Structure of a proteolytically resistant analogue of (NLys) 5SFTI-1 in complex with trypsin: evidence for the direct participation

of the Ser214 carbonyl group in serine protease-mediated proteolysis. *Acta Crystallographica Section D: Biological Crystallography* **2014**, *70* (3), 668-675.

(144) Kettle, J. G.; Alwan, H.; Bista, M.; Breed, J.; Davies, N. L.; Eckersley, K.; Fillery, S.; Foote, K. M.; Goodwin, L.; Jones, D. R. Potent and selective inhibitors of MTH1 probe its role in cancer cell survival. *Journal of medicinal chemistry* **2016**, *59* (6), 2346-2361.

(145) Bowen, H. J. M.; Sutton, L.; Sutton, L. E. *Tables of interatomic distances and configuration in molecules and ions*; Chemical Society, 1958.

(146) Dance, I. The pathway for serial proton supply to the active site of nitrogenase: enhanced density functional modeling of the Grotthuss mechanism. *Dalton Transactions* **2015**, *44* (41), 18167-18186.

(147) Liu, J.; He, X.; Zhang, J. Z.; Qi, L.-W. Hydrogen-bond structure dynamics in bulk water: insights from ab initio simulations with coupled cluster theory. *Chemical science* **2018**, *9* (8), 2065-2073.

(148) Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. **2012**.

(149) Sun, D.; Liao, D.-I.; Remington, S. J. Electrostatic fields in the active sites of lysozymes. *Proceedings of the National Academy of Sciences* **1989**, *86* (14), 5361-5365.

(150) Vaissier, V.; Sharma, S. C.; Schaettle, K.; Zhang, T.; Head-Gordon, T. Computational optimization of electric fields for improving catalysis of a designed Kemp eliminase. *ACS catalysis* **2018**, *8* (1), 219-227.

(151) Shehadi, I. A.; Yang, H.; Ondrechen, M. J. Future directions in protein function prediction. *Molecular Biology Reports* **2002**, *29* (4), 329-335. DOI: 10.1023/A:1021220208562.

(152) Suydam, I. T.; Snow, C. D.; Pande, V. S.; Boxer, S. G. Electric fields at the active site of an enzyme: Direct comparison of experiment with theory. *science* **2006**, *313* (5784), 200-204.

(153) Siddiqui, S. A.; Stuyver, T.; Shaik, S.; Dubey, K. D. Designed Local Electric Fields─Promising Tools for Enzyme Engineering. *JACS Au* **2023**, *3* (12), 3259-3269.

(154) *The PyMOL Molecular Graphics System*; (accessed 2024)

(155) DeLano, W. L.; Bromberg, S. PyMOL user's guide. *DeLano Scientific LLC* **2004**, *629*.

(156) Mennucci, B. Polarizable continuum model. *WIREs Computational Molecular Science* **2012**, *2* (3), 386-404. DOI: https://doi.org/10.1002/wcms.1086.

(157) *Gaussian 16, Revision C.01*; 2016.

(158) *GaussView, Version 6.1*; 2016.

(159) Bekker, H.; Berendsen, H.; Dijkstra, E.; Achterop, S.; Vondrumen, R. v.; Vanderspoel, D.; Sijbers, A.; Keegstra, H.; Renardus, M. Gromacs-a parallel computer for molecular-dynamics simulations. In *4th international conference on computational physics (PC 92)*, 1993; World Scientific Publishing: pp 252-256.

(160) Brown, F. K.; Kollman, P. A. Molecular dynamics simulations of "loop closing" in the enzyme triose phosphate isomerase. *Journal of Molecular Biology* **1987**, *198* (3), 533-546. DOI: https://doi.org/10.1016/0022-2836(87)90298-1.

(161) Kulkarni, Y. S.; Liao, Q.; Petrović, D. a.; Krüger, D. M.; Strodel, B.; Amyes, T. L.; Richard, J. P.; Kamerlin, S. C. Enzyme architecture: Modeling the operation of a hydrophobic clamp in catalysis by triosephosphate isomerase. *Journal of the American Chemical Society* **2017**, *139* (30), 10514-10525.

(162) Richard, J. P.; Amyes, T. L.; Malabanan, M. M.; Zhai, X.; Kim, K. J.; Reinhardt, C. J.; Wierenga, R. K.; Drake, E. J.; Gulick, A. M. Structure–Function Studies of Hydrophobic Residues That Clamp a Basic Glutamate Side Chain during Catalysis by Triosephosphate Isomerase. *Biochemistry* **2016**, *55* (21), 3036-3047. DOI: 10.1021/acs.biochem.6b00311.

(163) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural population analysis. *The Journal of Chemical Physics* **1985**, *83* (2), 735-746. DOI: 10.1063/1.449486 (acccessed 5/24/2024).

(164) Davidson, E. R.; Feller, D. Basis set selection for molecular calculations. *Chemical Reviews* **1986**, *86* (4), 681-696.

(165) Hehre, W. J. Ab initio molecular orbital theory. *Accounts of Chemical Research* **1976**, *9* (11), 399-406.

(166) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *The Journal of chemical physics* **1992**, *96* (9), 6796-6806.

(167) Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of chemical physics* **1989**, *90* (2), 1007-1023.

(168) Woon, D. E.; Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *The Journal of chemical physics* **1993**, *98* (2), 1358-1371.

(169) Svensson, M.; Humbel, S.; Froese, R. D.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: a multilayered integrated MO+ MM method for geometry optimizations and single point energy predictions. A test for Diels− Alder reactions and Pt (P (t-Bu) 3) 2+ H2 oxidative addition. *The Journal of Physical Chemistry* **1996**, *100* (50), 19357-19363.

(170) Lundberg, M.; Morokuma; Keiji. The Oniom method and its applications to enzymatic reactions. *Multi-scale quantum models for biocatalysis: modern techniques and applications* **2009**, 21-55.

(171) Serdakowski, A. L.; Dordick, J. S. Enzyme activation for organic solvents made easy. *Trends in Biotechnology* **2008**, *26* (1), 48-54.

(172) Rogers, R. D.; Seddon, K. R. Ionic liquids--solvents of the future? *Science* **2003**, *302* (5646), 792-793.

(173) James, E. I.; Murphree, T. A.; Vorauer, C.; Engen, J. R.; Guttman, M. Advances in hydrogen/deuterium exchange mass spectrometry and the pursuit of challenging biological systems. *Chemical reviews* **2021**, *122* (8), 7562-7623.

(174) Garcia, R. A.; Pantazatos, D.; Villarreal, F. J. Hydrogen/deuterium exchange mass spectrometry for investigating protein-ligand interactions. *Assay and drug development technologies* **2004**, *2* (1), 81-91.

(175) Barker, E.; Sleator, W. The infrared spectrum of heavy water. *The Journal of Chemical Physics* **1935**, *3* (11), 660-663.

(176) Torii, H.; Akazawa, T. Modeling of the Hydrogen Bond-Induced Frequency Shifts of the HOH and HOD Bending Modes of Water. *The Journal of Physical Chemistry A* **2024**.